© 2016 IJSRSET | Volume 2 | Issue 5 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Band Selection in Hyperspectral Images using Independent Component Analysis

Vinay Chandragiri

Department of Computer Science and Engineering, IIT Guwahati, Guwahati, Assam, India

ABSTRACT

In Hyperspectral Imaging is the new modality in medical applications which is probably being used in Remote sensing applications. The image is generally of high dimension with spectral bands for a pixel. The main idea of the segmentation is to identify cancerous cells among the tissues. Here I am trying to address the problem of classifying cells by gland segmentation for cancer detection in the given colon tissue. The dimensionality problem has been tackled by Band Selection based on Independent Component Analysis.

Keywords : Hyperspectral Imaging, ICA, Clustering

I. INTRODUCTION

Hyperspectral Imaging is a new and emerging technology in the field of Medical Imaging. It combines both digital imaging and spectroscopy. Instead of RGB values for each pixel, every pixel in the image contains a continuous electromagnetic spectrum and is very useful in characterizing the objects with immense accuracy and precision. Basically HSI is a stack of images (fig 1). This new technology of Imaging is developed by NASA for Earth Imaging (Remote Sensing) and Space Observation. Hyperspectral Images are produced by instruments called Imaging Spectrometers. Medical Image analysis helps contributing to the betterment of lives of people by prevention and cure of deadly diseases. These Image Processing and Machine Learning techniques will be the key behind the success.

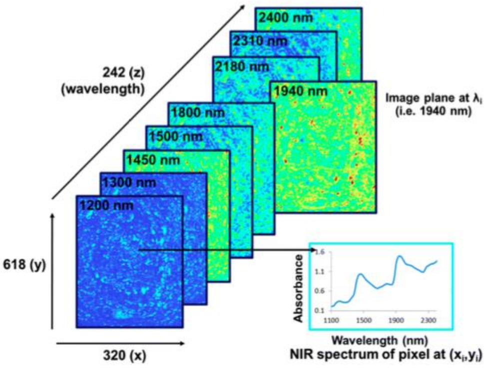

Fig 1

II. METHODS AND MATERIAL

1. Hyperspectral Images

Based Exploiting the property of rich information in Hyper- spectral Images (HSI), is a consistent development in the field of Medical Imaging and Health Care for various types of classification. It is found that this technology has important application in the field of Cancer detection. In medical image analysis of hyperspectral images provide a greater accuracy in determining the affected regions and reasons behind them.

HSI is capable of capturing both spectral information as well as spatial information in one shot. Hyperspectral Imaging provides a unique spectral signature, which can be used by processing techniques and discriminate materials.

The colon and rectum comprise the Large Intestine. Colon cancer is the third most common type of cancer after lung cancer and breast cancer. This is a disease in which normal cells in the lining of colon or rectum begin to change, grow without control and no longer die. It may cause bleeding which is painless and is not visible to eye.

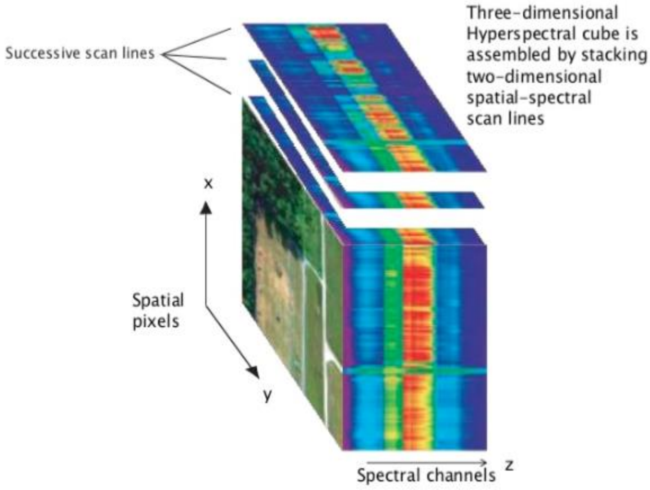

Fig 2

HSI helps in segmentation of glands and as well as classification of cells based on the properties of reflectance of different materials involved.

In the epithelial lining of large intestine, intestinal glands are found. These glands contain different types of cells in that epithelium lining such as goblet cells, enterocytes etc. Stroma is a part of tissue that contributes to the connection and structures.

Segmentation of these and grouping the similar types of cells together will help determine whether the tissue is benign or malignant if it is cancerous.

Statistical methods were used for classification (clustering) which are unsupervised. Supervised methods can yield better results with manual segmentation and using classifiers such as SVM's or Artificial Neural Networks.

Few recent advancements determine the extensive use of this technique in medical image analysis as the results are accurate. Prostate Cancer detection, Breast Cancer Detection and Skin Cancer Detection prevail among them.

2. Working with Hyperspectral images

The term "Curse of Dimensionality" refers to the difficulties in processing the high dimensional data. In a single Hyperspectral Image (say of dimension A x B x C) , where A x B represents number of pixels and C represents the number of spectral. Considering the medical data used, A and B were in order of 500 - 600 and C in the order 220 - 240. So each pixel in that particular image is a C-dimensional vector and there are more than 2,00,00 pixels in one image.
This property of an HSI makes it tough to process the raw data for analysis for various purposes. This is the underlying reason leading to "dimensionality reduction" of Hyper- Spectral Images a very intensive research topic to work upon.

2.1 What are features?

Features are attributes from raw data such that there value make an instance. These determinant values determine about the belonging of an instance to a particular class. We can classify features based on Relevance, Irrelevance and Redundancy.

As the dimensionality problem has already been discussed earlier, to process the raw data, we can employ two different paradigms which are helpful in not using the redundant data for our processing. Redundancy exists when one feature takes place the role of another feature.

2.2 Which features are better and why?

Feature selection is a process by which we choose certain features (a smaller subset) from raw data using certain evaluation criteria. Different statistical methods such as clustering, minimizing Mutual Information, computing KL divergence, minimizing Entropy are explored. The main objectives of this process is to remove redundant data and irrelevant data.

Feature reduction (also termed as Feature extraction) is a process of mapping the High dimensional data to lower dimension such that the redundant and irrelevant data is minimized. Principal component analysis would be perfect example for this technique.

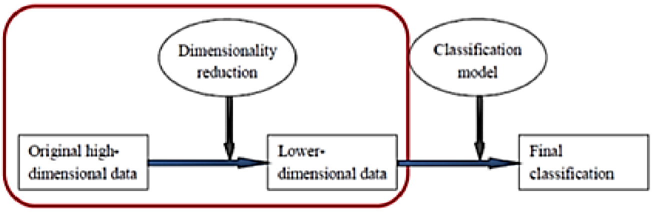

Fig 3

Optimally choosing a subset of features would be a better option instead of extracting new subsets of features as in the extraction will assure of loss of data and the process is ir- reversible. After doing some literature review about techniques that involve dimensionality reduction I turned myself over Feature selection.

2.3 Other Techniques which have been explored

Principal Component Analysis, Independent Component Analysis and Multiple Discriminant Analysis are few of them. I have used ICA as a means to select most relevant bands from raw hyperspectral data instead of PCA and MDA.

3. Independent Component Analysis

3.1 When we do ICA ?

- Raw data appears to be noisy
- When data is non Gaussian i.e cannot be grouped via central limit theorem
- The sensor involved collects several source signals simultaneously

3.2 Why ICA and Why not PCA?

ICA minimizes the higher order statistics such or Kurtosis which will essentially minimize the mutual information in the output.

PCA minimizes the covariance of the raw data. PCA is most likely not suitable with Hyperspectral Images and these are highly correlated i.e the nearby pixels are highly correlated.

The reason behind this is simple. With the uncorrelated variables in the data, the contributions by these in the lower dimension i.e to components will be almost equal in most of the cases. This does not happen with the variables which are correlated. We might skip the variables with most information with the process

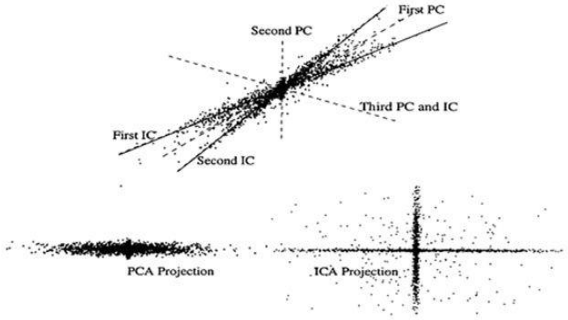

Fig 4

4.3 ICA Algorithm

Given a measurement X, assuming it as a linear combination of independent sources S and the mixing matrix A are to be determined such that **X = AS.**
The independent sources are determined from the equation

S = (inv)A * X i.e S = WX

where W is termed as the mixing matrix.

- Input: X of (N x M) represents N-Dimensional Sample. We can also fix the number of components we want to(say C).
- Output: W of (N x N) the Un-mixing matrix.

for p **in** 1 to C:
  $w_p \leftarrow$ *Random vector of length N*
  while w_p changes

$$\mathbf{w_p} \leftarrow \frac{1}{M}\mathbf{X}g(\mathbf{w_p}^T\mathbf{X}) - \frac{1}{M}g'(\mathbf{w_p}^T\mathbf{X})\mathbf{1}\mathbf{w_p}$$

$$\mathbf{w_p} \leftarrow \mathbf{w_p} - \sum_{j=1}^{p-1}\mathbf{w_p}^T\mathbf{w_j}\mathbf{w_j}$$

$$\mathbf{w_p} \leftarrow \frac{\mathbf{w_p}}{\|\mathbf{w_p}\|}$$

Output: $\mathbf{W} = \begin{bmatrix} \mathbf{w_1} \\ \vdots \\ \mathbf{w_C} \end{bmatrix}$

Output: $\mathbf{S} = \mathbf{WX}$

The key lies in determining the matrix W and a measure of non-Gaussianity.

3.4 Disadvantages

- We cannot determine the order of dominant components such as in PCA
- We cannot determine the variances of Independent components

4. Band Selection using ICA

After obtaining the W, the absolute weight coefficients corresponding to each to each row are sorted and a band sequence is obtained because the weight matrix determines how a particular band contributes to each material in the hyperspectral image.

From the sequence, the bands with greater absolute weight coefficients will contribute more to the Independent transformations than other bands. This means that the bands with higher weights will have more spectral information than the other bands.

Thus from ICA, the bands are obtained and only those bands are selected from the original image and proceeded further as we ignored the irrelevant and redundant information is neglected.

These selected bands are called Independent bands and thus the unsupervised classification i.e clustering can be carried over with these selected bands. The new spectral image is constructed with these independent bands.

III. RESULTS AND DISCUSSION

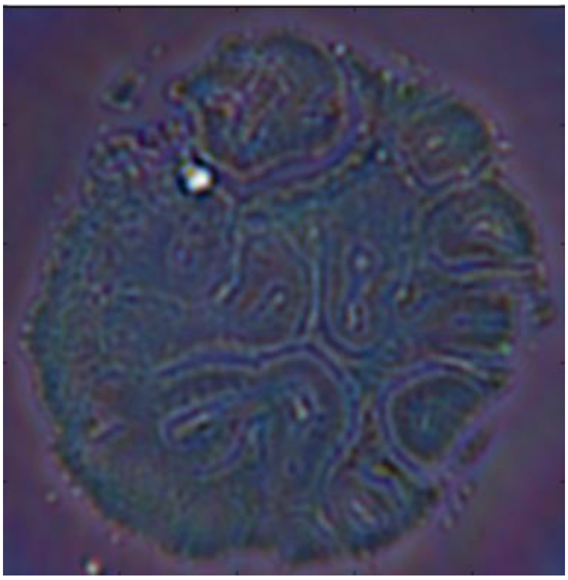

Figure 4: Original Image

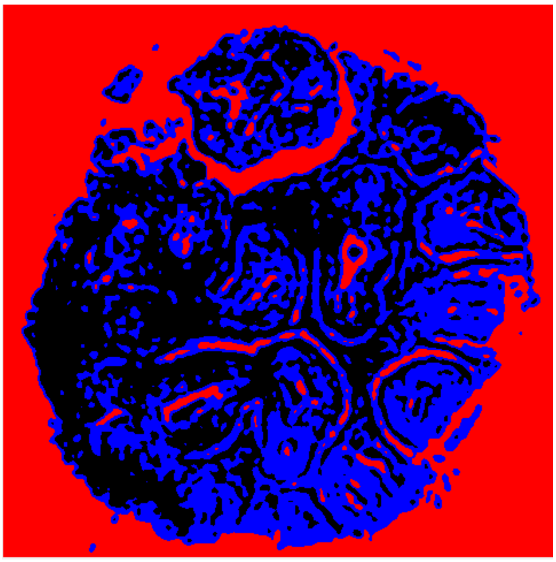

Figure 5 : Clustered Image (3 clusters) with 2 Spectral Bands 192 & 194 from all 226

IV. REFERENCES

[1] Independent Component Analysis: Algorithms and Applications Aapo Hyvrinen and Erkki Oja Neural Networks Research Centre Helsinki University of Technology, Finland

[2] Spatial Mutual Information Based Hyperspectral Band Selection for Classification

[3] A Comparative Analysis of Dimension Reduction Algorithms on Hyperspectral Data Kate Burgers, Yohannes Fessehatsion, Jia Yin Seo, Sheida Rahmani