

## INFORMATION EXTRACTION IN TOMB PIT USING HYPERSPECTRAL DATA

Xueyun Yang<sup>1</sup>, Miaole Hou<sup>1,2,\*</sup>, Shuqiang Lyu<sup>1,2,\*</sup>, Sheng Ma<sup>3</sup>, Zhenhua Gao<sup>3</sup>, Shuzhang Bai<sup>3</sup>, Mingyan GU<sup>2</sup>, Yiyi Liu<sup>2</sup>

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

### Commission III, WG III/4

**KEY WORDS:** Hyperspectral data, Principal component analysis, Historical remains, Band math, Density slice, Morphology

### ABSTRACT:

Hyperspectral data has characteristics of multiple bands and continuous, large amount of data, redundancy, and non-destructive. These characteristics make it possible to use hyperspectral data to study cultural relics. In this paper, the hyperspectral imaging technology is adopted to recognize the bottom images of an ancient tomb located in Shanxi province. There are many black remains on the bottom surface of the tomb, which are suspected to be some meaningful texts or paintings. Firstly, the hyperspectral data is preprocessing to get the reflectance of the region of interesting. For the convenient of compute and storage, the original reflectance value is multiplied by 10000. Secondly, this article uses three methods to extract the symbols at the bottom of the ancient tomb. Finally we tried to use morphology to connect the symbols and gave fifteen reference images. The results show that the extraction of information based on hyperspectral data can obtain a better visual experience, which is beneficial to the study of ancient tombs by researchers, and provides some references for archaeological research findings.

### 1. INTRODUCTION

Exploring information of the tomb helps to probe the progress of human civilization, reflects social civilization of different periods from different aspects, and is valuable information for studying historical events, which has high research value. At the same time, some symbols cannot be identified by the naked eye due to the serious effects of natural and human factors during the discovery and excavation of the tomb. However, the traditional extraction method is quite complicated and it is difficult to extract information without artificial destruction. Therefore, a non-destructive technique is needed to extract information effectively, accurately, and scientifically without destroying the tomb structure.

Hyperspectral is one of the remote sensing science and technology. Compared with traditional remote sensing technologies, hyperspectral is a comprehensive remote sensing technology. It has a great number of bands from visible to near infrared range (Tan K L, 2005). The ancient tomb can be explored using hyperspectral techniques without destroying the tomb material. At present, hyperspectral technology has been used in the relics studying and made a lot of research results. Tan et al. systematically explained these innovative application research results of hyperspectral remote sensing technology in archaeology (Tan K L, 2005). Zhou et al. used the classification method of Neural Network to get covered information in grease regions and established the linear regression model. She effectively extracted the grease hidden information (Zhou P P, 2017). Athos Agapiou et al. used hyperspectral techniques to detect underground structures and determine the distribution of underground buildings by analysed the anomalous spectra (Agapiou A, 2010). Hou et al. automatically extracted the mural

manuscript information through combining principle component analysis (PCA) and maximum likelihood classification based on hyperspectral data, but the accuracy of extraction should be improved (Hou M L, 2014). Sun M et al. propose to evaluate the flaking by automatically analysed hyperspectral images that were scanned at the site. The average spectral data from each ROI and its group label were used to train classification model. The results show that the experiment is effective (Sun M, 2015). Rohani N et al. cast the problem of pigment identification in a novel way by decomposing the spectrum into pure pigments. The results clearly demonstrated more accurate estimation of pigment composition than purely distance-based methods such as spectral angle mapping (SAM) and spectral correlation mapping (SCM) (Rohani N, 2016). George S. used a hyperspectral image dataset to estimate the functional distortion of the scan angle and distance using a geometric model to study the geometric distortion of one of several quality parameters in the spectral image caused by the camera translation phase (George S, 2016). Gong et al. used SAM to classify and identify the pigments used in paintings. Gong explored the spectral angle mapping of pigments by practicing spectral angle mapping of known pigments (Gong M T, 2014).

This paper uses hyperspectral data to extract black symbols that are suspected of ancient texts or patterns at the bottom of ancient tombs. And then, we try to use morphological processing methods to process the extracted images to obtain multiple images, which will provide reference for archaeologists' research. The paper is organized as follows: In section 2, the experimental data and preprocessing are introduced. In section 3, the research methods and principles

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

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

are discussed. In section 4, the experimental results and analyse are presented. Finally, a conclusion is given in section 5.

## 2. EXPERIMENTAL DATA AND PREPROCESSING

### 2.1 Experimental data

The experimental area is an ancient tomb in the south of Shanxi province, as shown in Figure1. The data was captured by a hyperspectral imaging (HSI) camera, VNIR400H. The wavelength of the instrument is from 400 nm to 1000 nm with the spectral resolution of 0.6 nm.

Due to the excavation process and the influence of the environment itself, the black symbols at the bottom of ancient tombs are difficult to identify with the naked eye. Therefore, this area is selected as the study area of this paper. Figure 2(a) is a camera photograph, and Figure 2(b) is a hyperspectral image of the study area.



Figure1. Bottom of the ancient tomb picture



(a) Camera picture (b) Hyperspectral true-color image  
 Figure2. Experimental Area

### 2.2 Preprocessing

The original curve of each pixel which is collected by the hyperspectral imaging camera, reflects the light intensity (DN) of this pixel in each band (Wu T, 2016). The collected data is mixed with instrument noise, light illumination and other factors, so the acquired hyperspectral data needs to be preprocessed. The acquired hyperspectral raw data is subjected to radiation correction to remove the instrument noise, and the gray value data of the image is converted into reflectance data. For the convenient of computing and storage, the reflectivity retains 2 decimal places and then multiplies by 10,000 to form integer data. The correction formula is as follows:

$$Ref = \frac{Date - Dark}{White - Dark} \times 10000 \quad (1)$$

Where Ref = Reflectance data  
 Date = really collected data  
 Dark = Dark current data  
 White = Whiteboard data

## 3. PRINCIPLE AND METHOD

The main steps of data processing are shown in Figure 3:

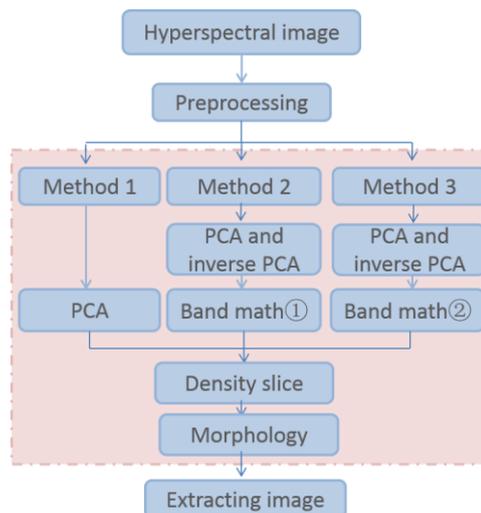


Figure3. Data processing steps

### 3.1 Principal Component Analysis

The study selects a hyperspectral image which taken from the bottom of an ancient tomb. The black symbols are scattered over the image. It is the area of interest in this article. In the 1040 bands, 90 bands with excessive noise are excluded and the remaining 950 bands are involved. The PCA method is used to extract most of the useful information of the image on first 10 principal components, which can concentrate a large amount of information on a small amount of principal component information. This method not only can retain and extract valid information but also reduce the complexity of the work.

### 3.2 Band math

After PCA, the image is restored to RGB color space by using inverse transformation. As shown in Figure 4, the first five principal components contain most of the information, so the paper selects the top five principal components for inverse PCA.

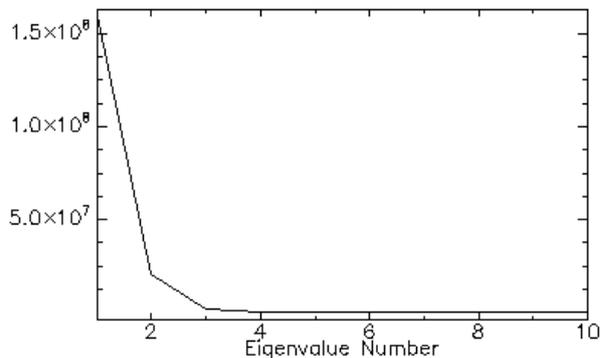


Figure4. Principal component eigenvalues

This paper extracts spectral curves of several points of interest and other parts. It is observed that the point of interest has low reflectivity in the visible and near-infrared bands, and other parts have low reflectivity in the visible light band but have high reflectivity in the near infrared wave band. Therefore, in order to extract the required information, we find out the bands with more obvious reflection contrasts of the symbol information at the bottom of the grave to perform band

operations. The spectral curve is shown in Figure 5. Among them, rose, sky blue, dark blue and green curves represent the spectral curve of the background point. And yellow, black and red curves are the spectral curves of points of interest.

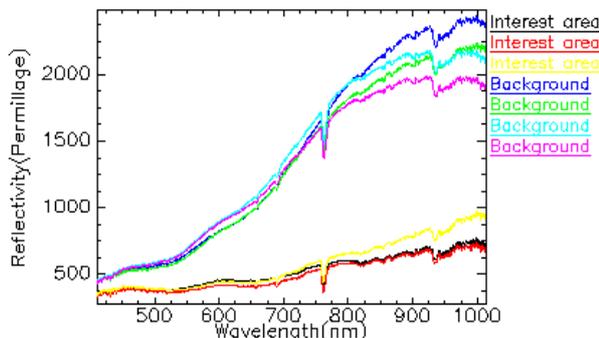


Figure5. Spectral curve

As shown in Figure.5, it can be observed that the spectral curves of the regions of interest in the 900nm to 1000nm bands differ greatly. After the band math operation, we can make the region of interest becomes negative, other parts are positive.:

$$\frac{((B1-1000) + (B2-1000))}{(B1+B2)} = \frac{(B1+B2-2000)}{(B1+B2)} \quad (2)$$

Where B1 = 900nm band  
 B2 is the 550nm band.

According to the characteristics of the spectral curve, we imitate the normalized vegetation index to distinguish between the region of interest and other parts, and use the band calculation formula as equation (3):

$$\frac{((B1-1000) + (B2-1000))}{(B1+B2)} \quad (3)$$

Where B1= 900nm band  
 B2 = 550nm band.

### 3.3 Density slice

Density slice is a classification method applied to single-band grayscale images. Based on the single band image after PCA and band operation, this paper uses density slice to make the contour of the region of interest clearer. The pixels in the region of interest are separated from the image to form a class and set to black. The pixels in other parts are separated to form a class and set to white. The profile of the region of interest is highlighted, leading to a contrast comparison with other parts. Density segmentation in remote sensing is generally equally spaced. However, the density segmentation applied in this paper is to separate the region of interest from other portions and it is not equally spaced.

### 3.4 Morphology

Mathematical morphology basic operations include dilation, erosion, opening and closing. According to the characteristics of these operations, this paper attempts to combine the operations to connect or disconnect the contours of the neighbouring elements to the contours of the region of interest. Through different combinations of operations, different results are obtained to provide reference for archaeologists to identify the symbols at the bottom of ancient tombs.

## 4. RESULTS AND ANALYSIS

### 4.1 Principal Component Analysis

The true-color image and the first component image after PCA are shown in Figure 6. We observe that the outline of the region of interest is enhanced in the first component image. However, it is greatly reduced in the true-color image due to the damage from soil and other materials.

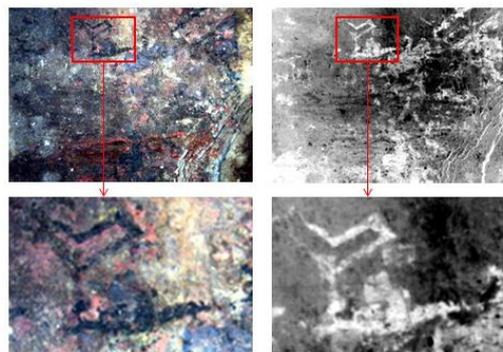


Figure6. Comparison of PCA before and after

### 4.2 Band math

As shown in Figure 7, (a) is a hyperspectral true-color image, the region of interest is mixed with other color substances, and (b) is an image obtained through band operation  $(B1+B2-2000)/(B1+B2)$ . It can be seen that the brightness of the region of interest is low and the brightness of the other portions is high, making the region of interest distinguishable from other regions.

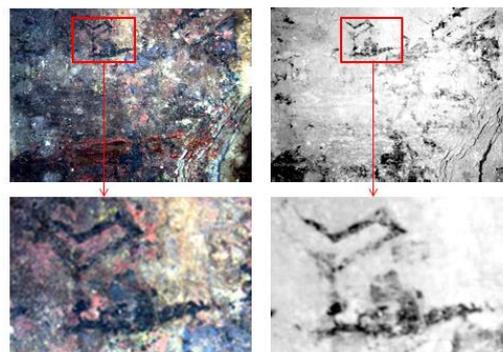


Figure7. Comparison of formula (2) before and after

As shown in Figure 8, (a) is a hyperspectral true color image, and (b) is an image obtained after the band operation  $((10000-B1)-B2)/((10000-B1)+B2)$ . The brightness of the region of interest is high, and the brightness of the other portions is low, highlighting the contour of the region of interest.

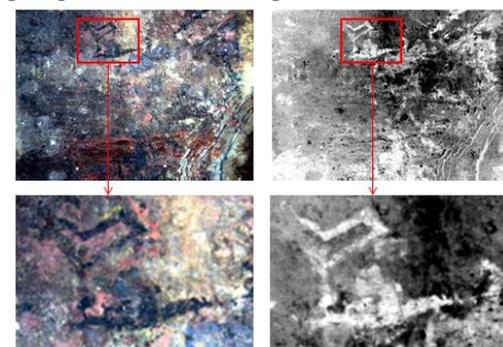


Figure8. Comparison of formula (3) before and after

### 4.3 Density slice

Use density slice to process the image after PCA, formula (2) and formula (3), the pixels of the region of interest and other portions of the region of interest are separated from the image and set as black and white, respectively. The results of the processing are shown in Figure 9 (a), (b), and (c) below.

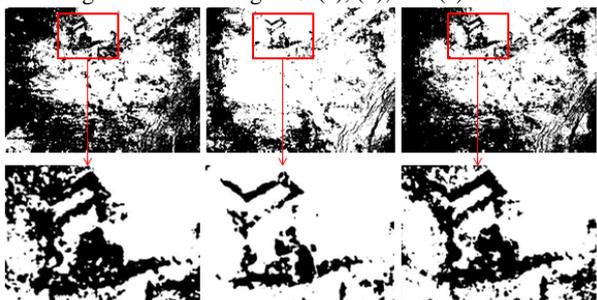


Figure 9. Image after density slice based on (a) the first principal component image after PCA, (b) band math image using equation 1, (c) band math image using equation 2

### 4.4 Morphology

The purpose of this paper is to extract the black materials which are suspected to be ancient texts or patterns at the bottom of the tomb, and to help archaeologists identify them by providing various methods. Thus, it is necessary to try more morphological combinatorial operations to process the images extracted. Figure 10 shows the processing of the PCA processing image. (a) One opening and closing operation. (b) One opening and closing operation, and two erosion and dilation operations. (c) Two erosion and expansion operations and one opening and closing operation. (d) Two opening and closing operations and one erosion and expansion operation. (e) Two erosion and expansion operations.

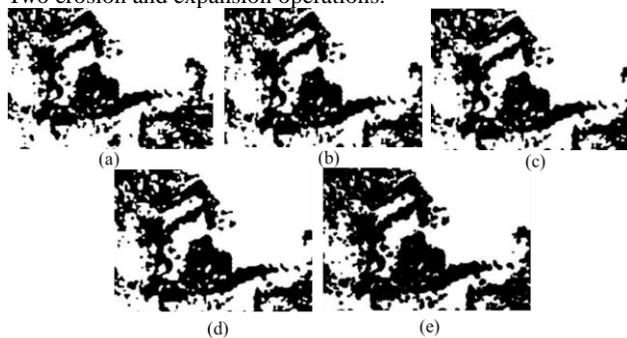


Figure10. PCA processing results

As shown in Figure 11, the image after formula (2) density slice is processed. (a) One opening and closing operation. (b) Three times of erosion and dilation operations and one opening and closing operation. (c) One opening and closing operation and one erosion and expansion operation. (d) Two erosion and expansion operations and two opening and closing operations. (e) Three erosion and expansion operations.

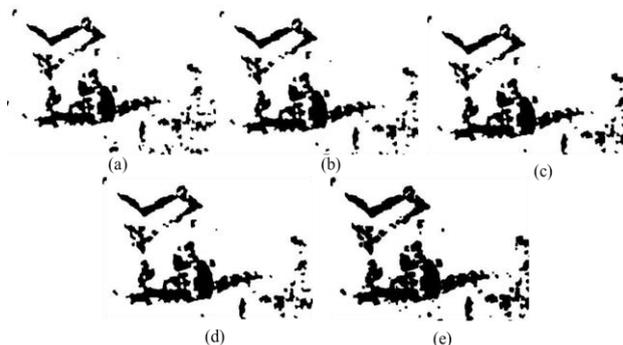


Figure11. Formula (2) processing results

As shown in Figure 12, the image after formula (3) density slice is processed. (a) One erosion and dilation operation. (b) One opening and closing operation and one erosion and dilation operation. (c) Three times of erosion and one expansion operation. (d) One opening and closing operation. (e) Two opening and closing operations.

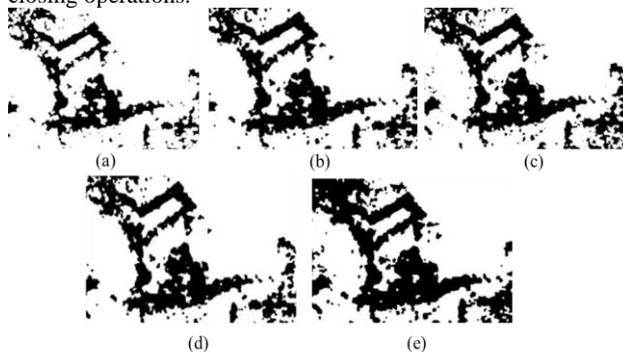


Figure12. Formula (3) processing results

## 5. CONCLUSION

This article analyses the principal component analysis, the band operation and the density slice method, and then observes the spectral curves of the regions of interest and those of other parts. Two equations for wave band operation are proposed and compared with the principal component analysis method. Finally, the most suitable method for the extraction of symbol information of ancient tombs is determined. The results show that principal component analysis can retain valid information, but the extraction results are not good. The result of the band operation equation  $((B1-1000) + (B2-1000)) / (B1+B2)$  is not very accurate, resulting in indistinct distinction between the interest region and the other portion values when the extracted density is segmented. The band operation equation  $((10000-B1)-B2) / ((10000-B1)+B2)$  makes the value of the region of interest close to 1 and the value of other portions close to 0, finding the critical value is a good way to extract it. Finally, we try to connect the symbols with morphology and give fifteen possibilities for archaeologists.

## 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

Agapiou, A., Hadjimitsis, D., Sarris A., Themistocleous, K., & Papadavid, G., 2010. Hyperspectral Ground Truth Data for the Detection of Buried Architectural Remains, Digital Heritage. Springer Berlin Heidelberg, pp.318-331.

George S., Hardeberg J Y., 2016. Estimation and Correction of Geometric Distortion in Pushbroom Hyperspectral System for Imaging Art Paintings. *Electronic Imaging*, 2016(12), pp.1-4.

Gong, M. T., & Feng, P. L. (2014). Preliminary study on the application of hyperspectral imaging in the classification of and identification Chinese traditional pigments classification—a case study of spectral angle mapper. *Sciences of Conservation & Archaeology*.

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.89-92.

Rohani, N., Salvant, J., Bahaadini, S., Cossairt, O., Walton, M., & Katsaggelos, A., 2016. Automatic pigment identification on roman Egyptian paintings by using sparse modeling of hyperspectral images. *Signal Processing Conference*, .pp.2111-2115.

Sun, M., Zhang, D., Wang, Z., Ren, J., Chai, B., & Sun, J., 2015. What's Wrong with the Murals at the Mogao Grottoes: A Near-Infrared Hyperspectral Imaging Method,*Scientific Reports*,pp. 5:14371.

Tan, K. L., 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.

Wu, T., Li, G., Yang, Z., Zhang, H., Lei, Y., & Wang, N., 2017. Shortwave Infrared Imaging Spectroscopy for Analysis of Ancient Paintings. *Applied Spectroscopy*, 71(5), pp.977.

Zhou, P. P., Hou, M. L., Zhao, X., 2017. Virtual restoration of ancient painting stains based on classified linear regression of hyper-spectral image. *Geomatics World*. 24(3), pp.113-118.