NWU Raman Spectral Imaging Toolbox (NWU-RSIT) Operator manual

Yishen Cheng, Shuang Wang*

Institute of Photonics and Photonics Technology, Northwest University, Xi'an 710127, Shaanxi, China

Declaration

NWU-RSIT toolbox is developed by Advanced Spectral Imaging Group, Institute of Photonics and Photonics Technology, Northwest University, Xi'an 710127, Shaanxi, China. NWU-RSIT is developed for scientific research, NOT for commercial use.

If you need more detailed information about the software usage and operation or have certain suggestions for software improvement, please feel free to contact Dr. Shuang Wang, Institute of Photonics and Photonics Technology (Email: swang@nwu.edu.cn; wsnwuphy@163.com).

The NWU-RSIT can be downloaded at https://github.com/wsnwuphy/NWU- Raman spectral imaging toolbox.git.

1. Introduction

NWU-RSIT toolbox is designed for Raman spectral multivariate analysis imaging. Its main functions include the Univariate imaging, hierarchical clustering analysis (HCA), K-means clustering (KCA), Principal component analysis imaging and Spectral unmixing algorithms, such as vertex component analysis (VCA) and N-FINDR.

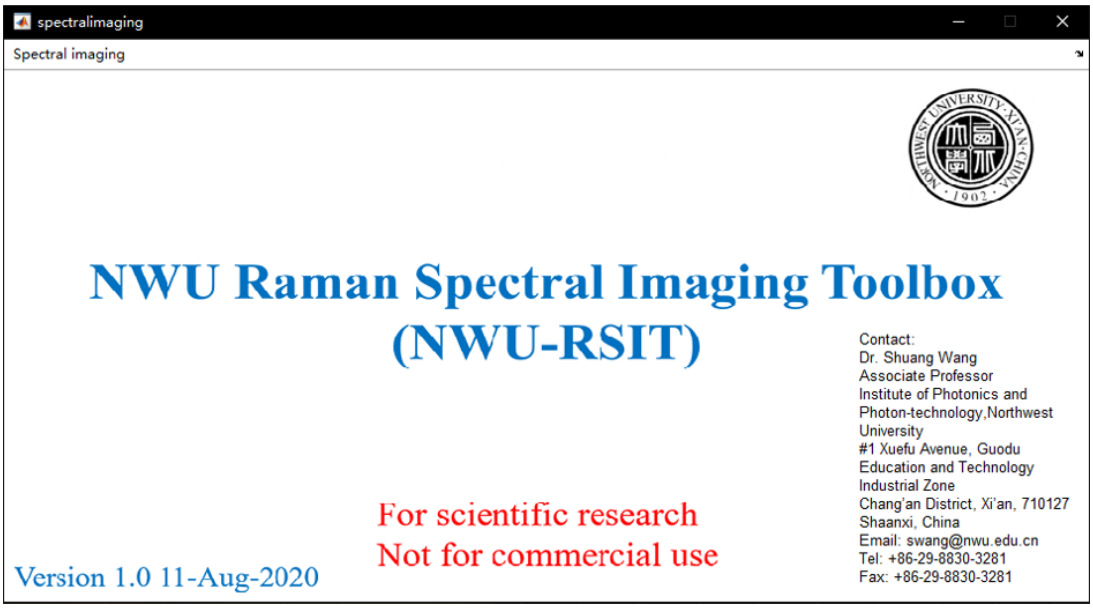


Figure 1. The main interface of the NWU-RSIT toolbox.

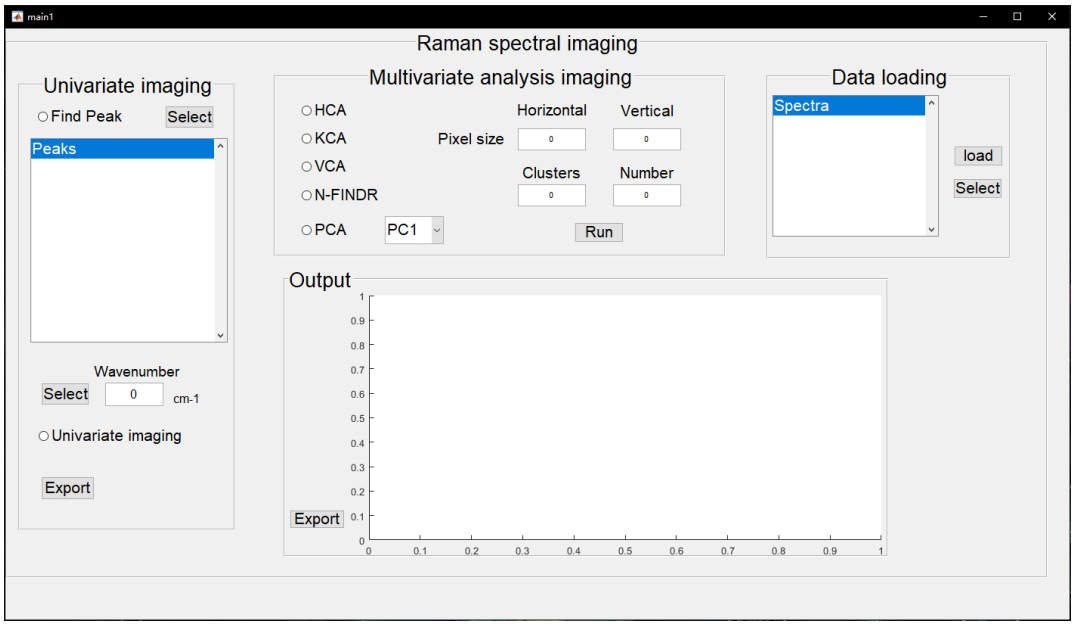


Figure 2. The main interface of the Raman spectral imaging.

2.1 Principal Component Analysis

The PCA represents a multivariable analysis technique. It simplifies a complex dataset by reducing the data dimension and extracts significant features from the dataset. Therefore, the PCA can identify subtle differences between different types of samples. Usually, the first principal component (PC) can explain the largest variance in a dataset, and variances explained by subsequent PCs decrease gradually. In addition, the PC scores reflects a difference between different categories, and each PC loading is related to the characteristic spectrum through the PC scores. In order to show the spatial distribution of scores, each vector of the score matrix $\boldsymbol{F}$ can be reshaped into p images of dimension $N_r \times N_l \times 1$, which represent the scores of p principal components in each pixel. Generate a pseudo-color image based on the score.

2.2 Hierarchical Clustering Analysis

Hierarchical clustering analysis is a commonly used algorithm for generating false color maps from hyperspectral data sets. This algorithm is implemented by generating a distance diagonal matrix from the data set to quantify the spectral similarity. It classifies the spectra with high similarity or the small distance, into a cluster, and finally generates a cluster tree representation through iteration. Then, according to the numbers of different clusters, each cluster is assigned a color to generate a pseudo-color image.

2.3 K-means Clustering Analysis

The well-known HCA are single-link and complete-link, however, the most popular and simplest partitional algorithm is KCA, which have a rich application in different scientific fields. Similar to HCA, the Euclidean distance among the spectra can be calculated and used satisfactorily in KCA. The acquired hyperspectral dataset is divided into k groups, and k objects are randomly selected as the initial clustering center. Then, the distance between each object and each sub-cluster center is calculated for assigning a cluster center to each spectrum. For each sample assigned, the center of the cluster is recalculated according to the existing spectrums in the cluster, which will be repeated until the cluster centers exhibit no changes. Using SSE (Sum of the Squared Error) as the objective function of clustering, two different clusters generated by running KCA

twice, select the minimum value of SSE. Assign colors to spectra with the same centroid to generate a pseudo-color map.

2.4 Vertex Component Analysis

Being an unsupervised hyperspectral unmixing algorithm, VCA algorithm assumes that in a linear hybrid model, each pixel can be regarded as an N-dimensional vector in an N-dimensional Euclidean space, where the wavenumber of the data set is defined as a coordinate in this space. When all spectra are defined in the N-dimensional space, they are contained in an N-dimensional simplex, which is a convex polyhedron with (n+1) nodes and a generalization of a triangle and a tetrahedron. It utilizes the fact that each spectrum belongs to a simplex, and the vertices of the simplex represent the most extreme spectrum in the data set, which presented a pure spectrum.

2.5 N-FINDR

N-FINDR finds all endmembers by looking for simplex with maximum volume(*3, 4*). In a N-dimensional space, each spectrum is treated as a point, which forms a convex simplex, and the endmembers are among the vertices. In a simplex with p vertices, the volume of a monomer consisting of p end elements is the largest. The algorithm iterates to find the endmember spectrum with the maximum volume. When the position of the terminal element is determined, the original spectrum as the endmember is found by looking up the original data set. The spectral contribution of the end elements at each pixel is calculated by the least square method, and then each pixel can be represented by a linear combination of the endmember.

2. Operation Guide

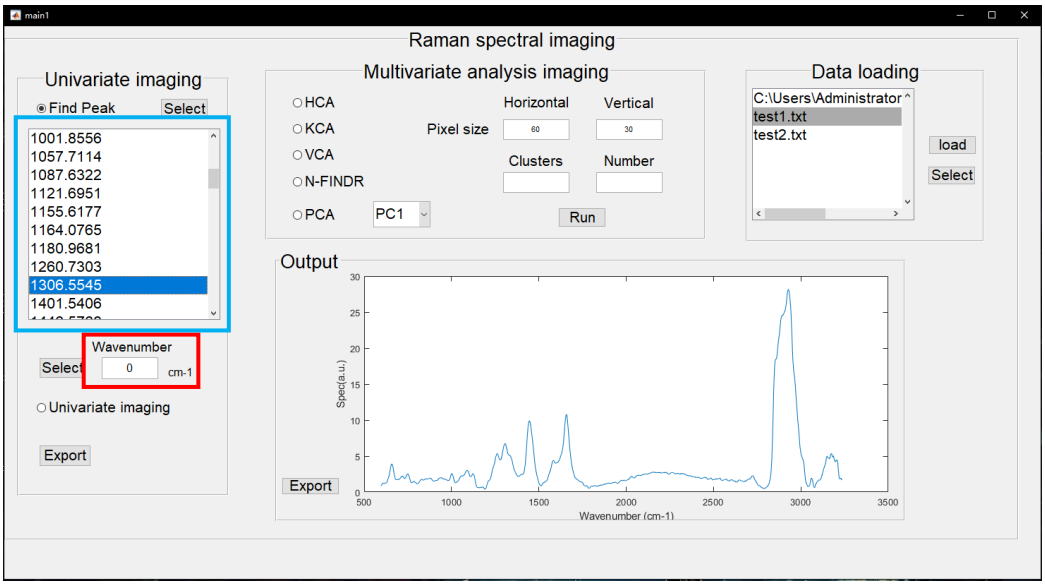
3.1 Data Loading

After opening the software, the operator can click “File > load data” button on the main interface to enter the data loading interface. At present, the NWU-RSIT accepts spectra data in *.mat and *.txt file formats. If the spectral data are given in a .mat file format, it needs to include the spectra intensity matrix ($N \times M$), a wavenumber variable ($M \times 1$). where M denotes the number of spectral features, and N represents the total

number of spectra. If the spectral data are given a *.txt file format, it is allowed to load multiple. The txt type spectrum data usually only has two columns, one column is the wavenumber, and the other column is the Raman spectrum intensity. It is worth noting that the beginning and end of the selected wavenumber range in all spectra should be consistent.

3.2 Univariate imaging

First, input the image size parameters in "Horizontal" and "Vertical", and then input the peak position in the red box to generate the image; Or import a single spectrum and do the following: "Find peak >Click on the peak>Select>Univariate Imaging ", using the peak finding function to select the peak position in the blue box to generate the image.



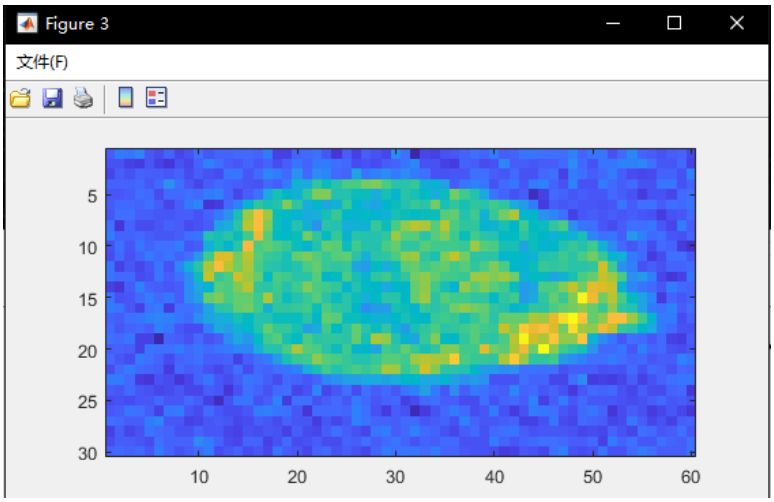


Figure 3. Univariate imaging module, single spectral peak search and univariate image.

3.3 Multivariate analysis imaging

All algorithms need to input the pixel size in “Horizontal” and “Vertical”. In the blue box, the clustering algorithm need to assign the number of input clusters in “Clusters” box, and the spectral decomposition algorithm need to assign the number of input endnumbers in “Number” box, select the desired function in the red box (select only one, select multiple will not run). PCs needs to be selected for principal component analysis. Click “Run” after input parameters. The generated image will be displayed in a pop-up window, and the resulting spectrum will be displayed in "Output". You can click “Export” to Export the resulting spectrum.

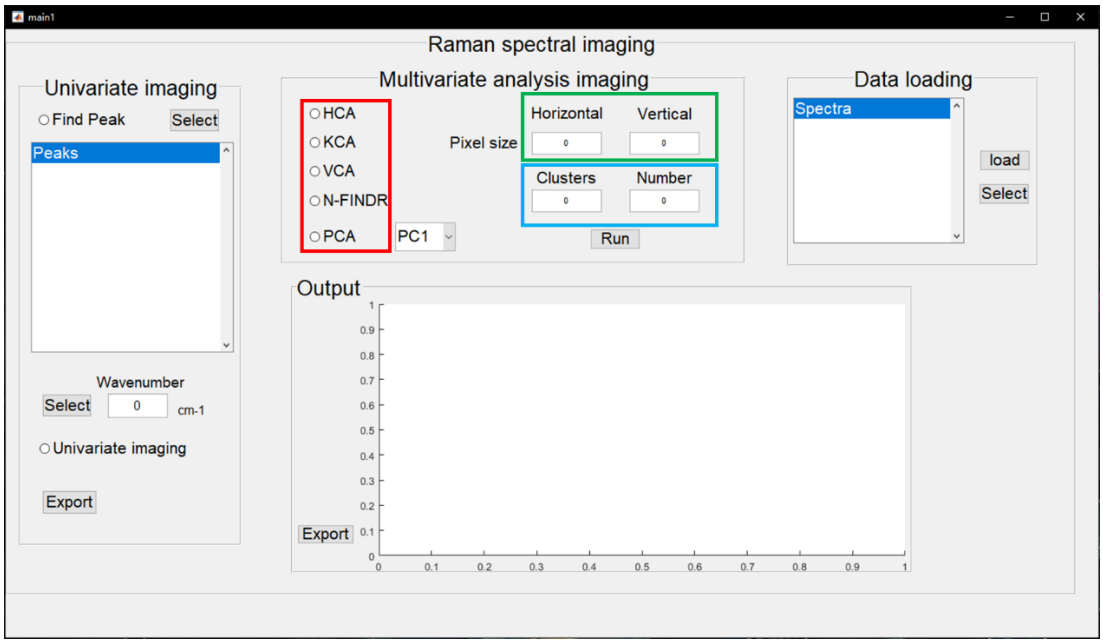
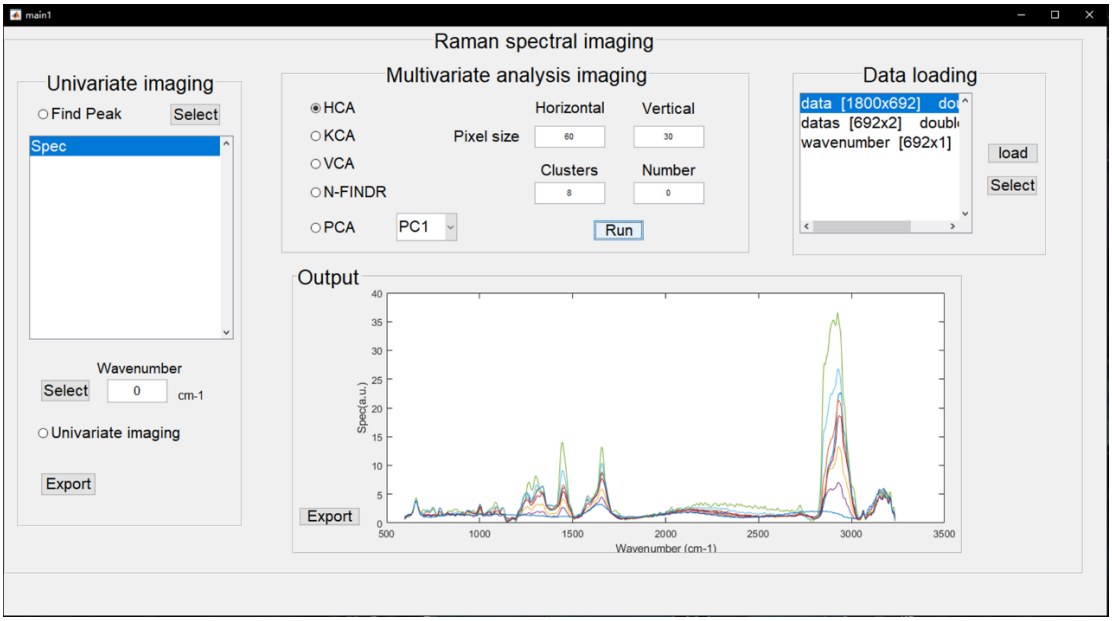


Figure 4. The green and blue boxes are the input parameters, and the red box are the algorithm selection



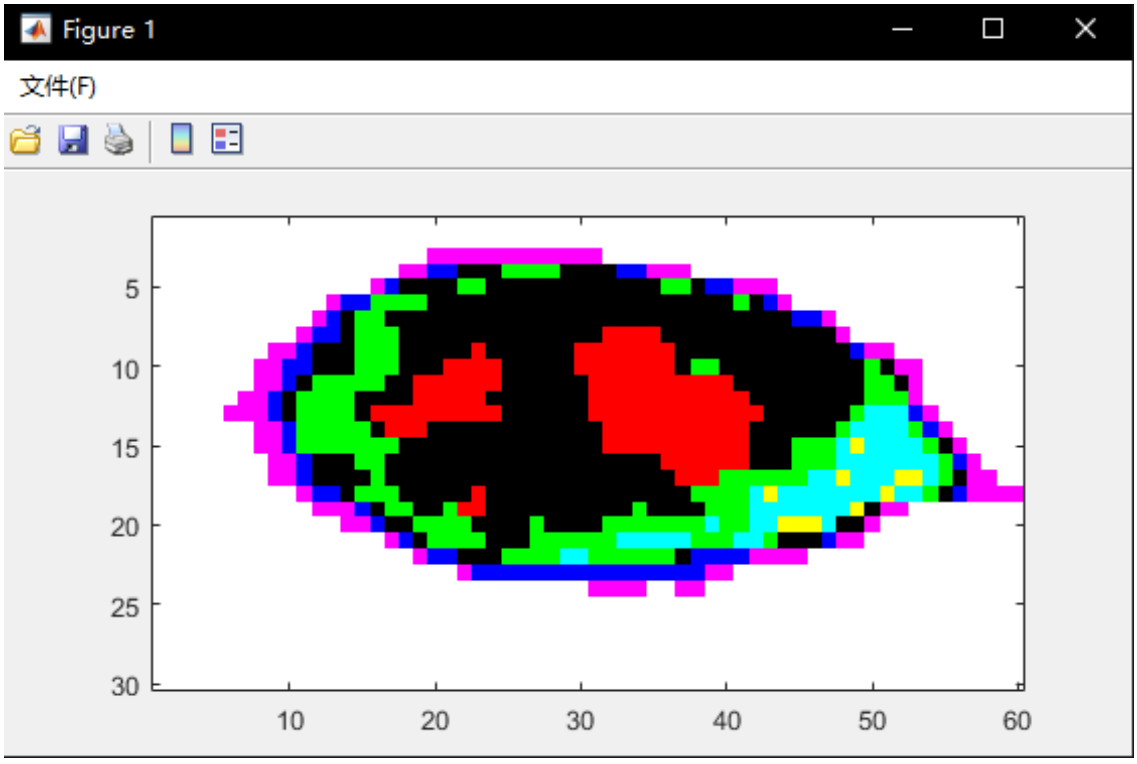
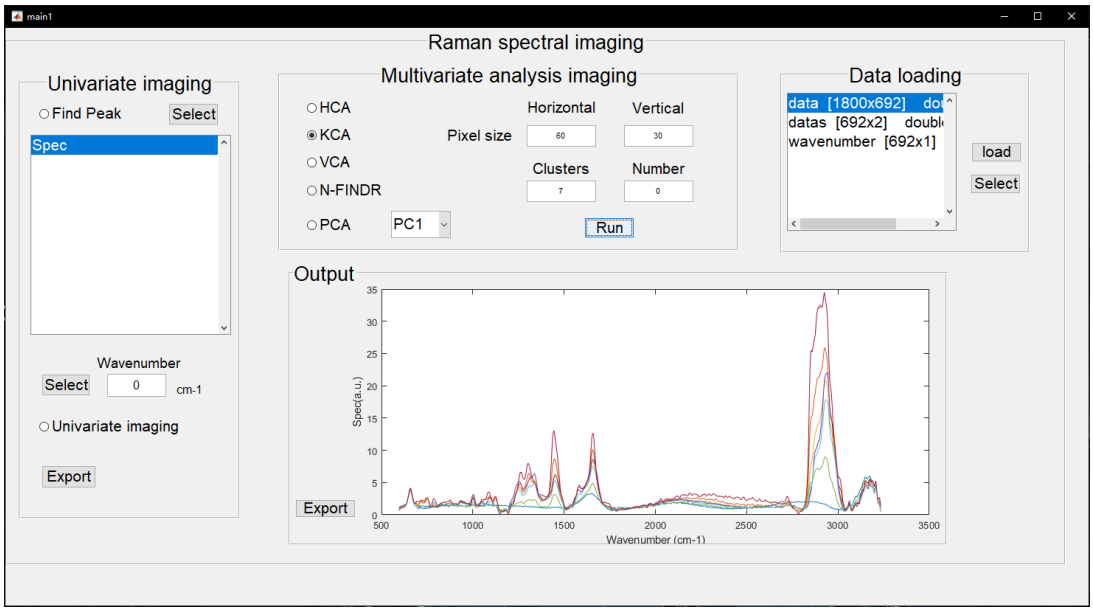


Figure 5. The HCA input interface and output results include false color images and cluster mean spectra.



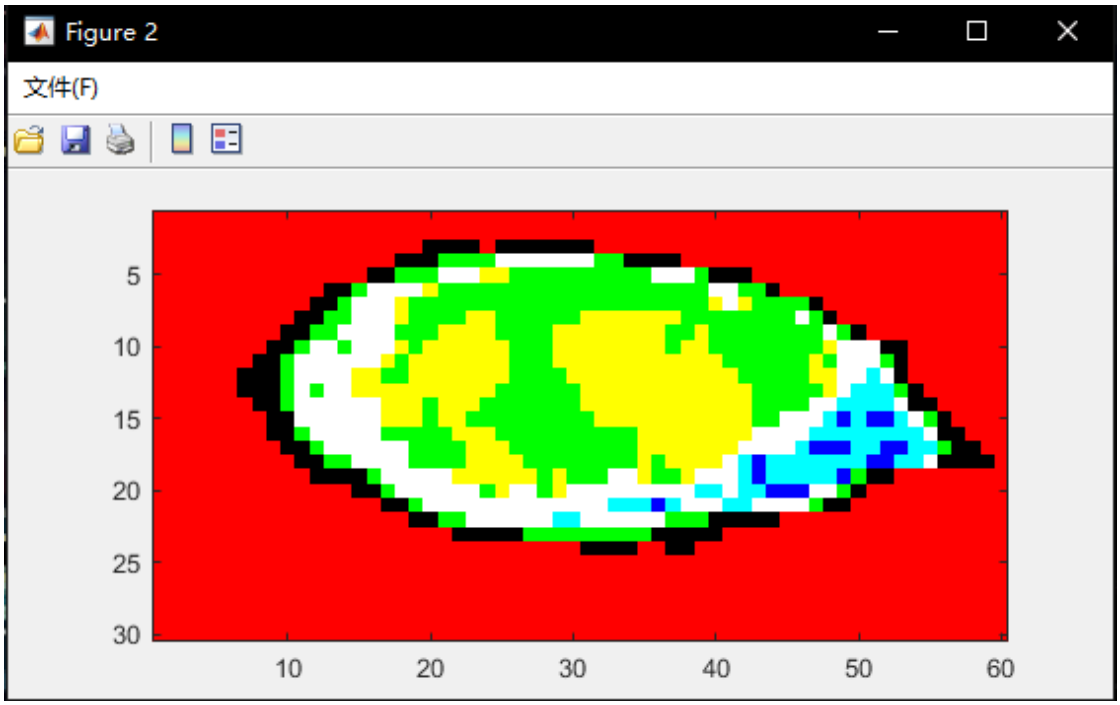


Figure 6. The KCA input interface and output results include false color images and cluster mean spectra.

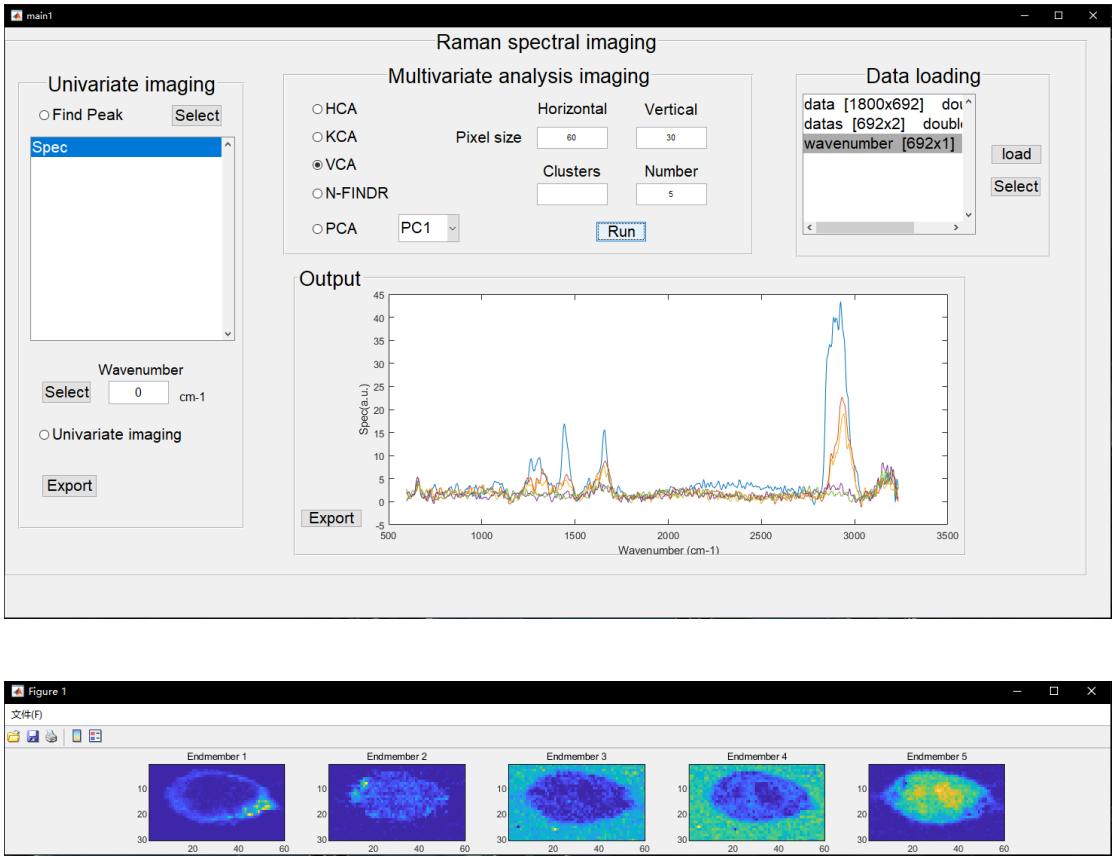


Figure 7. VCA input interface and output results include endmember spectra and endmember abundance images.

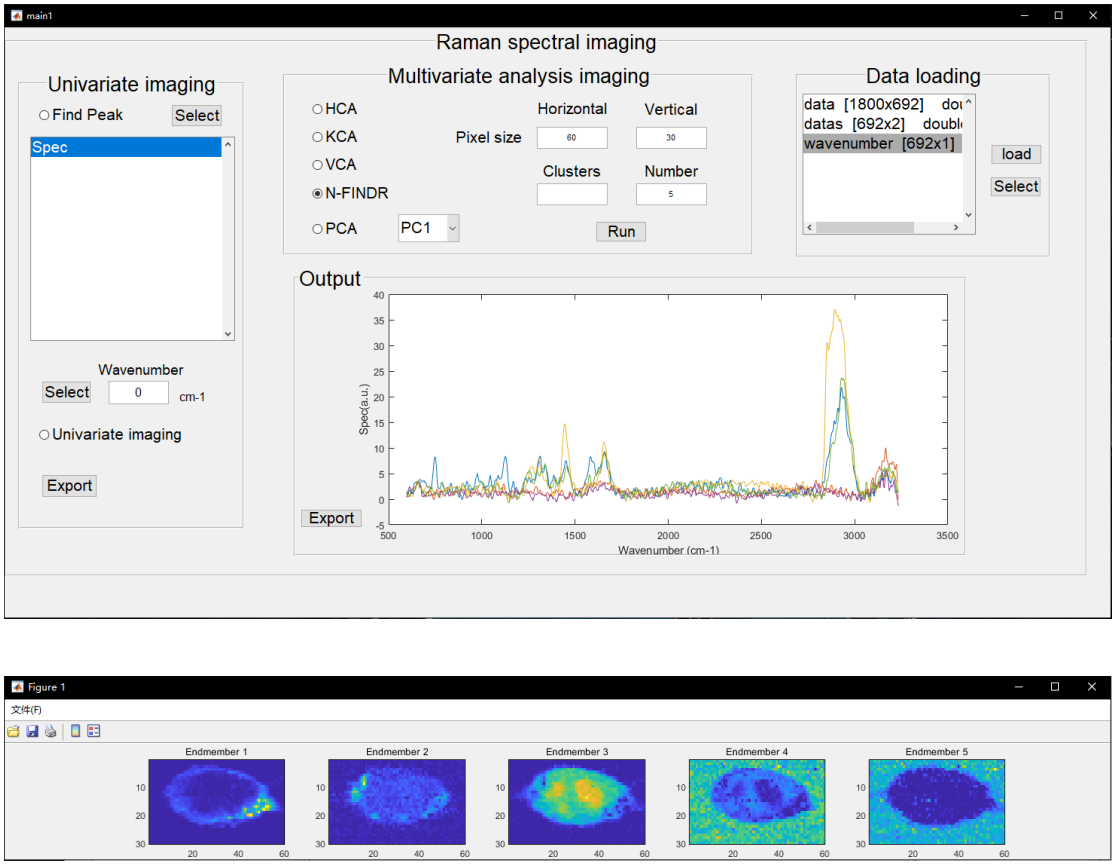
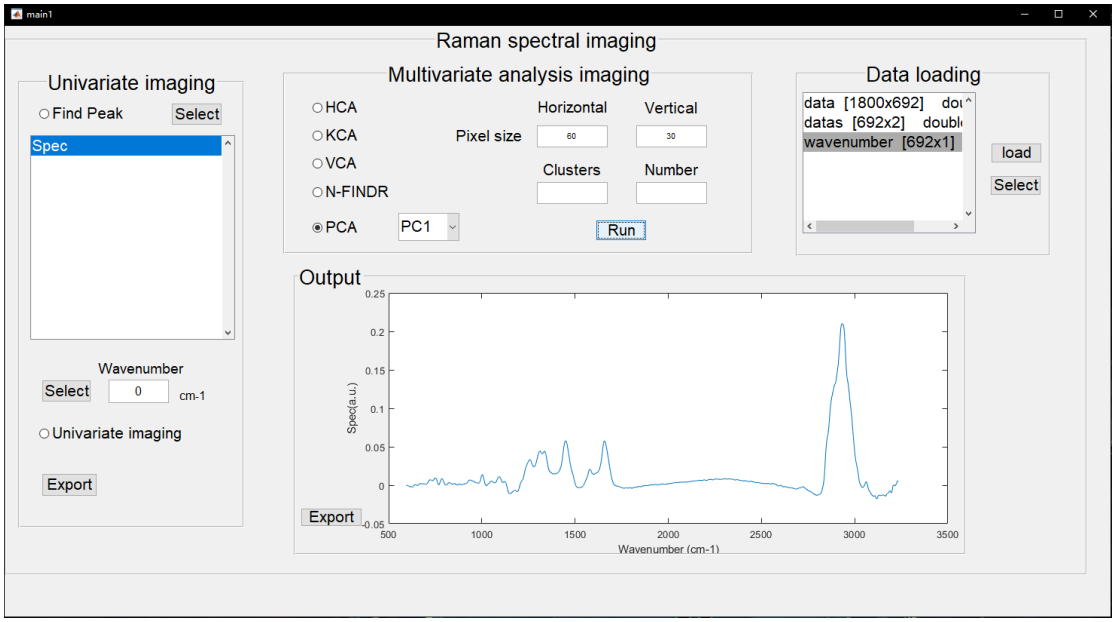


Figure 8. N-FINDR input interface and output results include endmember spectra and endmember abundance images.



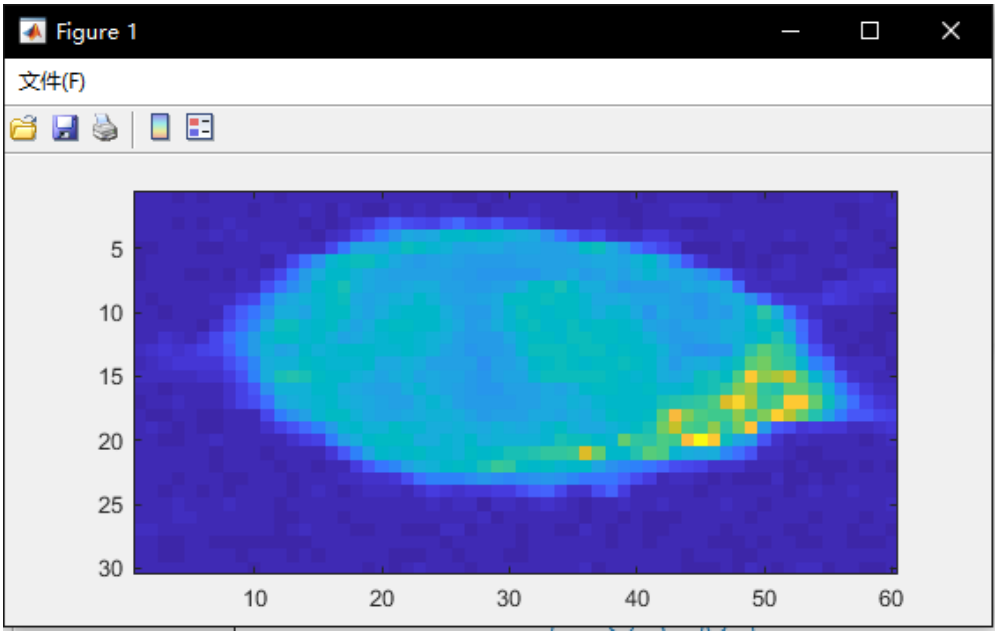


Figure 9. PCA results include PC1 loading and PC1 loading contribution imaging.

3. References

1. D Song, Y Chen, J Li,H Wang, T Ning, and S. Wang "A graphical user interface (NWUSA) for Raman spectral processing, analysis and feature recognition." J Journal of Biophotonics, e202000456 (2021).
2. J. Li, J. Qin, X. Zhang, R. Wang, Z. Liang, Q. He, Z. Wang, K. Wang, and S. Wang, "Label-free Raman imaging of live osteosarcoma cells with multivariate analysis," Appl Microbiol Biotechnol **103**, 6759-6769 (2019).
3. J. Li, Z. Liang, S. Wang, Z. Wang, X. Zhang, X. Hu, K. Wang, Q. He, and J. Bai, "Study on the pathological and biomedical characteristics of spinal cord injury by confocal Raman microspectral imaging," Spectrochim Acta A Mol Biomol Spectrosc **210**, 148-158 (2019).
4. J. Li, R. Wang, J. Qin, H. Zeng, K. Wang, Q. He, D. Wang, and S. Wang, "Confocal Raman Spectral Imaging Study of DAPT, a gamma-secretase Inhibitor, Induced Physiological and Biochemical Reponses in Osteosarcoma Cells," Int J Med Sci **17**, 577-590 (2020).
5. D. Song, T. Chen, S. Wang, S. Chen, H. Li, F. Yu, J. Zhang, and Z. Zhang, "Study on the biochemical mechanisms of the micro-wave ablation treatment of lung cancer by ex vivo confocal Raman microspectral imaging," Analyst **145**, 626-635 (2020).
6. A.-K. Kniggendorf, T. W. Gaul, and M. Meinhardt-Wollweber, "Hierarchical Cluster Analysis (HCA) of Microorganisms: An Assessment of Algorithms for Resonance Raman Spectra," Applied Spectroscopy **65**, 165-173 (2011).
7. H.-F. Köhn and L. J. Hubert, "Hierarchical Cluster Analysis," in *Wiley StatsRef: Statistics Reference Online* (2015), pp. 1-13.
8. J. Li, J. Li, J. Qin, H. Zeng, K. Wang, D. Wang, and S. Wang, "Confocal Raman microspectroscopic analysis on the time-dependent impact of DAPT, a gamma-secretase inhibitor, to osteosarcoma cells," Spectrochim Acta A Mol Biomol Spectrosc **239**, 118372 (2020).
9. J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," IEEE Transactions on Geoscience and Remote Sensing **43**, 898-910 (2005).
10. M. E. Winter, S. S. Shen, and P. E. Lewis, "A proof of the N-FINDR algorithm for the automated detection of endmembers in a hyperspectral image," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X.* (2004).