CIS-Kodak New Collaboration Research Grant Report

Visualization of High-dimensional Remote-Sensing Data Products

Hongqin Zhang, Ethan Montag, Dave Messinger
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
Rochester, NY, USA

October 2005

Abstract

This study investigated appropriate methodologies for displaying hyperspectral imagery based on knowledge of human color vision as applied to Hyperion and AVIRIS data. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) were used to reduce the data dimensionality, and these two methods were chosen also because of their underlying relationships to the opponent color model of human color perception. PCA and ICA-based strategies were then explored by mapping the first three PC or IC to several opponent color spaces including CIELAB, HSV, YCbCr, and YIQ. The gray world assumption, which states that given an image with sufficient amount of color variations, the average color should be gray, was used to set the mapping origins. The rendered images are well color balanced and can offer a first look capability or initial classification for a wide variety of spectral scenes.

I. Introduction

Hyper-Spectral imagers that can measure spectra at more than 200 narrow, contiguous bands have been developed in recent years. With the increasing number of bands, the remotely sensed hyperspectral imagery provides finer spectral resolution, yet presents challenges for data visualization techniques for display of the information content of such datasets. Color, as one of the attributes to display information, may be used effectively for more efficient data presentation and interpretation.

The goal of this research is to investigate the application of color and human vision system in the visualization of high-dimensional remote sensing products and develop visualization techniques based on principles of human vision, color appearance, and color perception for more efficient data exploitation and improved data mining.

II. Background

How can the hyperspectral information from a portion of the spectrum not perceived by humans be presented in a manner that is more easily understood? At first, in order to display the data on a standard tristimulus display, the dimensionality of the data bands must be reduced to three or less. Traditionally, one can map three widely spaced bands to RGB in hopes that they are minimally correlated or choose three specific bands to highlight a particular spectral feature, for example, the absorption peak of water. In addition to the popular simple methods, however, data projection is one of the common visual ways to get the interesting subsets of the original data by projecting high-dimensional data to lower dimensions. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are two widely used methods for such a purpose. In this study, these two methods were used to reduce the data dimensionality in order to make the data more amenable to visualization in three-dimensional color space. In addition, these two methods were chosen because of their underlying relationships to the opponent color model of human color perception.

Hyperspectral data bands are often highly correlated. The Principal Component (PC) Transformation can be used to produce uncorrelated output bands. This is done by finding a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized. PC bands are linear combinations of the original spectral bands and are uncorrelated. Unlike the intensity distribution in any particular band, the PCs provide spectral information spanning the entire spectral range of interest, preventing any one feature from being completely missed. Generally, the first 3 PC bands can explain more than 90% of the entire variance. While a common method for presenting PC images in a pseudocolor display is to map the first three PCs into an RGB image, a certain amount of research has been performed on perceptually based color displays for use with scientific data [Tyo, etc., Beauchemin, etc.]. Incorporating knowledge of human color processing into display strategies for spectral imagery was thought to be helpful in order to create images that present the rich information available from spectral sensors in a readily interpretable manner.

It is well known that the human vision system processes color by means of an achromatic channel and two opponent-color channels, which is called the opponent-color model of human color perception. It was demonstrated that the achromatic, red-green opponent, and blue-yellow opponent-color processing channels represent statistically noncovariant information pathways derivable from a PC analysis of the spectral sensitivities of the three classes of photoreceptors [Buchsbaum and Gottschalk, 1983]. Their result was the first mathematical explanation for why the human vision system processes color via opponent channels. Therefore, the opponent color formulation can be used to develop a display strategy for high-dimensional spectral imagery based on a similar PC analysis.

ICA is a statistical and computational technique that extracts independent source signals by searching for a linear or nonlinear transformation which minimizes the statistical dependence between components [Du, et al.]. It finds a nonorthogonal coordinate system in any multivariate data that minimizes mutual information among the axial projections of the input data. The directions of the axes of this coordinate system are determined by both second-order and higher-order statistics of the original data. The goal of ICA is to perform a transform that makes the resulting source outputs as statistically independent of each other as possible.

ICA was first proposed by Pierre Comon in 1994, and has been used in a variety of applications such as blind source separation, recognition, etc. As opposed to an orthogonal PCA basis, ICA has also been considered as a method to analyze natural scenes. Much interesting research have been conducted to investigate the spatial and chromatic structures of natural scenes by decomposing the spectral images into a set of linear basis functions using ICA. Wachtler, et al. [Wachtler] applied ICA to hyperspectral images to learn an efficient representation of color in natural scenes. Their founding suggests that nonorthogonal opponent encoding of photoreceptor signals leads to higher coding efficiency and is a result of the properties of natural spectra. ICA may be used to reveal the underlying statistical properties of color information in natural scenes.

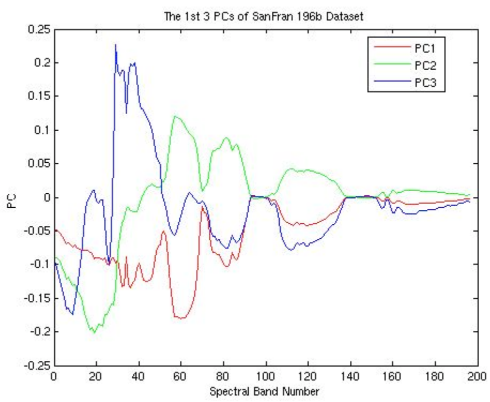
III. PCA and ICA processing

Imagery from the Hyperion[6] sensor on the Earth Observing 1 (EO-1) spacecraft was used as the initial data set for this project. This sensor can image a 7.5 km by 100 km land area per image in 220 spectral bands ranging from 0.4 to 2.5 μm with a 30-meter resolution per pixel. An urban[7] image of an area near San Francisco was analyzed using techniques developed by co-PI David Messinger to create 15 abundance maps. The raw radiance data have been converted to reflectance through an atmospheric compensation scheme to aid the un-mixing process.

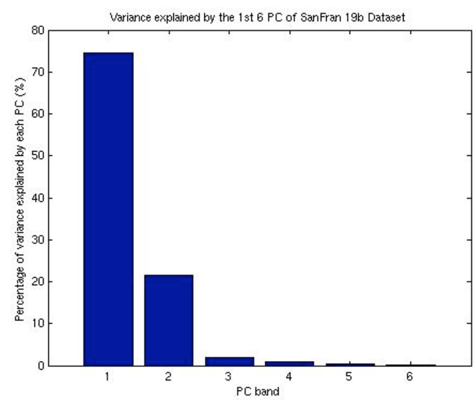
Additional imagery is available from the Hyperion site as well as other imagery provided by DIRS.

In the PCA and ICA processing step, the raw radiance data with removal of bad bands was used, and another Nevada scene from AVIRIS sensor with 50 bands selected (from band 172 to 221) was also tested. These datasets were designated as SanFran_196b, SanFran_100b, and Nevada_50b. The PCA and ICA results of these datasets are listed in the following tables.

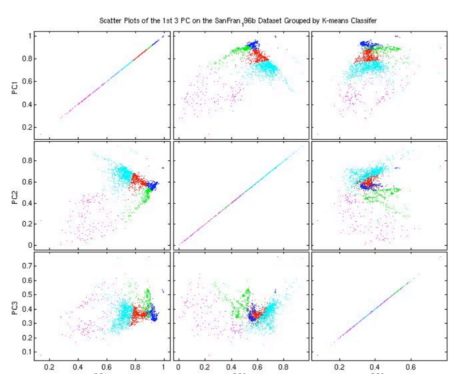
1. SanFran_196b_PCA



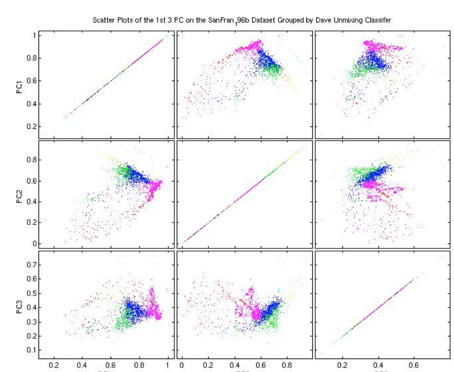
The 1st 3 PC (basis vectors)



PC1 -> 74.56%, PC2 -> 21.54%, PC3 -> 1.95%
Total by the 1st 3 PCs -> 98.05%

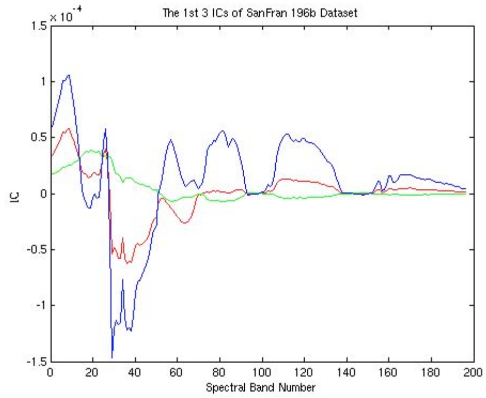


Scatter plots of the data projection onto the 1st 3 PC grouped by the K-means classifier results (5 classes)

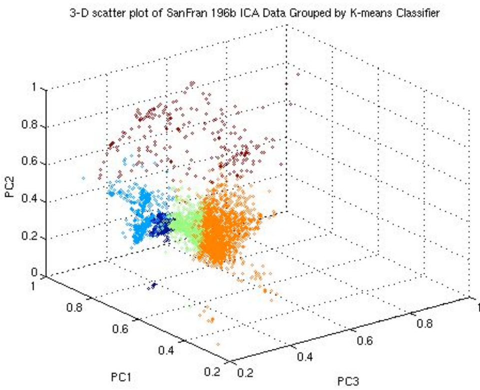


Scatter plots of the data projection onto the 1st 3 PC grouped by Dave's unmixing classifier results (15 classes)

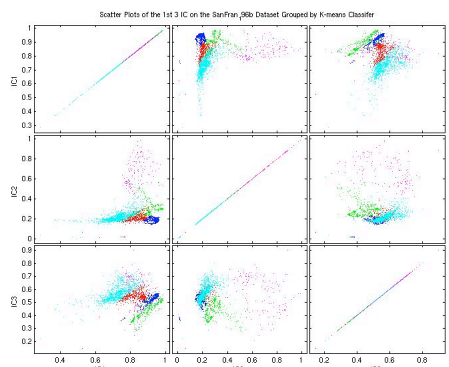
2. SanFran_196b_ICA



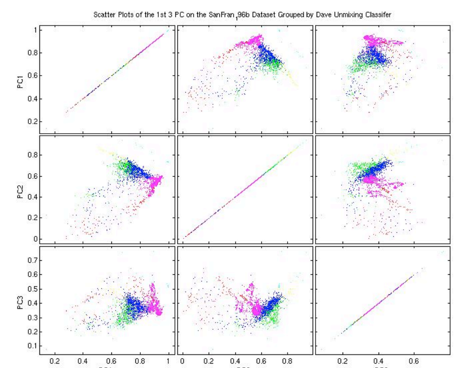
The 1st 3 Basis Vector of ICA



3-D scatter plot of the 1st 3 IC data grouped by K-means classifier results (5 classes)

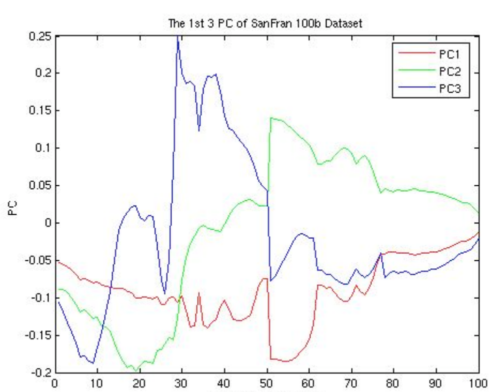


Scatter plots of the data projection onto the 1st 3 IC grouped by the K-means classifier results (5 classes)

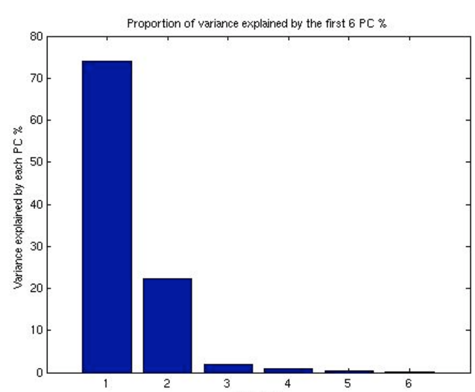


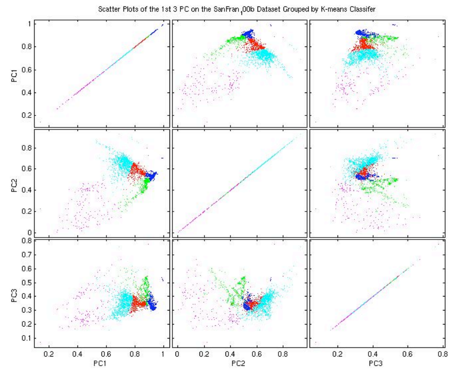
Scatter plots of the data projection onto the 1st 3 IC grouped by Dave's unmixing classifier results (15 classes)

3. SanFran_100b_PCA

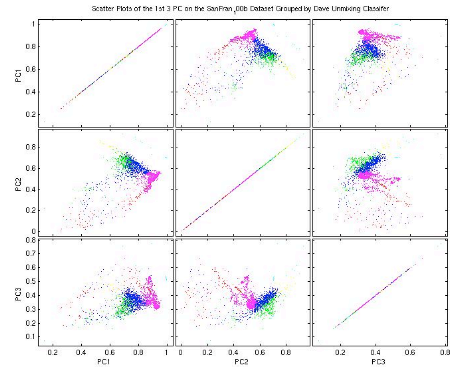


The 1st 3 Basis Vectors of PCA



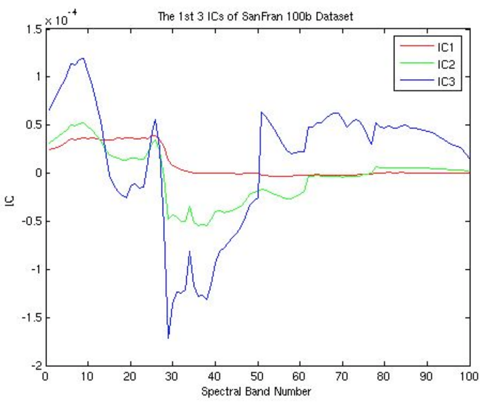


Scatter plots of the data projection onto the 1st 3 IC grouped by the K-means classifier results (5 classes)

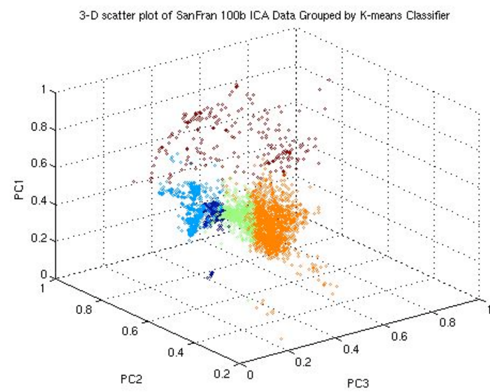


Scatter plots of the data projection onto the 1st 3 IC grouped Dave's unmixing classifier results (15 classes)

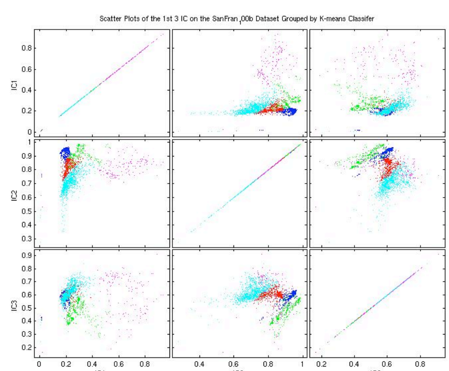
4. SanFran_100b_ICA



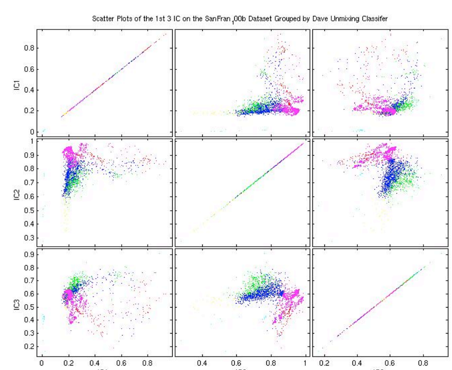
The 1st 3 Basis Vectors of ICA



3-D scatter plot of the 1st 3 IC data grouped by K-means classifier results

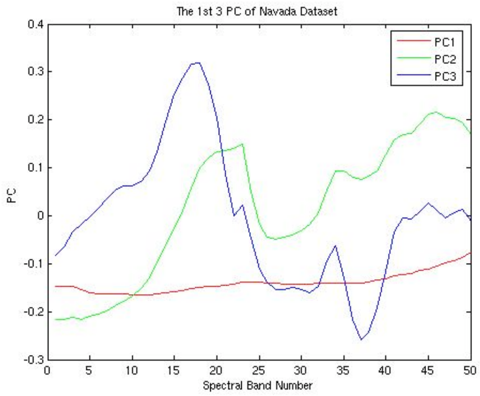


Scatter plots of the data projection onto the 1st 3 IC grouped by the K-means classifier results (5 classes)

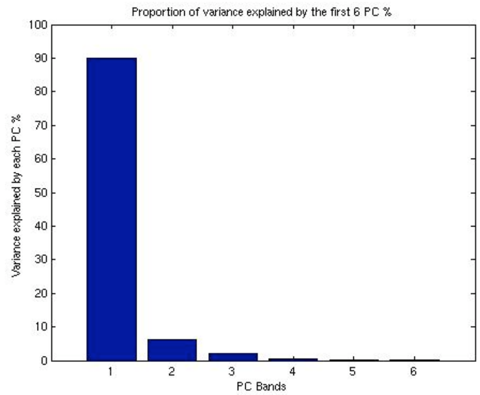


Scatter plots of the data projection onto the 1st 3 IC grouped by Dave's unmixing classifier results (15 classes)

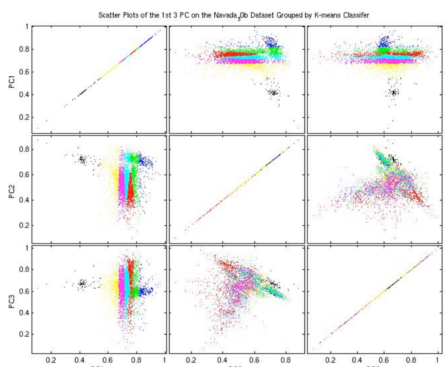
5. Navada_50b_PCA



The 1st 3 basis vectors of PCA

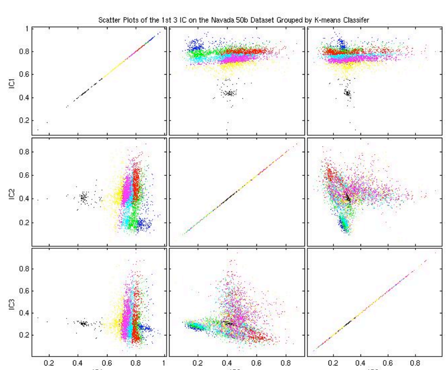


PC1: 89.996%, PC2: 6.35%, PC3: 2.16%



Scatter plots of the data projection onto the 1st 3 PC grouped by the K-means classifier results (7 classes)

6. Navada_50b_ICA



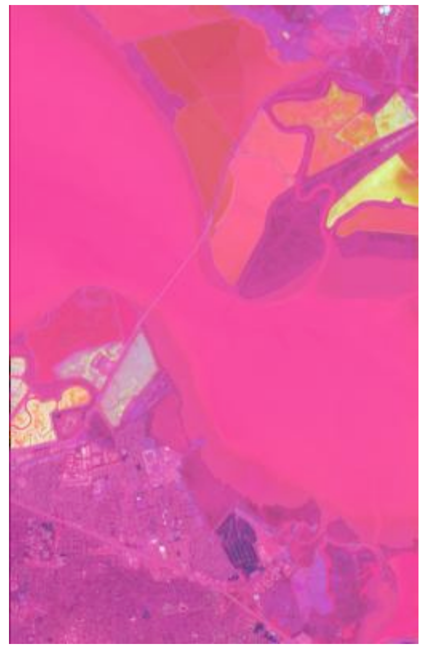
Scatter plots of the data projection onto the 1st 3 IC grouped by the K-means classifier results (7 classes)

IV. Color mapping strategies

1. Conventional method

A common method for presenting PC images in a pseudocolor display is to directly map the first three PCs into an RGB image as shown below.

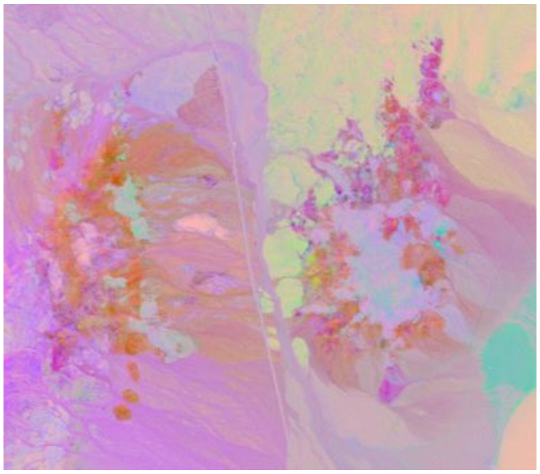
SanFran_196b_PCA – directly in RGB (normalized)



SanFran_196b_ICA – directly in RGB (normalized)

SanFran_100b_PCA– directly in RGB (normalized)

SanFran_100b_ICA– directly in RGB (normalized)



Navada_50b_PCA- directly in RGB
(normalized)

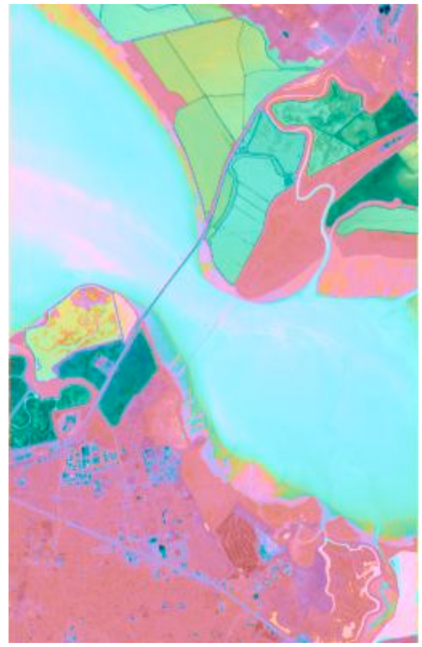
Navada_50b_ICA- directly in RGB
(normalized)

2. Perceptual rendering

PCA and ICA-based perceptual visualization strategies were then explored by mapping the first three PC or IC to several opponent color spaces including CIELAB, HSV, YCbCr, and YIQ. The gray world assumption, which states that given an image with sufficient amount of color variations, the average color should be gray, was used to set the mapping origins. The rendered images are well color balanced and can offer a first look capability or initial classification for a wide variety of spectral scenes. The resulting images in different opponent color spaces are shown below for each dataset and each dimensionality reduction method.

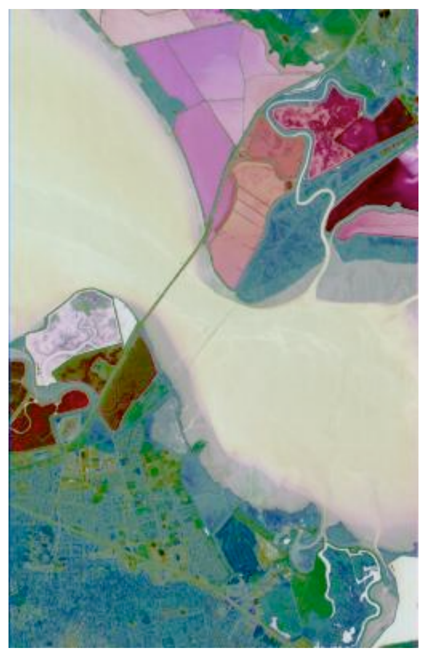
2.1 SanFran_196b_PCA

SanFran_196b_PCA_CIECAM02_opp2

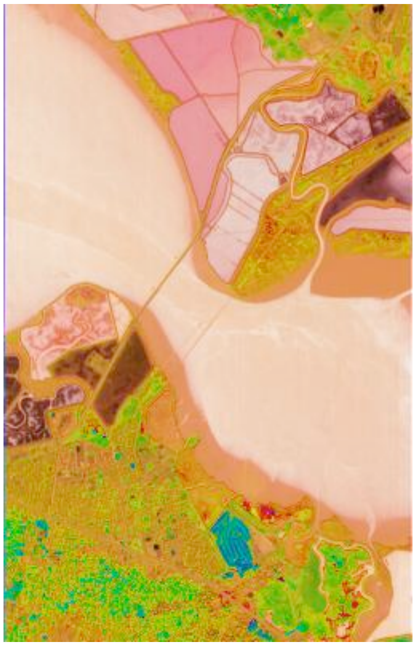


SanFran_196b_PCA_CIECAM02_pc2

SanFran_196b_PCA_CIELAB



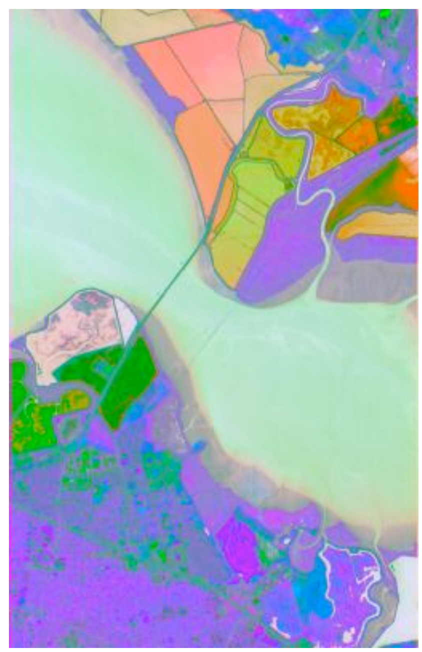
SanFran_196b_PCA_HSV_opp2_ClipS



SanFran_196b_PCA_HSV_pc1_NormS

SanFran_196b_PCA_MatlabYCbCr

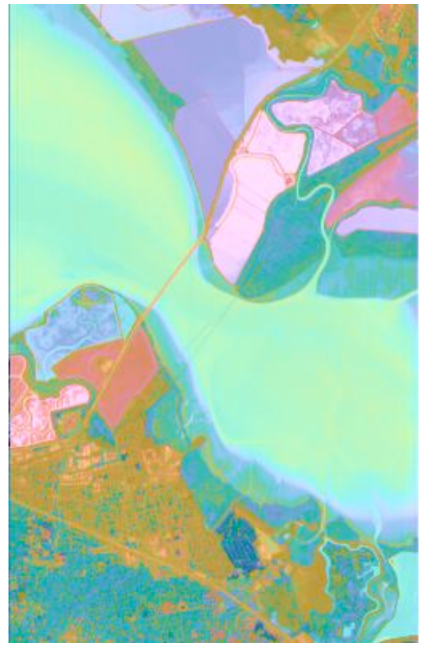
SanFran_196b_PCA_yiq



SanFran_196b_PCA_yuv

2.2 SanFran_196b_ICA

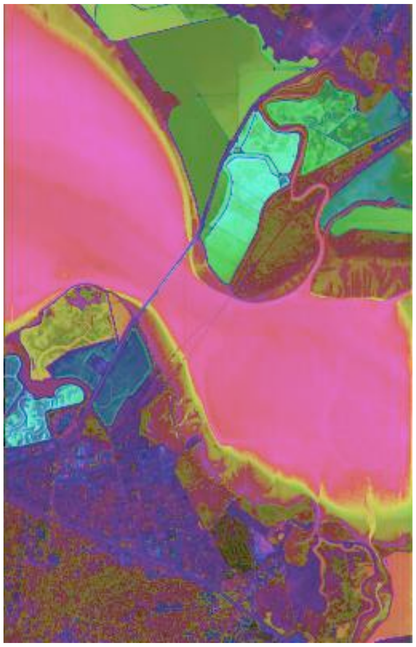
SanFran_196b_ICA_CIECAM02_opp2



SanFran_196b_ICA_CIECAM02_pc2

SanFran_196b_ICA_CIELAB2

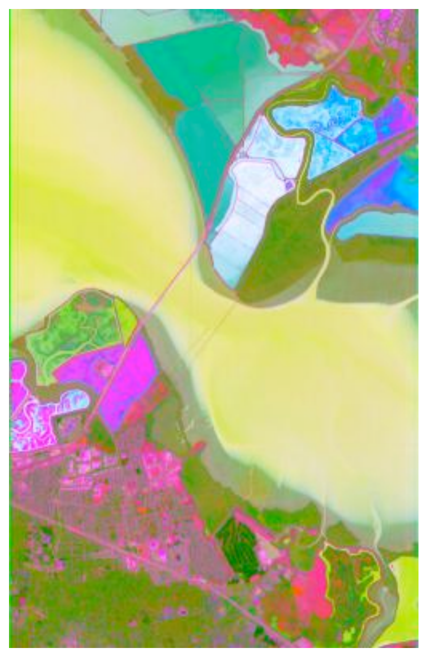
SanFran_196b_ICA_HSV_opp2_ClipsS



SanFran_196b_ICA_HSV_pc2_NormS

SanFran_196b_ICA_MatlabYCbCr

SanFran_196b_ICA_yiq



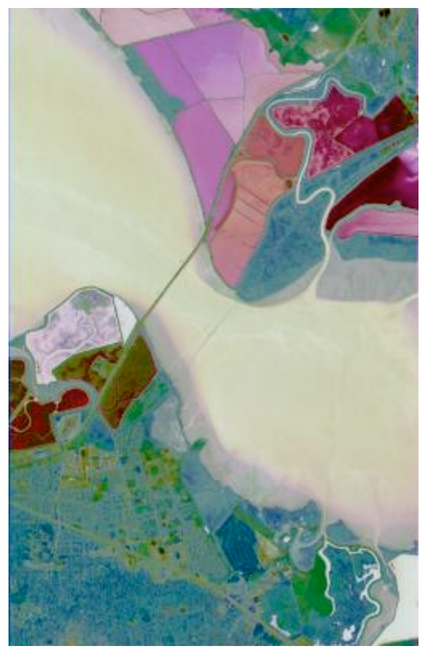
SanFran_196b_ICA_yuv

2.3 SanFran_100b_PCA

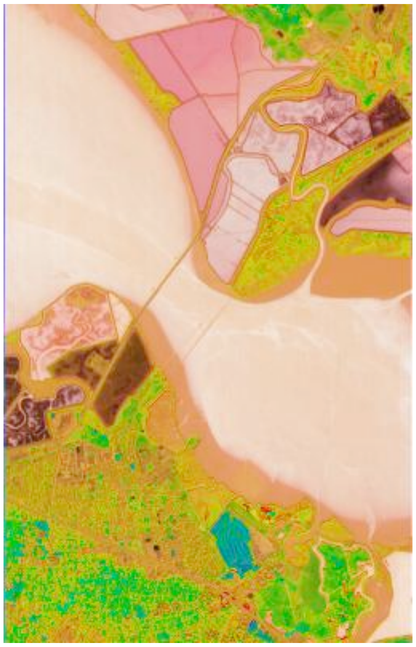
SanFran_100b_PCA_CIECAM02_opp2

SanFran_100b_PCA_CIELAB

SanFran_100b_PCA_opp1_ClipS



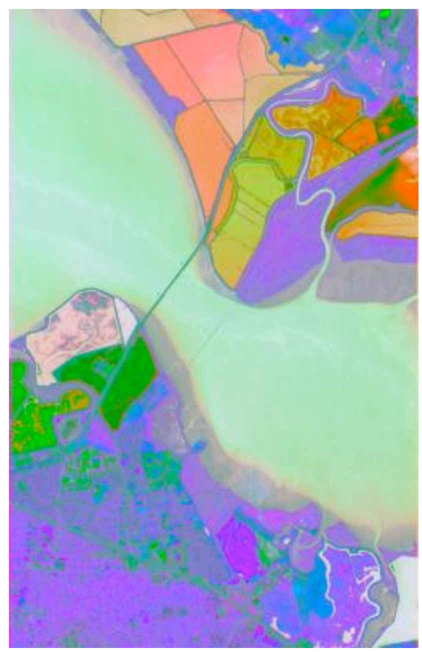
SanFran_100b_PCA_opp2_ClipS



SanFran_100b_PCA_HSV_pc1_NormS

SanFran_100b_PCA_MatlabYCbCr

SanFran_100b_PCA_yiq



SanFran_100b_PCA_yuv

2.4 SanFran_100b_ICA

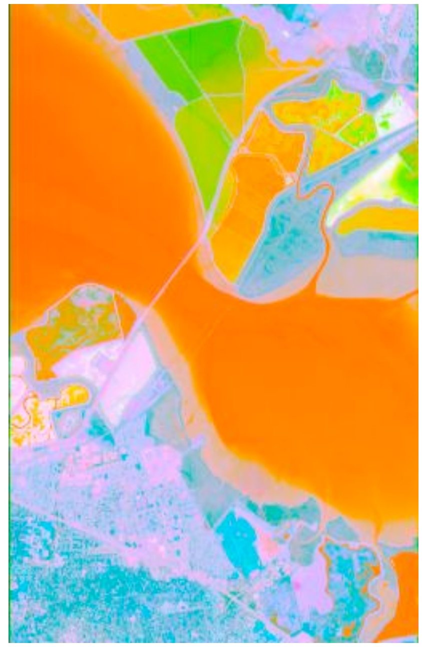
SanFran_100b_ICA_CIECAM02_opp2

SanFran_100b_ICA_yuv

SanFran_100b_ICA_CIELAB

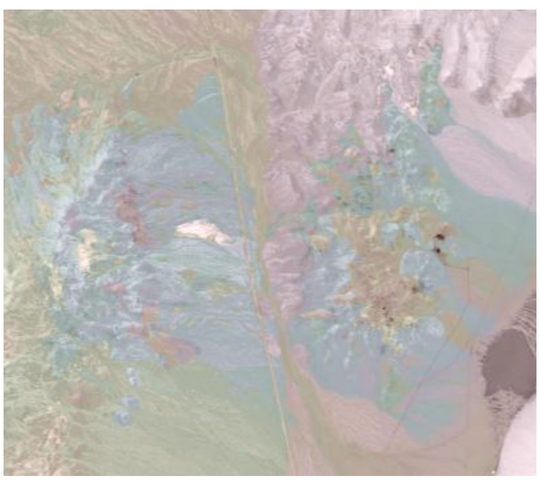
SanFran_100b_ICA_HSV_opp2_ClipS

SanFran_100b_ICA_MatlabYCbCr



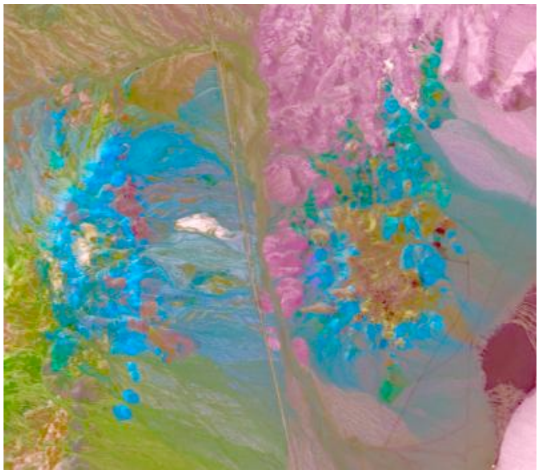
SanFran_100b_ICA_yiq

2.5 Nevada_50b_PCA



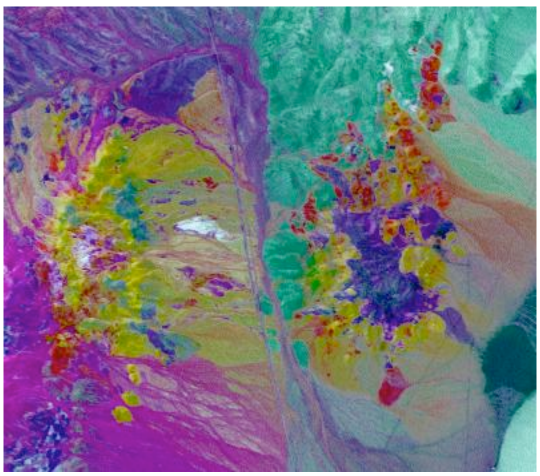
Nevada_50b_PCA_CIECAM02_opp2

Nevada_50b_PCA_CIECAM02_pc2

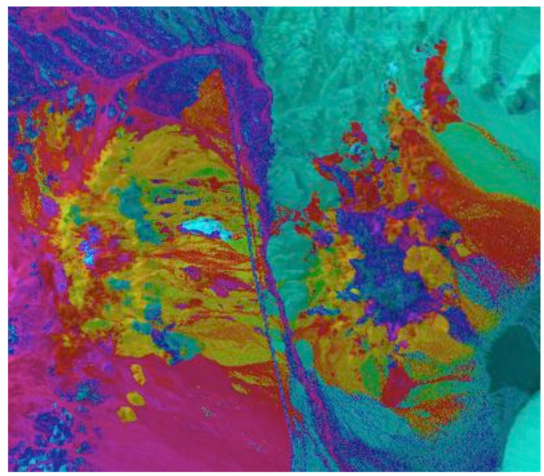


Nevada_50b_PCA_CIELAB

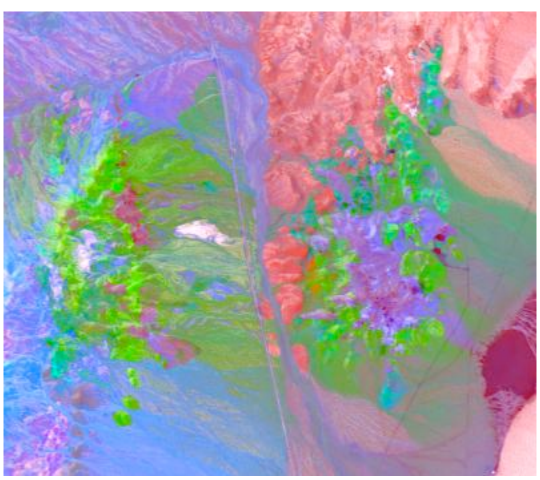
Nevada_50b_PCA_HSV_opp1_ClipS



Nevada_50b_PCA_HSV_opp2_ClipS

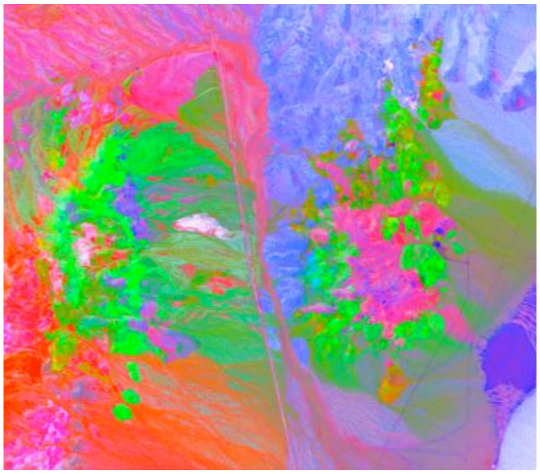


Nevada_50b_PCA_HSV_pc_NormS



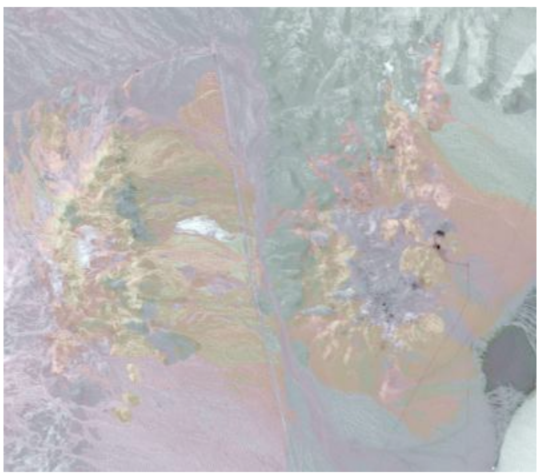
Nevada_50b_PCA_MatlabYCbCr

Nevada_50b_PCA_yiq

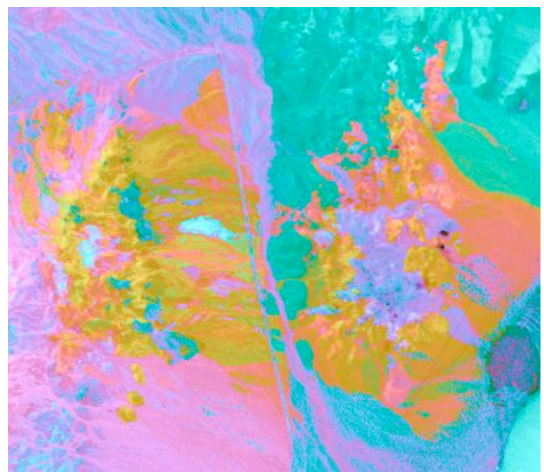


Nevada_50b_PCA_yuv

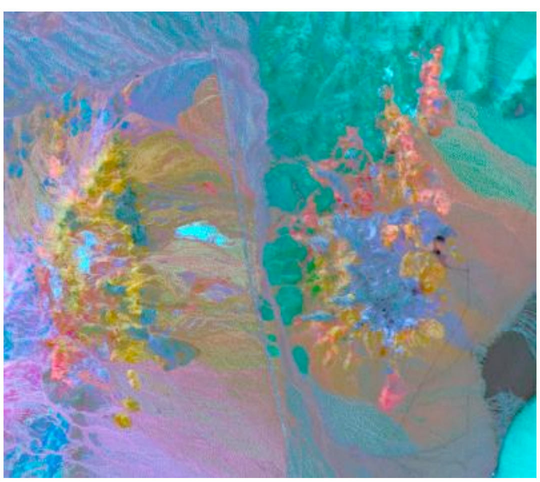
2.6 Nevada_50b_ICA



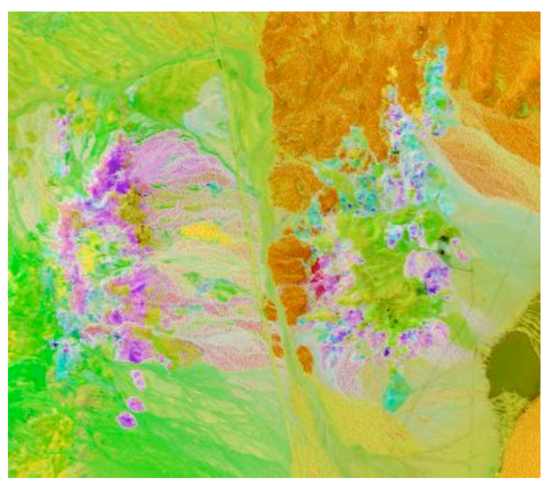
Nevada_50b_ICA_CIECAM02_opp2



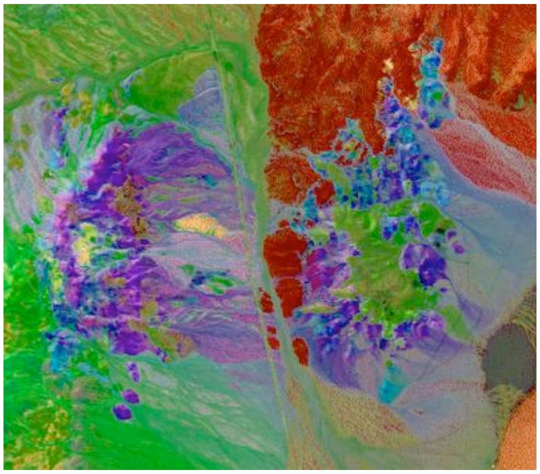
Nevada_50b_ICA_CIECAM02_pc2

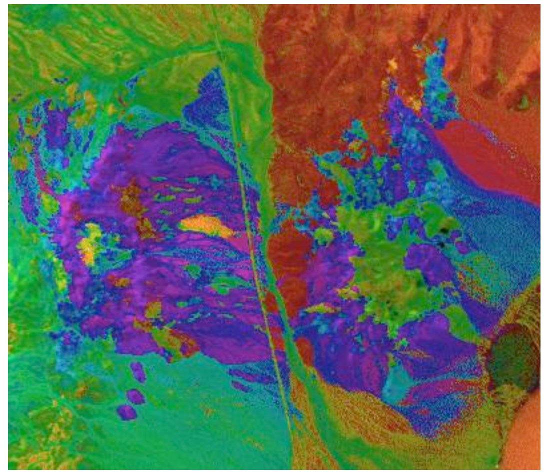


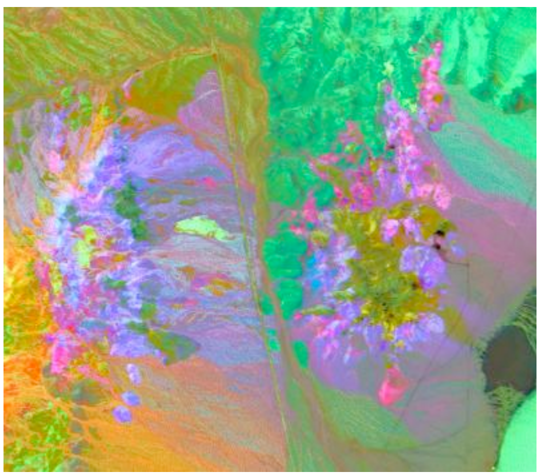
Nevada_50b_ICA_CIELAB



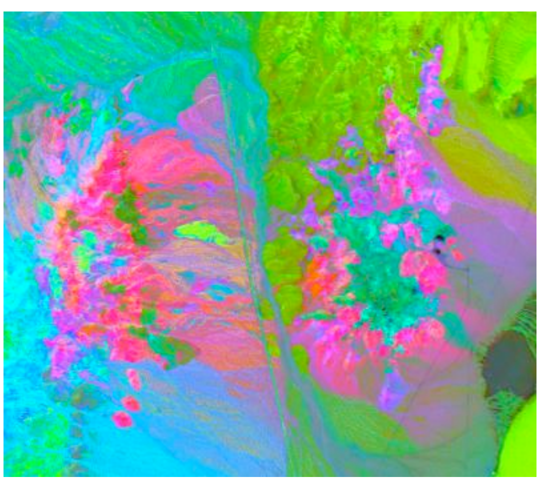
Nevada_50b_ICA_HSV_opp1_ClipS


Nevada_50b_ICA_opp2_ClipS


Nevada_50b_ICA_HSV_pc2_NormS


Nevada_50b_ICA_MatlabYCbCr


Nevada_50b_ICA_yiq


Nevada_50b_ICA_yuv

V. Conclusion

These document summaries the results of the project. At this stage of the research, we can see from the above section IV that the rendered images are well color balanced and can offer a first look capability or initial classification for a wide variety of spectral scenes.

Though the resulting images may be visually appealing, they may assign different hues for pixels that are identified as belonging to the same class. Looking at the scatter plots in section III of the first three components grouped by classification results, it's seen that both PCA and ICA have the ability to capture the cluster structure in the dataset by just using the first few components which can be used prior to cluster analysis, e.g. before performing K-means clustering to determine a good value for K. A good visualization technique based on PCA or ICA should be able to illustrate this cluster feature. According to the research on the perception and understanding of different color attributes [Zhang and Montag], we desire a different hue for each class. Therefore, further research will be to develop supervised methods by examining scatter plots of PCA or ICA data and the corresponding classification map with the goal of showing a certain class with a certain hue while maintaining a good hue separation among classes.

The usefulness of these visualization strategies may be evaluated and compared with other visualization techniques found in the literature by psychophysical experiments.

VI. Publication

Results from the work has been submitted to the SPIE Defense and Security Symposium, Conference on "Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery", April 2006.

VII. Acknowledgment

We would like to take this opportunity to thank Kodak for their financial support for this project.

Reference

1. Beauchemin, M. and Fung, K. B., Intensity-Hue-Saturation Color Display Transform For Hyperspectral data. The 21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada, 21-24 June 1999.
2. Buchsbaum, G. and Gottschalk, A., 1983. Trichromacy, Opponent Colors Coding and Optimum Color Information Transmission in the Retina. Proc. R. Soc. Lond. B, Vol, 220, pp. 89-113.
3. Du, H., et al., Band Selection Using Independent Component Analysis for Hyperspectral Image Processing, Dept. of Elec. And Comp. Engineering, University of Tennessee.
4. Tyo, J. S., Diersen, D. I., and Olsen, R. C., Principal-Components-Based Display Strategy for Spectral Imagery, IEEE Transactions on Geoscience and Remote Sensing, Vol.41, No.3: 708-718.
5. Zhang, H., Montag, E. D., How Well Can People Use Different Color Attributes? The 12th Color Imaging Conference, Scottsdale, AZ, 2004.
6. Wachtler, T., Lee, T. and Sejnowski, T., 2001. Chromatic Structure of Natural Scenes, J. Opt. Soc. Am. A/Vol.18, No.1: 65-76.
7. http://eol.usgs.gov/hyperion.php
8. http://eol.usgs.gov/sampleurban.php