

Spectral quality metrics for terrain classification*

John P. Kerekes^{a#} and Su May Hsu^b

^aCenter for Imaging Science, Rochester Institute of Technology, Rochester, New York USA

^bLincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts USA

ABSTRACT

Current image quality approaches are designed to assess the utility of single band images by trained image analysts. While analysts today are certainly involved in the exploitation of spectral imagery, automated tools are generally used as aids in the analysis and offer hope in the future of significantly reducing the analysis timeline and analyst work load. Thus, there is a recognized need for spectral image quality metrics that include the effects of automated algorithms.

Previously, we have reported on candidate approaches for spectral quality metrics in the context of unresolved object detection. We have continued these efforts through the use of empirical trade studies in the context of ground cover terrain classification. HYDICE airborne hyperspectral imagery have been analyzed for the effects on scene classification accuracy of spatial resolution, signal-to-noise ratio, and number of spectral channels. Various classification algorithms including Gaussian maximum likelihood, spectral angle mapper, and Euclidean minimum distance have been considered. Performance metrics included classification accuracy, confusion matrices, and the Kappa coefficient. An extension of the previously developed Spectral Quality Equation (SQE) has been developed for the terrain classification application.

As expected, the accuracy of terrain classification shows only modest sensitivity to the parameters considered, except at the extreme cases of high noise, few bands, and small ground resolution. However, these results are useful in continuing to develop the quantitative relationships necessary for characterizing the quality of spectral imagery in various applications.

Keywords: Spectral imaging, spectral quality, terrain classification accuracy

1. INTRODUCTION AND CONTEXT

The ability to quantitatively assess the quality of a multispectral or hyperspectral image is desirable for many reasons including instrument comparisons and trade studies, image archive retrieval and tasking of data collections. However, the notion of the “quality,” of a spectral image will depend upon many disparate factors including characteristics of the scene, the sensor, the algorithms applied, and the desired product. While an image with just a few bands but high spatial resolution may have high quality judged by someone looking at spatial information, an image with many bands but moderate spatial resolution may have even higher quality when judged by an analyst looking at spectral information. It is precisely these tradeoffs that one seeks to quantify in the development of a spectral quality measure.

The myriad of applications of spectral imagery and the dependence of perceived quality on the interrelationships of parameters make the task of assigning a quality measure to spectral imagery very daunting. It is desirable that a spectral quality measure describe the potential utility of an image in an application of interest. Previously, we have reported on candidate approaches for spectral quality metrics in the context of unresolved object detection [1] using spectral imaging sensors operating in the reflective portion of the optical spectrum. In this paper we present a continuation of these efforts through the use of empirical trade studies in the context of ground cover terrain classification.

* This work was sponsored by the Department of Defense under contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and not necessarily endorsed by the United States Government.

[#] kerekes@cis.rit.edu; Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, New York, 14623

There have been other efforts at assigning a quantitative quality measure to hyperspectral imagery including ones based on analysts' interpretation of quality and utility [2] [3], spectral similarity [4], and the relationship of detection and false alarm probabilities with collection parameters [5]. These efforts offer alternative approaches to this difficult problem and demonstrate promise in the context of their work.

This paper describes our approach and initial results in developing a candidate spectral quality metric in the context of terrain classification. Classification accuracy as a quality metric has been used to measure the effects of lossy compression of spectral imagery [6], [7]. In our work, we study how the accuracy varies as a function of spectral sensor parameters, and then use those results to explore ideas to develop a quantitative spectral quality metric for that application.

First, we review our previous discussions of the issues regarding spectral quality and describe a notional concept of a spectral quality equation and a spectral quality rating scale. We then review the General Image Quality Equation (GIQE) that was used as a model for the proposed Spectral Quality Equation (SQE). Parameter tradeoff sensitivities are investigated using airborne hyperspectral imagery and the results presented. We conclude with discussions of some ideas on spectral quality and topics for further research.

2. SPECTRAL QUALITY PARAMETERS AND NOTIONAL CONCEPT

2.1 Spectral Quality Discussion

The notion of the "quality," of a multispectral or hyperspectral image deserves some discussion, as there is not a universally accepted definition of the term. In this usage, the dictionary defines quality as "degree or grade of excellence.,, Thus, a spectral quality measure should contain a monotonically increasing scale that represents the degree of excellence (in some sense) of a spectral image. The use of a numerical scale to describe the quality is a convenient way of ordering the values and comparing disparate images. It is intuitive that the higher the numerical value, the higher the quality.

As mentioned in the Introduction, the degree of excellence, or quality, of an image can depend on the particular use of the image. A farmer looking for signs of a non-localized disease outbreak in his crops may not be concerned whether the imagery has 5 or 10 meter spatial resolution, but rather whether it has the spectral resolution and band locations to detect the impending stress. On the other hand, a military analyst looking for small objects may be more concerned about the spatial resolution than the spectral content. Thus one must consider the application when discussing spectral image quality.

Another aspect of assigning a numerical value for the quality of a spectral image is to distinguish between potential and actual value. A spectral image may have the signal-to-noise ratio and spectral resolution to identify a crop type, but there may be a cloud present over the field of interest in the image. Thus, while the image has high potential quality, its actual utility is low for the problem of interest.

In this and our previous work on the topic, we have elected to constrain the investigation to one application at a time before trying to generalize the findings. Also, we will use the term quality to describe the *potential* value of the spectral image and set aside issues regarding obscuration in a given image.

2.2 Parameters Affecting Spectral Quality

Given the constraints we have adopted in this work, there still remain a large number of parameters of the remote sensing system that will affect the quality of the imagery. Table 1 lists a number of these parameters grouped according to their place in the end-to-end remote sensing system.

While this list is not comprehensive, it still contains many diverse parameters. Again, to make this effort tractable, we further reduce the parameters studied to the three most likely to affect quality: spatial resolution, spectral resolution, and signal-to-noise ratio (highlighted in bold in Table 1).

Table 1. Example parameters affecting spectral image quality in current context. **Bold** parameters are the focus in this work.

Scene	Target of interest (size and spectral complexity) Background class(es) and their complexity Atmospheric state Solar zenith angle
Sensor	Spatial resolution and sampling Spectral resolution and sampling Spectral region Signal-to-noise ratio Calibration accuracy (radiometric and spectral) Total field of view View angle
Processing Algorithms	Spectral selection and/or feature extraction Atmospheric compensation Classification algorithm

2.3 Spectral Quality Notional Concept

Now that we have discussed the definition, context, and parameters affecting spectral quality, let us review the approach we have pursued with a simple diagram as shown in Figure 1. Here, we see three axes representing the primary parameters of a spectral imaging system selected above. Ground resolved distance (GRD) represents spatial resolution. Channel spectral bandwidth, or $\Delta\lambda$, represents spectral resolution. Noise level is used as a surrogate for signal-to-noise ratio.

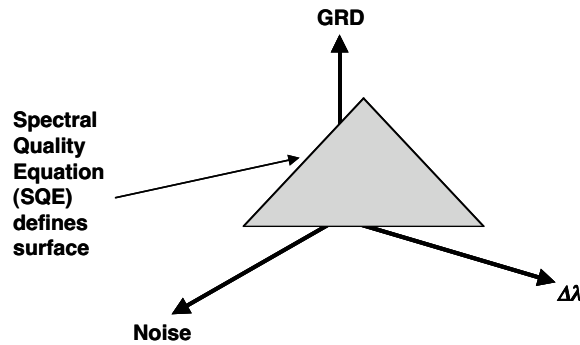


Figure 1. Notional concept of the Spectral Quality Equation defining a surface of constant quality.

Note that as one moves away from the origin in Figure 1 the values of the parameters increase and lead to a degradation of the image quality in an intuitive sense. The notional surface drawn in the figure represents the concept of tradeoffs in these parameters that lead to a constant quality level. The surface defines the concept of a spectral quality equation (SQE), which describes these tradeoffs in quantitative terms. Parameter combinations that lie on the surface described by the SQE will have equivalent quality. This concept is extended to an idea of a Spectral Quality Rating Scale (SQRS), which leads to multiple surfaces at various distances from the origin, with surfaces closer to the origin having a higher numerical value (greater quality). The tasks that can be accomplished at the various SQRS levels can then be defined using a qualitative description as with the National Imagery Interpretability Rating Scale (NIIRS) [8] developed for single band imagery.

3. REVIEW OF DETECTION SPECTRAL QUALITY EQUATION

As mentioned above, our previous efforts at developing a spectral quality equation were focused on the application of unresolved object detection with spectral imagery. To this end, we developed some initial spectral quality equations modeled after the General Image Quality Equation (GIQE) [9]. The GIQE was developed to map single band electro-

optic imaging sensor parameters to the NIIRS levels. Since the GIQE and NIIRS are well known and established in the literature, we felt they were a good foundation on which to build an extension for spectral image quality efforts.

This previous work investigated the tradeoffs between the parameters identified in Section 2 (GRD, spectral resolution, and SNR), using an analytical spectral performance prediction model [10] that ran quickly and allowed a large number of parameter combinations to be studied efficiently. A set of observation scenarios with three targets and three backgrounds were defined, and then a large number of trade studies conducted. A constant performance criterion (e.g., $P_D = 0.8$, $P_{FA} = 10^{-5}$) was adopted and then surfaces constructed along the lines of the notional one shown in Figure 1. Then, a linear function was fit to these surfaces using regression techniques, and draft spectral quality equation developed. The resulting SQE for subpixel detection is shown in equation 1 where GRD represents the ground resolved distance in cm, SNR is the signal-to-noise ratio, and N is the number of channels in the reflective solar spectral region (0.4 to 2.5 μm).

$$SQRS_{model_detection} = 9.65 - 3.22 \log_{10}[GRD(cm)] + 0.44 \log_{10}[SNR] + 0.81 \log_{10}[N] \quad (1)$$

In developing equation 1, we selected one significant constraint. Since the spectral resolution comes in through the specification of the number of channels N in the reflective solar region, the equation could mathematically be used to predict the SQRS for a single band panchromatic EO image by setting $N = 1$. Given that there is the established NIIRS scale for this case, it would seem useful to have the SQE predict similar values as the GIQE. This is also an intuitive constraint since a spectral image could be used to synthesize a panchromatic image by adding all spectral images together to form one panchromatic image. Another constraint we employed was to place the typical state of the art in airborne imaging spectrometers at about the middle of a 0 – 9 rating scale.

While the model-based analysis provided a convenient and quick way to study the parameter sensitivities, we also conducted a similar study using empirical data collected by a real sensor. Data collected by the HYDICE [11] airborne imaging spectrometer over forest and desert backgrounds, and similar objects as the model case, were analyzed. The various SQE's (one for each target/background pair) were derived from the detection results in a manner analogous to the model-based approach.

Several constraints of the empirical data restricted us from performing identical analyses to the model-based approach. In particular, the empirical data have sensor noise, artifacts, and residual calibration errors, which set an upper bound on the SNR. Since we add random noise to the data when performing the SNR sensitivity analyses, we limited the range of SNR studied to 25 – 100. In addition, the finite number of background samples in each image limited the false alarm rate used in the constant performance criterion to 10^{-4} . Equation 2 presents the empirical based equation while equation 3 contains another model-based one constrained to the same SNR dynamic range and false alarm rate.

$$SQRS_{empirical_detection} = 9.70 - 3.32 \log_{10}[GRD(cm)] + 0.67 \log_{10}[SNR] + 0.48 \log_{10}[N] \quad (2)$$

$$SQRS_{model_detection} = 9.91 - 3.23 \log_{10}[GRD(cm)] + 0.66 \log_{10}[SNR] + 0.60 \log_{10}[N] \quad (3)$$

Equations 2 and 3 show similar relative values of the coefficients demonstrating a nice consistency between the model-based and empirical-based analyses. The differences between these equations and equation 1 which was developed using results over a larger range of parameter values demonstrate the “softness,” of this approach and raise a flag of caution to not use these equations verbatim. Clearly, this is an area in need of continued active research. However, these results do demonstrate the viability of the overall approach of defining contours of constant performance and fitting functions to describe the parameter dependencies.

4. TERRAIN CLASSIFICATION SPECTRAL QUALITY

The same HYDICE data sets used for the previously reported detection analysis was used for the terrain classification study, but with all manmade objects masked out. Images of a forested area and a desert area were selected for analysis.

In preparation for the parameter trade study of impacts on classification accuracy, the following approach was developed to establish the “ground truth,„. Rather than have an analyst define limited training and testing areas, an unsupervised algorithmic approach was adopted to assign each pixel to one of six terrain cover classes. First, the data were transformed using Principal Components and the first 20 PC’s used in an ISODATA clustering step to produce a first cut at the six classes. This was followed by two iterations of a Gaussian Maximum Likelihood classifier to refine the class assignments.

As a way of examining the sensitivity to classification algorithm, we used two computational simple and well known spectral vector distance metrics to assign each pixel to a class in the trade study: Euclidean minimum distance (EMD) and spectral angle mapper (SAM). For each algorithm, the class mean spectra was used to compute the distance metric for each pixel and then the assignment made to the class closest to the mean vector.

In order to have a single scalar as a measure of performance for the classification results, we adopted the Kappa statistic [12] as our metric. The Kappa statistic combines the entries of the standard confusion matrix and is based on the difference between the actual agreement in the confusion matrix and the chance agreement. Figure 2 presents the technique for its computation and a sample interpretation [13].

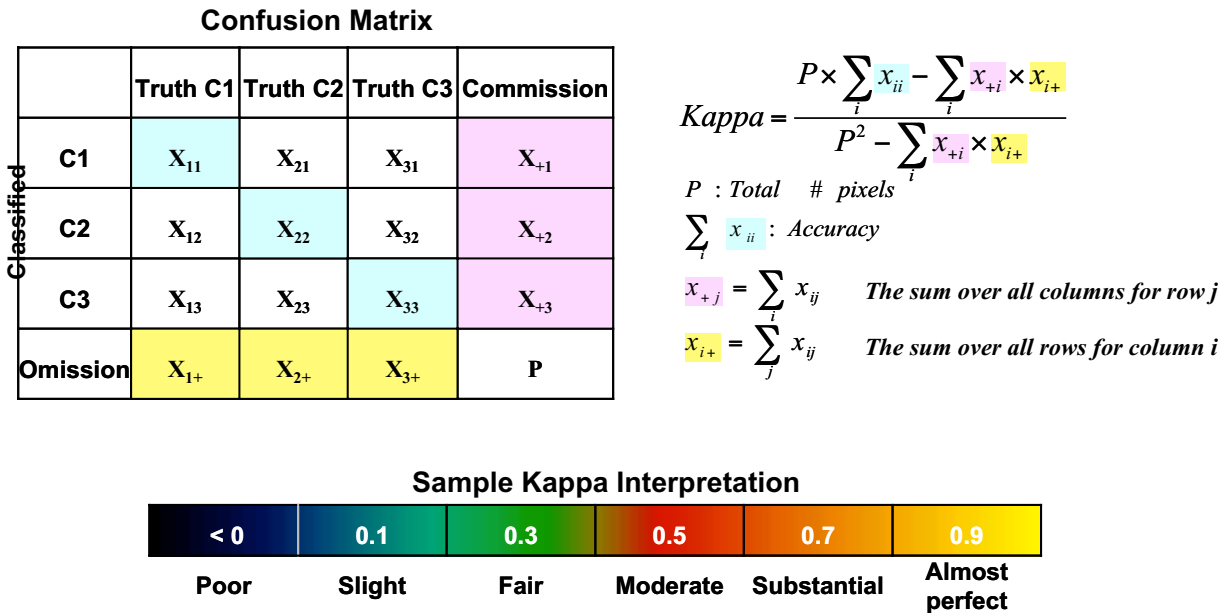


Figure 2. Description of the computation and an interpretation of the Kappa statistic.

To develop the quantitative relationships among the parameters of interest, a large trade study was conducted using the Forest and Desert images and a range of parameter values. The classification algorithms were applied and the Kappa statistic estimated for each parameter combination used in the trade study. Four levels of SNR (25, 50, 75, 100) were studied by adding random noise to each pixel with a standard deviation equal to the signal level divided by the specified SNR. Four spectral resolutions were studied by aggregating adjacent spectral channels by factors of two (120, 60, 30, 15 channels). Five spatial resolutions (1, 2, 3, 4, 5 meters) were studied by aggregating spatial pixels. Since the spectral classes were not always spatially adjacent, the pixels from each class were realigned to form contiguous spatial regions so that the spatial aggregation would occur within each class and ensure the maximum number of homogenous (single class) pixels at the various spatial resolutions.

Figure 3 presents an example set of results showing the surfaces of constant Kappa (at three values) for the EMD algorithm applied to the Forest data set. The surfaces appear to be relatively planar, but with little sensitivity to GRD (spatial resolution), except in conjunction with the Kappa value. That is, we see the well known relationship that classification accuracy (represented by the Kappa statistic) is inversely proportional to spatial resolution.

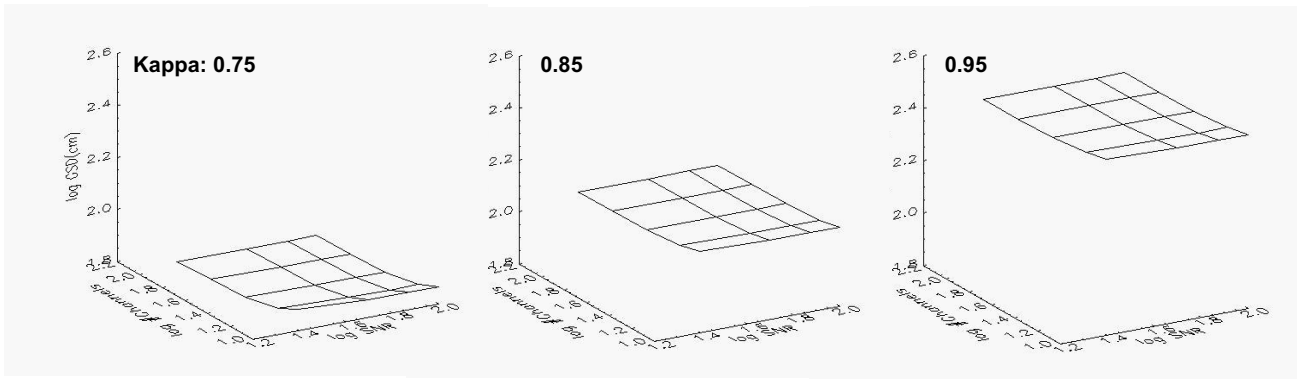


Figure 3. Constant Kappa surfaces at three Kappa values for EMD applied to the Forest scene.

This analysis was repeated for the Forest scene with the SAM algorithm, and then for the Desert with both EMD and SAM. These results were then fit to functions relating the Kappa statistic to the three parameters studied (GRD, SNR, and N) in a manner analogous to approach taken for the unresolved object detection analysis. Equations 4 through 7 show these functions. Note that these equations show the relationship between the sensor parameters and achievable Kappa statistic. Also, note that the equations are scaled to predict a value 10 times the normal Kappa statistic to achieve a 0 to 10 dynamic range.

$$Kappa_{EMD_Forest} \times 10 = 1.25 + 3.32 \log_{10}[GRD(cm)] + 0.04 \log_{10}[SNR] + 0.15 \log_{10}[N] \quad (4)$$

$$Kappa_{SAM_Forest} \times 10 = 0.20 + 3.52 \log_{10}[GRD(cm)] + 0.28 \log_{10}[SNR] + 0.21 \log_{10}[N] \quad (5)$$

$$Kappa_{EMD_Desert} \times 10 = 6.66 + 1.32 \log_{10}[GRD(cm)] + 0.17 \log_{10}[SNR] + 0.06 \log_{10}[N] \quad (6)$$

$$Kappa_{SAM_Desert} \times 10 = -4.76 + 2.80 \log_{10}[GRD(cm)] + 0.81 \log_{10}[SNR] + 0.38 \log_{10}[N] \quad (7)$$

One issue in producing these equations is that the resulting dynamic range of classification accuracy (and Kappa statistic) for the different scenes, algorithms, and trade study parameter values varied significantly. Some classes were particularly separable with some algorithms, and some not. Thus, the above equations were developed using different ranges of Kappa. Table 2 shows the ranges resulted during the study.

Table 2. Kappa values achieved over the range of the trade study for the various scenes and algorithms.

Scene/Algorithm	Kappa statistics achieved
Forest/EMD	0.75, 0.80, 0.85, 0.90, 0.95
Forest/SAM	0.75, 0.80, 0.85, 0.90, 0.95
Desert/EMD	0.95, 0.96, 0.97, 0.98, 0.99
Desert/SAM	0.30, 0.35, 0.40, 0.45, 0.50

Since there were different ranges of Kappa statistics from which the equations were developed, we should not expect them to necessarily be the same. This is particularly true for the offset constant since it defines the origin for that case. However, we do see a reasonable amount of consistency in the relative values of the coefficients.

5. DISCUSSION

Examination of equations 1-3 and 4-7 leads to some interesting observations. First, it should be pointed out they calculate two different, but related quantities. Equations 1-3 predict a SQRS value that is related to how small an object can be detected. Equations 4-7 predict the Kappa statistic that is related to overall classification accuracy. Both cover an approximate 0-10 range, with higher values indicating a higher level of performance.

Second, equations 1-3 and 4,5, and 7 have a similar magnitude for the coefficient of the $\log[GRD]$ term, although the sign is opposite between the detection and classification cases. As mentioned earlier, this reflects the intuitive

understanding that the smaller the pixel size, the easier it is to detect small objects, but larger pixels lead to higher classification accuracies through an averaging effect.

In considering a spectral quality metric and a scale that is unified across applications, these opposite trends with spatial resolution would have to be reconciled. One way is to use a single equation to calculate the metric, but then have the interpretation of the level be different depending on the application. Of course, this would have to be the case always since the descriptions would be different, but even to the extent of having a reversed trend with better performance in the classification application as the descriptor for lower scale values.

An alternative approach would be to have separate equations for each application, but allow the scale to be monotonic with performance. That is, as one goes up the spectral quality scale, the performance (object detection capability, classification accuracy, etc.) increases. Ultimately, this is likely to be the desirable case because it is reasonable to expect that the various applications of spectral imagery will have different sensitivities to sensor and scenario parameters.

Finally, it is interesting to note that the relative values between the SNR and number of channels, and the GRD coefficients are similar for both sets of equations, indicating performance is mostly driven by the spatial resolution for classification as well as detection applications.

6. SUMMARY AND FUTURE WORK

In summary, we have presented a continuation of efforts to capture the relative tradeoffs in application performance from sensing parameters with the goal of developing candidate spectral image quality metrics. Our previous work in applying these ideas to unresolved object detection was reviewed and extended to apply to the terrain characterization problem. We observed parameter relationships and dependencies with similar magnitudes between the two applications, but with the opposite dependence on spatial resolution. That is, detection performance increases with increasing resolution, but classification accuracy decreases with increasing resolution. Finally, we suggest that ultimately a set of equations will likely be necessary to characterize spectral image quality, but with a common scale that follows intuition with higher numerical values corresponding to higher quality.

Future work in this area involves extending these ideas to material identification applications and to explore additional data sets (and model analyses) to investigate the robustness of these initial formulations of spectral image quality metrics.

ACKNOWLEDGMENT

This work was performed while Dr. Kerekes was at MIT Lincoln Laboratory.

REFERENCES

1. J. Kerekes and S. Hsu, "Spectral Quality Metrics for VNIR and SWIR Hyperspectral Imagery," *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, SPIE Vol. 5425, 2004.
2. L. Martin, J. Vrabel, J. Leachtenauer, "Metrics for Assessment of Hyperspectral Image Quality and Utility," *Proceedings of International Symposium on Spectral Sensing Research*, 1999.
3. L. Bubion, K. Bennett, J. Bowles, and L. Martin, "Image Quality Evaluation of AOTF and Grating Based Hyperspectral Imaging Sensors," *Proceedings of International Symposium on Spectral Sensing Research*, 1999.
4. J. Sweet, J. Granahan, and M. Sharp, "An Objective Standard for Hyperspectral Image Quality," *Proceedings of AVIRIS Workshop*, Jet Propulsion Laboratory, Pasadena, California, 2000.
5. S. Shen, "Spectral Quality Equation Relating Collection Parameters to Object/Anomaly Detection Performance," *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, SPIE Vol. 5093, 2003.
6. J. Su, S. Hsu, and S. Orloff, "Assessment of Effects of Lossy Compression of Hyperspectral Image Data," *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, SPIE Vol. 5425, 2004.

7. P. Keranen, A. Kaarna, and P. Toivanen,,Spectral Similarity Measures for Classification in Lossy Compression of Hyperspectral Images,, *Proceedings of Image and Signal Processing for Remote Sensing VIII*, SPIE Vol. 4885, 2003.
8. J. Leachtenauer, "National Imagery Interpretability Rating Scales: Overview and Product Description,, *ASPRS/ACSM Annual Convention and Exhibition Technical Papers*, Vol. 1, pp. 262-272, 1996.
9. J. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "The General Image Quality Equation,, *Applied Optics*, pp. 8322-8328, 10 November 1997.
10. J. Kerekes and J. Baum, "Spectral Imaging System Analytical Model for Subpixel Object Detection,, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 5, pp. 1088-1101, May 2002.
11. L. Rickard, R. Basedow, E. Zalewski, P. Silvergate, and M. Landers, "HYDICE: An Airborne System for Hyperspectral Imaging,, *Proceedings of Imaging Spectrometry of the Terrestrial Environment*, SPIE Vol. 1937, pp. 173-179, 1993.
12. R. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis, 1999.
13. J. Landis and G. Koch, "The measurement of observer agreement for categorical data,, *Biometrics*, 33, p.159, 1977.