

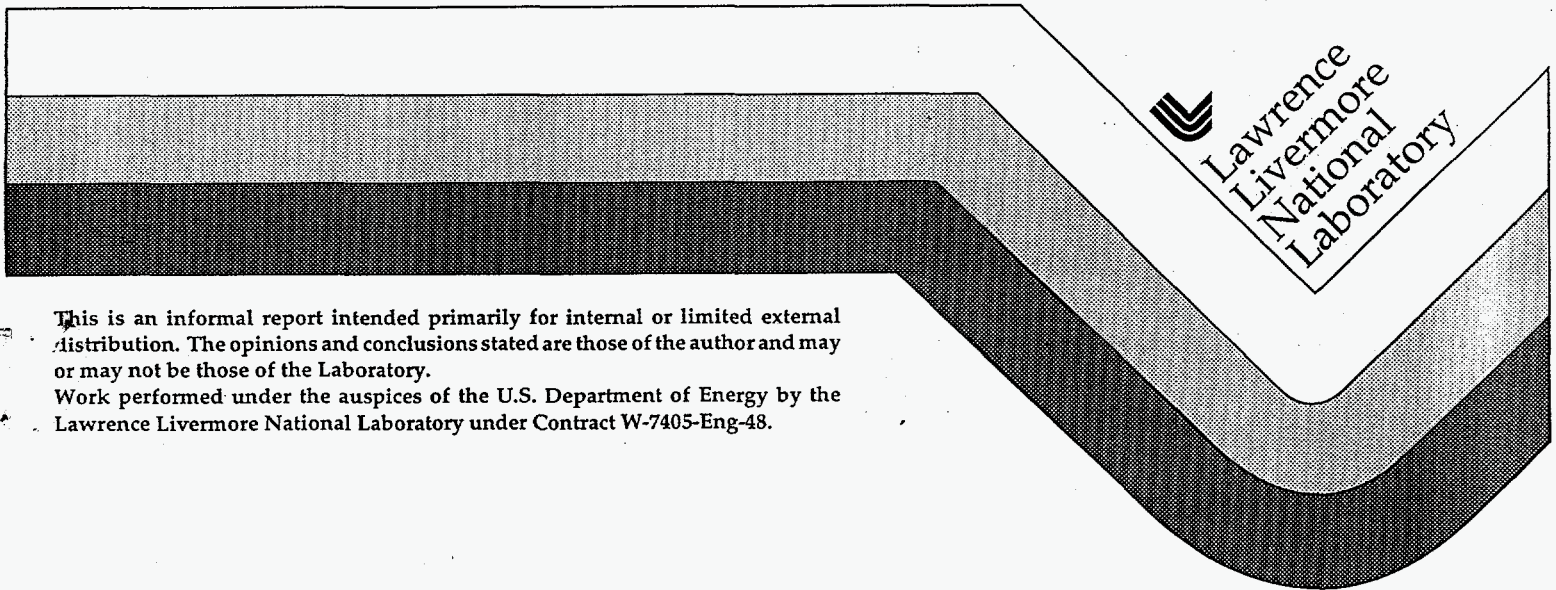
Analysis of Hyper-Spectral Data derived from an Imaging Fourier Transform: A Statistical Perspective

Authors

Sailes K. Sengupta, Gregory A. Clark, David J. Fields

RECEIVED
MAY 05 1996
OSTI

January 10, 1996



This is an informal report intended primarily for internal or limited external distribution. The opinions and conclusions stated are those of the author and may or may not be those of the Laboratory.

Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This report has been reproduced
directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161

Analysis of hyper-spectral data derived from an
imaging fourier transform: A statistical perspective

Sailes K. Sengupta ¹

Gregory A. Clark ¹

David J. Fields ²

January 10, 1996

¹ EE/ERD, LLNL/UC

² V-Div., LLNL/UC

Abstract

Fourier transform spectrometers (FTS) using optical sensors are increasingly being used in various branches of science. Typically, a FTS generates a three-dimensional data cube with two spatial dimensions and one frequency/wavelength dimension. The number of frequency dimensions in such data cubes is generally very large, often in the hundreds, making data analytical procedures extremely complex. In the present report, the problem is viewed from a statistical perspective. A set of procedures based on the high degree of inter-channel correlation structure often present in such hyper-spectral data, has been identified and applied to an example data set of dimension $100 \times 128 \times 128$ comprising 128 spectral bands. It is shown that in this case, the special eigen-structure of the correlation matrix has allowed us to extract just a few linear combinations of the channels (the significant principal vectors) that effectively contain almost all of the spectral information contained in the data set analyzed. This in turn, enables us to segment the 'objects' in the given spatial frame using, in a parsimonious yet highly effective way, most of the information contained in the data set.

1. Introduction

Modern Fourier transform spectrometers (FTS) generate hyper-spectral data that can be described as a data 'cube' having two spatial dimensions and a third interferogram dimension with a large number of spectral bands, usually numbering in the hundreds. Lawrence Livermore National Laboratory is in the process of developing such an FTS for chemical effluent monitoring [1]. The analysis and interpretation of such data present a formidable task to the data analyst. The foremost task in such an analysis is an effective reduction of data with a minimum loss of information and maximum ease in the extraction of relevant information from the reduced dataset. This kind of task has been accomplished in the past although in other contexts, using a variety of methods. Among such methods are:

the method based on eigen-structure of the covariance (correlation) matrix that goes by the name of principal component analysis (PCA) [2] or empirical orthogonal function analysis (EOF) [3] or Karhunen-Loeve (K-L) [4] expansion, the method based on projection-pursuit [5] that is a projection algorithm using statistics of high-dimensional data to find optimal projections on a significantly lower dimensional subspace in a repetitive fashion, and a class of methods based on neural networks [6] that are nonlinear and nonparametric, exploiting the parallel distributed processing capability inherent in these algorithms.

The present report and others that will be forthcoming will explore the development and enhancements of such methods in the context of hyper-spectral data analysis. Only the first method will be considered in this report. The methods indicated in the other two will be considered in later studies.

Significant data reduction can be accomplished by the extraction of principal components (PC) when the data are highly correlated. In such cases, the first few principal components can be effectively utilized for the purpose of segmentation and/or object identification purposes using some well known methods. Reduction of dimensionality resulting from PC extraction make the implementability of such algorithms relatively easy. The organization of the

report is as follows. In section 2, we describe the method of principal components and some of its advantages and shortcomings in hyper-spectral data analysis. We discuss in section 3 how the extracted PCs can be exploited for the purpose of segmentation of hyper-spectral imagery. Several methods of segmentation of multi-spectral imagery exist in the literature. We briefly examine a few of them and present in detail one of them for application to a hyper-spectral data set in section 4. There, we also discuss a preprocessing technique that proved useful in the particular data set we have analyzed. Section 5 concludes with indications for future work.

2. Principal component analysis.

Principal vectors determine an orthogonal representation scheme for hyper-spectral data in which the coordinate axes are ordered based on the variances of the projections of the data set on the axes (the principal components or PCs for short). The total variance of these projected data sets equal to the sum of the variances in each band. Furthermore these PCs are uncorrelated. This in effect provides us with a framework with which to study the problem of summarizing the variation in the hyper-spectral data cube to the study of a few chosen linear combinations of the spectral bands.

In an $M \times N \times B$ data set, where B is the number of spectral bands, there are $n = MN$ multivariate samples of B components each. We can compute the $B \times B$ sample covariance matrix C and analyze its eigenstructure for the principal component analysis. If the dynamic ranges of variation in different spectral bands are significantly different, the correlation matrix instead of the covariance matrix is often recommended for eigenstructure analysis in order to avoid the overwhelming contribution of the channels with large dynamic ranges on the dominant eigenvectors. This has the same effect as doing the covariance based PCA on the standardized variable. Due to the omnipresent noise in the measurements, it can be assumed that the covariance (correlation) matrix is positive definite. Hence, the eigenvalues are positive, the eigenvectors span the original space, and are orthogonal. Thus they provide a new coordinate system for the data. The k th PC or the k th coordinate in the new system is given by

the dot product of the original data and the k th principal vector. In practice, the principal vectors are computed first with or without the data standardization indicated earlier. The standardization, which is essentially a scaling operation, is normally called for when the data in different spectral bands appear to be heterogeneous in the sense of having different dynamic ranges of values. The standardization can be done in several ways depending on the nature of the data set. One possible way is to subtract the mean (over all bands) image from the images in each band and divide the result by the standard deviation over all bands on a pixel by pixel basis. This method of standardization will be termed *inter-spectral standardization*. This process will tend to enhance inter-spectral variability on a pixel by pixel basis.

On the other hand, one can apply the standardization procedure within the same band. Here, one would, for a given band, subtract the mean radiance in the band and divide the result by the standard deviation of the radiance in the same band. We shall refer to this as *intra-spectral standardization*. This method of standardization will tend to enhance the pixel to pixel contrast and especially the outliers in a given band. When applied within a PCA context, this is equivalent to doing a PCA based on the correlation matrix of the spectral bands. Also, this has the effect of forcing the dynamic range in each spectral band to stay within approximately the same interval. In the present work, we have indicated with examples that both methods of standardization may prove to be useful. It should however be noted that without some kind of scaling, the principal components analysis of a heterogeneous data set tend to pick out the channel noise variability in the first few 'significant' principal components. Hence it is desirable to remove the channel noise variability first before undertaking a principal component analysis. The issue of channel noise variability has not been addressed in this report.

Various approaches have been suggested for the retention of significant principal components. These can be based on a set of statistical criteria indicating the degree of 'lack of fit' and is determined by the root mean square scaled residuals. In another approach, one may employ statistical tests of significance based on the computed eigenvalues. This

however requires additional assumptions on the underlying distributions. In the present case we have adopted the practice of truncating at a level where the percentage of variance of the principal component under consideration falls below a predetermined low level.

While principal component representation followed by truncation often results in significant data reduction and consequent reduction in computational complexity in the subsequent analyses, it has some known drawbacks. In instances where the spectral bands are relatively uncorrelated, significant data reduction will not, in general, be possible. Also, the images represented by the principal components may not necessarily provide any direct physical interpretation. This makes any meaningful segmentation of the imagery somewhat difficult. In such cases, extensive experimentation would be necessary to establish empirically any correspondence that may exist between the significant PCs and the underlying phenomenon of physical significance.

3. Segmentation of multi-spectral imagery

We will briefly indicate three classes of methods employed for the purpose of segmentation of multi-band co-registered imagery. First, the method of *region merging* [7] in which one begins with an image array $I(x,y)$ containing $M \times N$ elements and end with a set of r regions R_k , $k=1,2,\dots,r$ where r is less than MN . Thus the original image is viewed as the disjoint union of regions where the pixels within a region are 'similar' and pixels in two different regions are dissimilar. In the context of multi-spectral data the similarity must be in all components simultaneously. In addition to similarity, the notion of *proximity* also must play a role in deciding whether one should merge two regions to make into one. In a different approach, one can use the notion of *region splitting* [8], where one starts with a single region, namely the whole image and splits it to create new entities or regions based on the nonuniformity of the region being split and proceed in turn with each of the split regions until a stage is arrived when the region is no longer heterogeneous. The measure of uniformity that is used as a criterion in this method is allowed to depend on multiple

bands of the imagery. Other methods exist where one uses a hybrid approach. Yet another method is through *clustering*. Here, one uses spectral, or other features of a pixel to put it in one of several groups of 'similar' pixels. Ordinarily, spatial proximity is not considered in clustering. There are several measures of similarity (and dissimilarity) in use. The most frequently used similarity (dissimilarity) measure for numerical features is the correlation (Euclidean or other 'distance') measure. There exist two general types of nonhierarchical clustering methods in statistical literature, depending on whether the analyst starts with, or without the knowledge of how many different clusters exist in the given dataset. In the first case, *K-Means algorithm* [9] is the method of choice while in the latter, the *Isodata* [10] algorithm is employed. In the following paragraph, we describe the former.

K-Means Algorithm

The algorithm computes K Euclidean distance based clusters starting from a set of initial estimates of the K cluster means (centroids). Let p be the number of variables (features) to be used for clustering. Let $X_{n \times p}$ be the data matrix with n pixels to be grouped into K clusters based on a feature vector of p components. The p components of the feature vector may be the spectral radiances in p distinct bands of the imagery. Alternatively, they may be derived features such as the first p principal components in the hyperspectral data. More generally, the derived features may even be computed statistics based on a neighborhood of suitable size around the pixel. The basic idea in the algorithm is to find a grouping of the pixels into K groups so as to *minimize the total within-cluster-sum-of-squares* S . The total sums-of-squares within each cluster is computed as the sum of the squared deviation over all values of each variable (feature) from the means of each variable within the cluster. Specifically, the criterion S to be minimized is given by

$$S = \sum_i \sum_j \sum_m (X_{\alpha(i,m)} - \text{Avg}(X_{i,j}))^2,$$

Where, $\text{Avg}(X_{i,j})$ is the average of the observations for the variable j in group i and $\alpha(i,m)$ is the row index of the m th observation in the i th cluster in the matrix X . The algorithm

sequentially processes each observation and reassigns it to a different group if in so doing there is a decrease in S . For a detailed description of the algorithm, see [11] for example. The following summarizes the three basic steps in the algorithm[12].

1. Partition all items into K preliminary groups or start with K initial centroids. This may be a randomly chosen set of K points. Alternatively, one may partition the pixels in the imagery into a set of K rectangular cells and choose their respective centroids as the starting point. This allows proximity to be one of the determining factors in the clustering, albeit in an indirect fashion.
2. For each item in the list, assign it to the cluster whose mean vector is nearest (with respect to some metric) to it. Typically, the metric is chosen as the Euclidean distance with standardized or unstandardized observations. Update the centroids of the cluster(s) that are affected due to this assignment (or reassignment, if any, during subsequent steps).
3. Repeat step 2 until no more reassignment takes place through a complete pass of the data.

While the K-Means algorithm provides an efficient and often effective method of segmentation via clustering, its primary disadvantage is that the number K and the corresponding seeds have to be provided in advance. There are strong arguments against this[12]. Some of these are listed below:

- (a) The algorithm is sensitive to the existence of an outlier possibly producing at least one cluster with highly dissimilar items.
- (b) In a practical application, when one is looking for exactly K objects in a scene, the absence of one or more of the objects will produce a nonsensical clustering.
- (c) If two or more seeds happen to fall in the same 'natural' cluster, their resulting clusters will be differentiated poorly.

Most of these objections can be tackled if we try the algorithm with several K's followed each time by statistical tests on cluster validity in an iterative fashion. An extensive battery of such tests is available[13].

Remark: Most of the classical statistical algorithms in Data Reduction and Clustering have their counterparts in some of the recent algorithms based on fuzzy set theory and/or artificial neural networks. These will be dealt with in our forthcoming reports.

4. Data and results

Science Applications International Corporation (SAIC) participated in a field experiment on December 8-9, 1993 at the Nevada test site on a series of plots operated by the Lawrence Livermore National Laboratory to demonstrate the potential advantages of visible hyper-spectral imaging in the detection of buried land mines. The data used in this study were generated by SAIC using the same type of imaging device mentioned in the above study although in a completely different environment. For a detailed description of this imaging system termed the *Hyper-Spectral Imager*, see the SAIC document [14].

The dimensions of the data set considered here are given by $100 \times 128 \times 128$ with the last dimension indicating the number of spectral bands in the cube. The 128 spectral bands range over the span of 400-950 nanometers of wavelength in the spectrum. Fig 1 shows a scene containing a building in the background and vegetation in the foreground in four separate spectral bands: 0, 48, 78 and 100. The choice was guided by the fact that different sets of objects in the scene appear more, or less highlighted in different spectral bands. For the purpose of identifying individual pixel characteristics based on a composite of the available spectral bands, we have computed the interspectral features: mean, standard deviation, minimum and the maximum over the 128 spectral bands for each pixel resulting in four images (fig 2). Fig 3 shows the 'centered' images corresponding to the aforementioned bands where for each pixel, the mean pixel intensity (over all bands) has been subtracted. And fig 4 shows the same four bands 'standardized' by dividing the images in fig 3 by the

pixel standard deviation for each pixel. While the centered images have enhanced some specific objects in them, the standardization has really highlighted the major classes of objects in the scene. The process of choosing the maximum and minimum for each pixel has been applied also to the standardized images and shown in fig 5. The highlighting effect has indeed been dramatic in this case especially with the maximum. The point in all this exercise has been to explore the extraction of inter-spectral features that may help in identifying a pixel as belonging to one of several possible classes in a scene albeit in an ad hoc manner.

The search for inter-spectral features can be made in a more systematic manner if we first examine the inter-spectral covariance (correlation) structure in the data cube. Fig 6 shows the pair-wise correlation of the 128 spectral bands in the SAIC building image displayed as an 'intensity' map. Specifically, we regard intensities at two bands as two variables, their observed values being the intensities at these two bands for a given pixel. The sample correlation coefficient between the two bands is computed based on the pixels of the imagery as observed samples. In the intensity map shown in fig 6 the intensity at the (i,j) th location is the computed correlation coefficient between the bands i and j . The correlation coefficients between two different bands are all positive and range approximately from 0.267 to 0.999 or more. The display also shows that the bands within the range 0 to 72 are rather highly correlated (yellow) providing spectral information of a redundant nature. It also shows that bands in the range 117-127 have relatively low correlation coefficients with those in the range 0 to 72 (0.25-0.35) indicating the presence of spectral information largely independent of those present in bands 0-72. Similar observations may be made for the spectral ranges 86-88 and 98-102 regarding their correlation coefficients with those in the range 0-72. The remainder of the range of bands within the range 73-127 bear moderate correlation with those in the range 0-72. Finally, within the band 0-72 a sub-band 0-8 appears to have relatively low correlation with the rest of the bands. In summary, the correlation intensity map is somewhat revealing in that it indicates the presence of two relatively uncorrelated broad bands in the regions 0-72 and 73-127. Within these broad bands there are several narrower

bands in the ranges 0-8 for the first and 86-88, 98-102 and 117-127 for the second where the degree of dependence with the broad bands 73-127 and 0-72 respectively is much weaker. The above observations about the inter-spectral correlation structure are largely qualitative in nature. To use the correlation structure in a constructive manner, we did an eigenstructure analysis on the correlation matrix. As perhaps could have been expected from the qualitative observations made earlier, the first four eigenvalues account for slightly more than 99% of the total of the eigenvalues and the first 9 about 99.96% (see fig 7). This indicates that with an orthogonal representation scheme along the principal vectors (the eigenvectors of the correlation matrix) more than 99.96% of the variation in the data could be accounted for by the set of projections of the data along the first nine principal vectors. The components in the first four principal vectors have been pictured in fig 8. These are the 'weights' of the different spectral bands in the respective principal vectors. The first principal vector (in red) indicates approximately equal weight to the bands indicating that it represents the approximate average intensity over the bands. The second principal vector (in blue) suggests an approximate difference over two broad bands 0-70 and 71-127. The third (in purple) suggests an averaging of two narrow bands 0-20 and 110-127. The fourth (in green) does not seem to have any simple interpretation. The data cube was projected along the first nine principal vectors and the resulting images shown in fig 9. Clearly each of the resulting PC's is showing different sets of object features highlighted. A detailed empirical study of this correspondence was not undertaken due to a lack of 'ground' truth regarding the data cube. However, it was felt that the principal components have shown sufficient object separation within the images for meaningful segmentation of the scene. Among several segmentation schemes that use multiple features of a pixel, we chose to implement the K-Means algorithm described in sec. 3. Fig 10 shows the results of the application of this algorithm to our data. We used the first nine principal components as pixel features for the clustering. The number of clusters assigned were 3, 5, 7, and 10 respectively. A comparison with the original images shown in fig 1 shows that the clustering in each case has provided meaningful segmentation of the scene into

physically distinct objects. Since no ground truth was available, no attempts were made to validate the clusters.

5. Conclusions and Future Work

We have outlined in this report some aspects of the analysis of hyper-spectral data obtained as outputs from Fourier transform imaging spectrometers currently under development at the Lawrence Livermore National Laboratory. The approach we have taken here is from a statistical perspective with data reduction in a meaningful way as the primary focus. Clearly, other approaches such as those based on *artificial neural networks* or *fuzzy set theory* or a combination of two or more of these are possible. In our future study we will concentrate on one or more of these approaches. Even the statistical approaches indicated here are only sketchy at best and need to be expanded further. For example, in the principal component approach, one could test for the significance of the smaller eigenvalues targeted for possible truncation with a view to data reduction. The clustering can be handled in various possible manners. Statistical tests for *cluster validity* after clustering or during the clustering in an interactive fashion, can be applied. *Quantitative studies* of individual clusters and/or segments can be made based on regression analysis. Such studies would assist in the development of methodologies for building chemical profiles of gas plumes or other atmospheric objects by 'remote sensing'. Development of methodologies for *sharpening* of image segment boundaries based on multichannel data are in progress. The *sensitivity* of the different aspects of hyperspectral data analysis to the *bandwidth* of the spectra can be studied on a statistical basis. All of these with applications to real experimental data sets are relegated to our future work with hyper-spectral data.

Acknowledgment

This work was supported by the LIFTIRS Project of the V-Div. (formerly L-Div.) of the Lawrence Livermore National Laboratory, University of California.

References

1. Carter, M.R., C. Bennett, D. Fields, J.Hernandez, Gaseous Effluent Monitoring and Identification Using an Imaging Fourier Transform Spectrometer, UCRL-JC-114619, Oct. 1993.
2. Johnson, R.A. and D.W. Wichern, Applied Multivariate Analysis, 2nd ed., Prentice Hall, Englewood Cliffs, NJ.,1988.
3. Preisendorfer, R.W. Principal Component Analysis in Meteorology and Oceanography, Elsevier Science Publishing Co, New York, 1988.
4. Rabbani, M. and P.W. Jones, Digital Image Compression techniques, Soc. Photo-opt.Instr.Engrs Bellingham Washington, 1991.
5. Jones, M.C. and R. Sibson, "What is Projection Pursuit," The journal of the American Statistical Association, 82 249-266, 1987.
6. Simpson, P.K., Artificial Neural Systems:Foundations, Paradigms, Applications and Implementation, Pergamon Press, NY 1989.
7. Levine, M.D., Vision in Man and Machine, McGraw Hill, New York, 1985.
8. Pavlidis, T., Structural Pattern Recognition, Springer Verlag, 1977.
9. MacQueen, J.B., 'Some Methods for Classification and Analysis of Multivariate observations, 'Proc. 5th Berkeley Symp. on Math. Stat. and Prob., 1, Berkeley, Calif.: Univ. of Calif. Pr., 281-297, 1967.
10. Hall, D.J. and D.K. Khanna, 'The ISODATA Method of Computation for the Relative Perception of Similarities and Differences in Complex and Real Data,' in Statistical Methods for Digital Computers, Eds. K. Enslein, A. Ralston, and H.S. Wilf, J. Wiley and Sons, New York, 1977.

11. Hartigan, J.A. and M.A. Wang, 'Algorithm AS 136: A K-Means Algorithm,' *Applied Statistics*, **28**, 100-108, 1979.
12. Johnson, R.A. and D.W. Wichern, *Applied Multivariate Statistical Analysis*, 2nd Ed., Prentice Hall, 1988.
13. Jain, A.K. and R.C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.
14. Anderson, Richard, 'Hyperspectral Imaging of Emplaced Land Mines,' SAIC Report prepared for the Lawrence Livermore National laboratory, Subcontract No.: B256727, Dec 31, 1993.

Figure Legends:

1. SAIC_2 data : some chosen spectral bands are shown. Counter-clockwise from bottom left are bands 0,48,78 and 100 respectively.
2. SAIC_2 data :Counter-clockwise from bottom left are the mean, standard deviation, minimum and maximum respectively, computed on a pixel by pixel basis over 128 spectral bands.
3. Same as 2 except centered at the respective means over the bands.
4. Same as 3 except scaled by the standard deviation over the bands.
5. SAIC_2 data :The minimum(left) and the maximum(right) over the bands(after standardization).
- 6.(a-b) SAIC_2 data :Interband correlation coefficients displayed as intensity image. $0.2676 \leq \rho(i, j) \leq 0.99995$, $0 \leq i, j \leq 127$, $i \neq j$. Fig 6a shows the contour map of the interband correlation. Fig 6b shows the correlation structure in multiple colors.Increasingly ordered in the order: Blue, Green, Red, and Yellow.
7. Cumulative percentage of variance 'explained' by successive principal components. The first 9 principal components explain 99.96% of total variation in the data.
8. Weights of the 128 bands in the first four principal components. Red,blue purple and green respectively represent the weights for the first, second, third and fourth principal components.
9. First 9 principal components for the SAIC_2 data based on the correlation matrix. Shown in the order bottom to top, left to right, starting from bottom left.
10. Segmentation of the SAIC_2 data by the K-means clustering algorithm. Shown counter-clockwise from bottom left are the segmentations with K=3,5,7 and 10 respectively.