

January 01, 2007

Component extraction from CRM skin images

Sol M. Cruz Rivera

University of Puerto Rico - Mayaguez

Vidya Manian

University of Puerto Rico - Mayaguez

Charles A. DiMarzio

Northeastern University

Recommended Citation

Cruz Rivera, Sol M.; Manian, Vidya; and DiMarzio, Charles A., "Component extraction from CRM skin images" (2007). *Research Thrust R2 Presentations*. Paper 12. <http://hdl.handle.net/2047/d10008974>

This work is available open access, hosted by Northeastern University.



Component Extraction from Hyperspectral CRM Images

Sol M. Cruz Rivera, MS Graduate Student UPRM, sol.cruz@ece.uprm.edu
 Dr. Vidya Manian, Assistant Professor UPRM, manian@ece.uprm.edu
 Dr. Charles DiMarzio, Professor NU, DiMarzio@ece.neu.edu
 Laboratory for Applied Remote Sensing and Image Processing
 University of Puerto Rico at Mayagüez, P. O. Box 9048, Mayagüez, Puerto Rico 00681-9048



Abstract

Confocal Reflectance Microscope images the skin in 4 dimensions. The third and fourth dimensions being the wavelength and the depth, respectively. We compare traditional methods with new techniques for component separation. The components from the skin images are cell/nuclei, mitochondria, melanin, organelles and other minor components. The spectral unmixing method N-finder is used to extract end members. The results of this method will be compared with ICA methods that will be developed for the 4-dimensional data. The components and their concentrations (abundances) will be output from these methods.

Independent Component Analysis

Independent Component Analysis (ICA) is a statistical and computational technique for finding the underlying factors or components from multidimensional data. The observed data \bar{y} are expressed as a linear transformation of latent variables \bar{x} that are non gaussian and mutually independent:

$$\begin{bmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{bmatrix} = A \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \Rightarrow \bar{y} = A\bar{x}$$

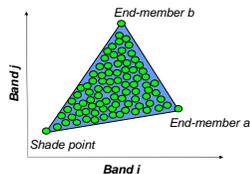
where $[y_1, y_2, \dots, y_n]^T$ is the vector of the observed random variables, $[x_1, x_2, \dots, x_n]^T$ is the vector of the independent latent variables or the *independent components*, and A an unknown constant matrix, also known as the *mixing matrix*. The goal of the ICA consists of estimating both the A matrix and \bar{x} when we only observe a given sequence of independent observations y_1, y_2, \dots, y_n .

N-FINDR

The *N-FINDR* algorithm is an automated technique for finding the *end-members* in an image. The resulting images shows the abundances of the corresponding end-member for that pixel. *N-FINDR* uses the fact that in general, the spectra of a particular pixel in an image is assumed to be a linear combination of the end-member spectra.

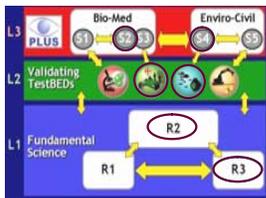
$$p_{ij} = \sum_k e_{ik} c_{kj} + \varepsilon$$

where p_{ij} is the i -th band of the j -th pixel, e_{ik} is the i -th band of the k -th end-member, c_{kj} is the mixing proportions for the j -th pixel from the k -th end-member, assumed to sum one, and ε is the Gaussian random error, assumed to be small. The vertices of a simplex, that is the simplest geometric shape that can enclose a space of a given dimension, are the end-member spectra. Hence, finding the pure pixels in an image (end-members), is nothing but, finding the points in the data that represent the vertices of the simplex containing the data.



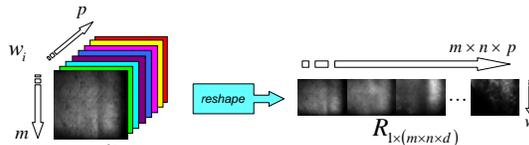
Technology Transfer Opportunities

This work will be useful for CenSSIS Researchers and Students from R2C, S1, S3, and S4 who make use of multi and hyperspectral images and will result in technology transfer to the industry in the form of tools and methodologies for spectral image processing.
 This work is of interest to: ITT, NGA, Lockheed Martin, ARMY, NASA

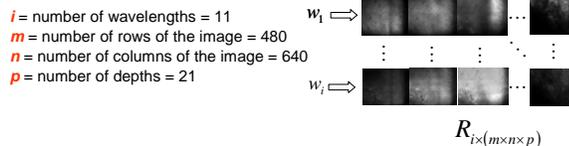


CRM Data

The CRM images has four dimensions (4-D). The first two dimensions formed the 480x640 images. The third and fourth dimensions are the depth and the wavelength, respectively. For a specific wavelength the 4-D data are arranged in the following way:



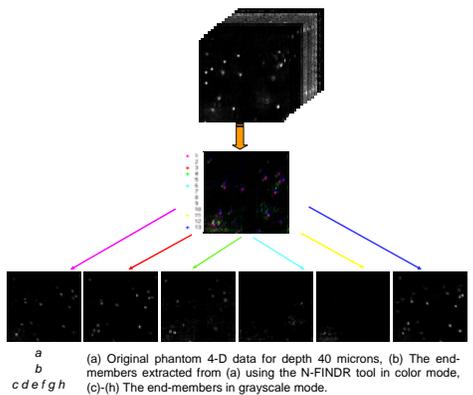
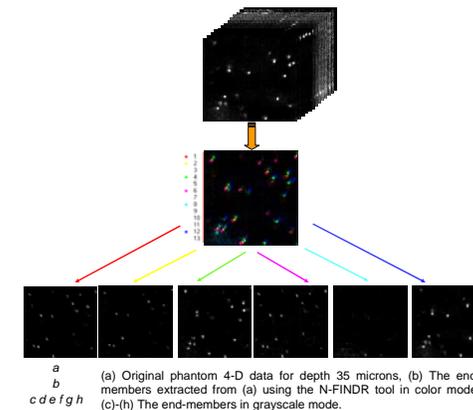
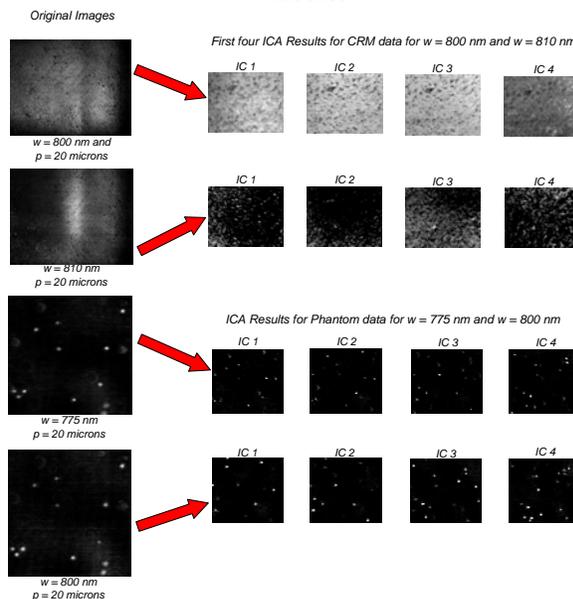
For all the wavelengths, the 4-D data are arranged as follows:



Phantom Data

Like CRM images, phantom images has four dimensions. They are arranged in the same manner as the CRM data, but the size of phantom images is different. In this case the number of wavelengths is 13, the number of row and columns of the image is 100.

Results



Conclusions and Future Work

1. Statistical techniques have been applied to hyperspectral microscope images for extracting components (endmembers).
2. Both linear and non-linear unmixing methods have been presented.
3. The number of endmembers extracted by these methods have to be verified with ground truth.
4. Future work will include implementing spatial processing in extents of 3x3, 5x5 windows to extract features that will be used to train a classifier.
5. Semi-supervised methods such as semi-SVM that are more suitable for this type of data that do not have ground truth, will be used for classification. These methods can work in high dimensional space and require very few training samples which can be extracted from pure pixel extraction methods such as N-FINDR or PPI (Pixel Purity Index) routines.

References

1. R. Duda, P. Hart and D. Storks, "Pattern Classification", Second Edition, pp. 1543-1551.
2. José M. P. Nascimento and José M. B. Dias, "Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?", *IEEE Trans. Geosci. Remote Sensing*, 2004
3. N-FINDR 3.0 Documentation

Acknowledgments

- Partially supported by the NSF Engineering Research Centers Program under grant ECC-9986821 and DoD under contract W911NF-06-1-0008.
- ICA Algorithm provided by Laboratory of Computer and Information Science of the Helsinki University of Technology, Finland