

Comparisons between spectral quality metrics and analyst performance in hyperspectral target detection

John P. Kerekes^{*a}, David W. Messinger^a, Paul Lee^a, Rulon E. Simmons^b

^aChester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY

^bITT Industries Space Systems, LLC, Rochester, NY

ABSTRACT

Quantitative methods to assess or predict the quality of a spectral image continue to be the subject of a number of current research activities. An accepted methodology would be highly desirable for use in data collection tasking or data archive searching in ways analogous to the current prediction of panchromatic image quality through the National Imagery Interpretation Rating Scale (NIIRS) using the General Image Quality Equation (GIQE). A number of approaches to the estimation of quality of a spectral image have been published, but most capture only the performance of automated algorithms applied to the spectral data. One recently introduced metric, however, the General Spectral Utility Metric (GSUM), provides for a framework to combine the performance from the spectral aspects together with the spatial aspects. In particular, this framework allows the metric to capture the utility of a spectral image resulting when the human analyst is included in the process. This is important since nearly all hyperspectral imagery analysis procedures include an analyst.

To investigate the relationships between candidate spectral metrics and task performance from volunteer human analysts in conjunction with the automated results, simulated images are generated and processed in a blind test. The performance achieved by the analysts is then compared to predictions made from various spectral quality metrics to determine how well the metrics function.

The task selected is one of finding a specific vehicle in a cluttered environment using a detection map produced from the hyperspectral image along with a panchromatic rendition of the image. Various combinations of spatial resolution, number of spectral bands, and signal-to-noise ratios are investigated as part of the effort.

Keywords: Spectral imaging, spectral quality, target detection, image quality

1. INTRODUCTION

The ability to quantitatively assess the quality and utility of a multispectral or hyperspectral image is desirable for many reasons including instrument comparisons and trade studies, archive image 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.

A previous publication¹ reviewed a number of different approaches to the concept of a quantitative quality metric. The various approaches were described and then compared numerically with automated analysis of airborne hyperspectral imagery in a target detection application. The results of that study were mixed with no single metric emerging as a clear winner, although the General Spectral Utility Metric (GSUM) approach² offered a framework that seemed intuitive and likely to be general. It was also observed that including a term describing the contrast between the target and the background lead to more consistent results across various targets in the detection application.

* kerekes@cis.rit.edu

One limitation of the previous work was that it compared the metric predictions to the results of an automated detection algorithm without consideration of the role of an Image Analyst (IA) in the process. It is increasingly common that spectral images are collected in conjunction with high resolution panchromatic imagery to provide enhanced spatial information and context in the analysis. It is also true that most commonly an analyst is involved with the ultimate interpretation of the data. It is clearly desirable that this role of a human analyst in the process be captured by a spectral metric and indeed that was part of the motivation for the framework of the GSUM.

The study reported here represents an initial attempt at capturing the contributions of the analyst in the empirical analysis of imagery for comparison to various spectral quality metrics. The task remains one of target detection but instead of relying solely on the quantitative results from the application of a detection algorithm to an image, we solicit the evaluations of volunteer analysts in reporting target locations based upon a high resolution panchromatic image in addition to the detection algorithm output. The reports of the analysts are then scored against the truth and these results compared to several spectral quality metric prediction models.

In Section 2 we review the spectral metrics included in this study. Section 3 describes the simulated scene used in the study. Section 4 describes the task assigned to the volunteer analysts and Section 5 presents the analysts results and their comparison to the spectral metrics. Section 6 summarizes our findings and points to possible future directions for this research.

2. SPECTRAL QUALITY METRICS

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.

For this experiment we selected three candidate spectral quality metrics published in the literature and reviewed in our previous publication. The following reviews the details and equations for these selected metrics.

2.2. Kerekes and Hsu³

This work was closely modeled after the NIIRS GIQE in that a number of analyses were conducted with spectral images (or model analyses) of varying quality, followed by the development of regression equations relating the spectral image parameters to analysis task performance.

In this approach model-based trade studies and empirical analyses of matched filter target detection in the reflective VNIR/SWIR spectral region were performed. Regression analysis was then applied to these results and the spectral image parameters deemed to be of most importance to detection performance, resulting in an objective equation for a Spectral Quality Rating Scale (SQRS) value. Slightly different equations were obtained between the model and the empirical analysis, but the relative ordering of images with different parameters remained the same. In our current work, we will use the empirically-derived result for comparison.

$$SQRS_{detection} = 9.70 - 3.32 \log_{10}[GSD(cm)] + 0.67 \log_{10}[SNR] + 0.48 \log_{10}[N] \quad (1)$$

Here, GSD is the ground sample distance, SNR is the traditional signal-to-noise ratio, and N is the number of spectral channels in the spectral region of interest. To normalize the 0-9 values produced by SQRS, we divide by ten so they can be compared to the other metrics which have a 0 to 1 scale.

2.3. Shen⁴

This work also followed the approach of specifying a performance metric, analyzing a large number of images with varying spectral image parameters, and then forming a regression between the metric and parameters. In this case, the metric of interest was detection probability at a specified false alarm rate. For the case of $P_{FA} = 0.001$, the following equation was found to have a good fit.

$$P_D = 6.25 - 0.81 \log_{10}[GSD(m)] + 0.12 \log_{10}[SNR] - 0.20 \log_{10}[\Delta\lambda(nm)] - 2.43 \log_{10}[\sigma_{scene}] \quad (2)$$

In this equation $\Delta\lambda$ is the average spectral resolution of the channels and σ_{scene} is the average across all spectral bands of the standard deviation of the pixels in the scene. The units for σ_{scene} were the HYDICE scaled radiance units which are equal to $4/3 \mu\text{W}/\text{cm}^2\text{-sr-}\mu\text{m}$. The coefficients were found to vary modestly for other false alarm rates, mostly affecting the constant term as an offset. This formulation has the advantage of considering the scene variability through the σ_{scene} parameter.

2.4. Simmons, et al²

This general solution to spectral utility prediction is based on two concepts: (1) the transforming of spatial information into a confidence value, and (2) the numerical combining of spatial and spectral confidence values into a single confidence number.

Image analyst confidence values have been related to NIIRS for specific problems called Essential Elements of Information (EEIs). A typical plot of NIIRS versus confidence for a given EEI is a sigmoid-shaped curve, with the curve being different for each EEI. In practice, the relationship between NIIRS and confidence is largely driven by the size of the target relative to the image resolution. In equation form these curves are given as:

$$C_{Spatial} = \frac{(N / N_{50})^E}{1 + (N / N_{50})^E} \quad (3)$$

where:

$$E = 2.7 + 0.7(N/N_{50}),$$

N = Number of resolution cycles per minimum dimension of target,
 N_{50} = Cycle criteria for 50 percent success,

and N_{50} has the following values for detection, recognition, and identification: 1.0 ± 0.25 , 4.1 ± 0.35 , and 6.4 ± 1.5 .

The spectral confidence ($C_{Spectral}$) can be found through an assessment of the separability of target and background spectral distributions or from results of hyperspectral image analysis techniques such as a spectral matched filter. For this work, we used the average of normalized matched filter values (the matched filter output divided by the maximum) for the spectral confidence of a given image. In general, it may be that results from spectral algorithms will require a weighting factor to convert them to a confidence value.

Given the spatial and spectral confidences it is desirable for them to be combined. This is not simply an additive process, since if you have a confidence of 1.0 (i.e., 100%) from either the spatial or spectral information, you are totally confident of your answer regardless of any further information from the other side. Ideas from development of "semantic transformations" lead to the method shown in (4).

$$C_{Total} = 1 - (1 - C_{Spatial}) \cdot (1 - C_{Spectral}) \quad (4)$$

Note that while a spectral image simultaneously gives both spectral and spatial information, this equation is applicable to cases where the information is not coincident in time or space such as fusing information from a high resolution panchromatic image with a lower resolution spectral image taken simultaneously or even at a different time.

3. SCENE AND IMAGERY DESCRIPTION

Ideally, one would like to use real spectral imagery collected under many conditions in order to understand the trends and parameter sensitivities in determining the quality associated with each image. However, empirical data collections can be quite expensive and time consuming and it is difficult to control or even know the true values of scene or sensor parameters. While they ultimately are necessary, the use of physics-based image simulation tools can provide imagery useful in these early stages of spectral quality research. Also, while real images offer limited experimental configurations and often suffer from sensor artifacts, simulated images can be rendered with different characteristics fairly easily and offer artifact-free and completely known data useful for these type of studies.

RIT's Digital Imaging and Remote Sensing laboratory's Image Generation (DIRSIG) model⁵ was used to simulate the images under a variety of conditions. The next sections briefly describe this model and the scene selected for analysis.

3.1. DIRSIG

The DIRSIG model produces radiometrically accurate high resolution spectral imagery spanning the visible through longwave infrared spectral regions. The process of simulating a spectral image starts with a geometric database describing the world in the form of faceted surfaces making up the terrain and objects on the surface. These faceted surfaces are characterized by material types having optical and thermodynamic properties as described in a material database. The geometry of surface objects is described using common Computer Aided Design (CAD) tools. The geometric data are sampled using rays cast by a ray tracing routine under direction of the sensor model which controls the fields of view, sampling pattern, etc. based on the sensor characteristics and the platform flight profile. The ray tracer is also used to determine the sun-target-sensor radiation paths, as well as the sources of reflected and self-emitted background radiation bi-directionally reflected toward the sensor. All of the target, path and background data are passed to the radiometry model which accounts for all sources of radiometric flux reaching the sensor using the MODTRAN⁶ atmospheric model. MODTRAN also defines the transmission through the atmosphere and upwelled/downwelled radiation emitted and scattered by the atmosphere. The spectral radiance reaching the sensor is multiplied by the sensor's spectral response functions for each band and integrated to yield the in-band radiance. This sensor radiance field is then convolved with the spatial point spread function of the imaging system and noise added to yield the final image.

3.2. Scene description

The scene selected as the background for this experiment is a portion (Tile 1) of the large DIRSIG scene known as MegaScene I. The MegaScene project⁷ produced synthetic images modeled on an area northeast of Rochester, New York. This scene was designed to be capable of being rendered at approximately 1 meter resolution. The scene covers a residential area including a school and its ball fields and encompasses an area 800 meters by 800 meters.

To this baseline scene we added 26 vehicles representing three types (SUV, pickup and hatchback) with reflectance spectra corresponding to seven different types of vehicle paints. Eleven vehicles were white SUV's that were designated as the targets of interest. Two vehicles (one hatchback, one pickup) were also assigned the same white paint as the targets, but were meant as "decoys" with their slightly smaller size. The other thirteen vehicles were a mix of the vehicle types and paint spectra from the six other colors (two blue, two green, one gray and one black). The vehicles were distributed between parking lots and roads, with some clearly in the open and some nearly hidden by tree cover.

3.3. Simulated images

A total of 27 images were rendered spanning combinations of spectral resolution, spatial resolution, and signal-to-noise ratio. Nine panchromatic images were simulated at 0.5m, 1.0m, and 2.0m ground resolutions, each at SNR's equal to 10, 100, and 1000. Nine multispectral images were simulated at 2.0m, 4.0m, and 8.0m resolution with each also at SNR=10, 100, and 1000. The multispectral images contained five spectral bands corresponding to the first five bands of the Landsat Thematic Mapper⁸. Nine hyperspectral images were simulated at the same spatial resolution and SNR as the multispectral images. The hyperspectral images contained 210 bands from 400 to 2500 nm corresponding to the HYDICE⁹ instrument spectral characteristics. Figure 1 shows the 0.5m, 1.0m and 2.0m (top to bottom) panchromatic and the 2.0m, 4.0m and 8.0m (top to bottom) hyperspectral images.

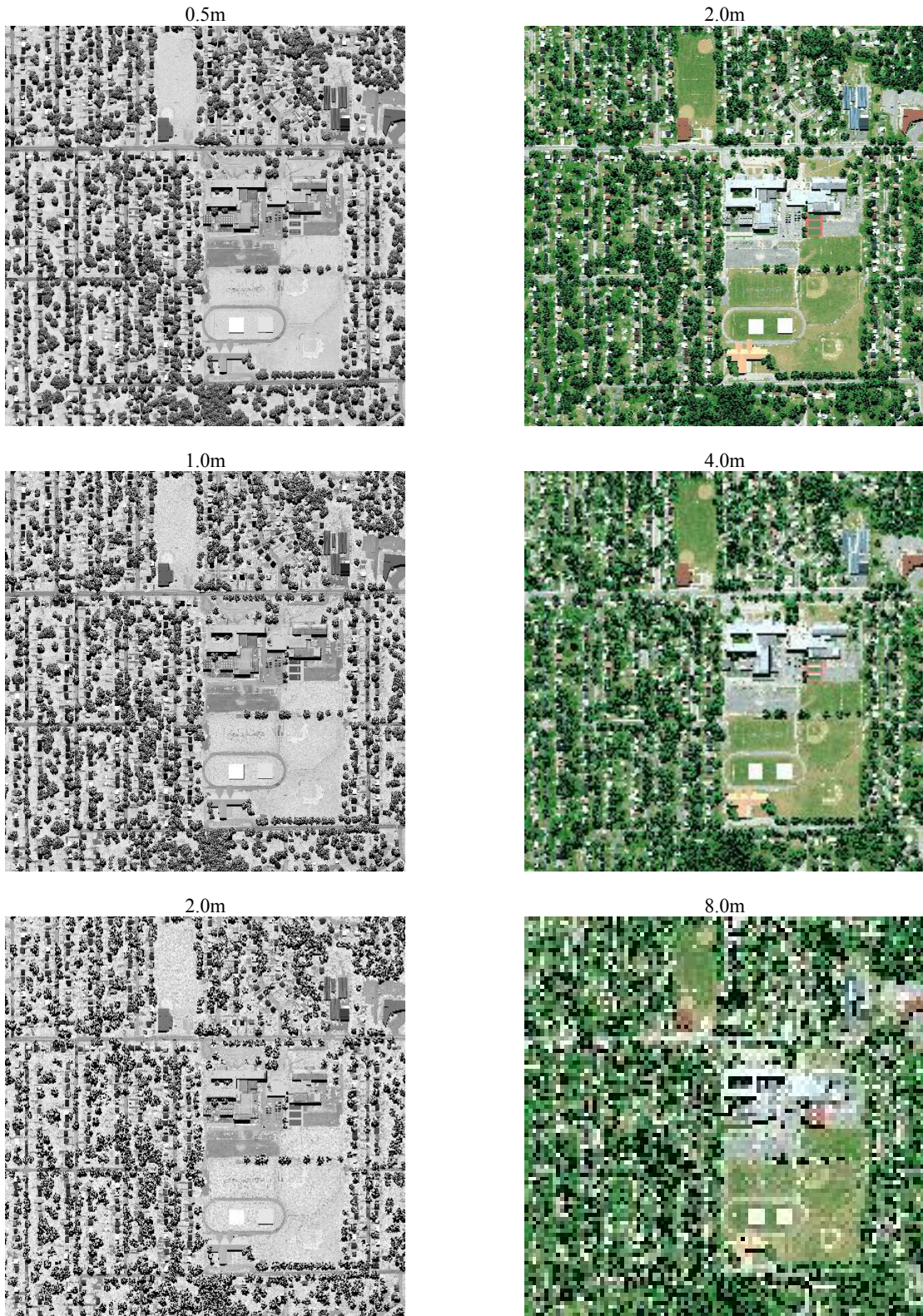


Figure 1. Simulated images used in the study. Panchromatic on the left and RGB's of the spectral images on the right.

3.4. Image processing

Rather than let the analysts process the spectral imagery themselves introducing additional variation in the results, the images were first processed in a uniform manner. Each of the spectral images were atmospherically compensated using the empirical line method and the two large gray-level panels deployed in the middle of the running track in the scene. The simple matched filter available in ENVI was then applied using the vehicle white paint as the target spectrum. In applying the matched filter bands with low signal in the hyperspectral image were avoided resulting in 144 of the 210 bands used in the filter. All five bands of the multispectral image were used in the matched filter.

4. IMAGE ANALYSIS TASK

The analysts were given the task of finding white SUV's in the imagery. For each combination of GSD, SNR, and spectral image type, the analysts were provided the grayscale panchromatic image and the matched filter output file derived from the spectral image, both in ENVI format. Figure 2 provides an example of the image pairs provided to the analysts. The analysts were told the parameters (MSI or HSI, SNR, and GSD) of the images to aid in their assessment.

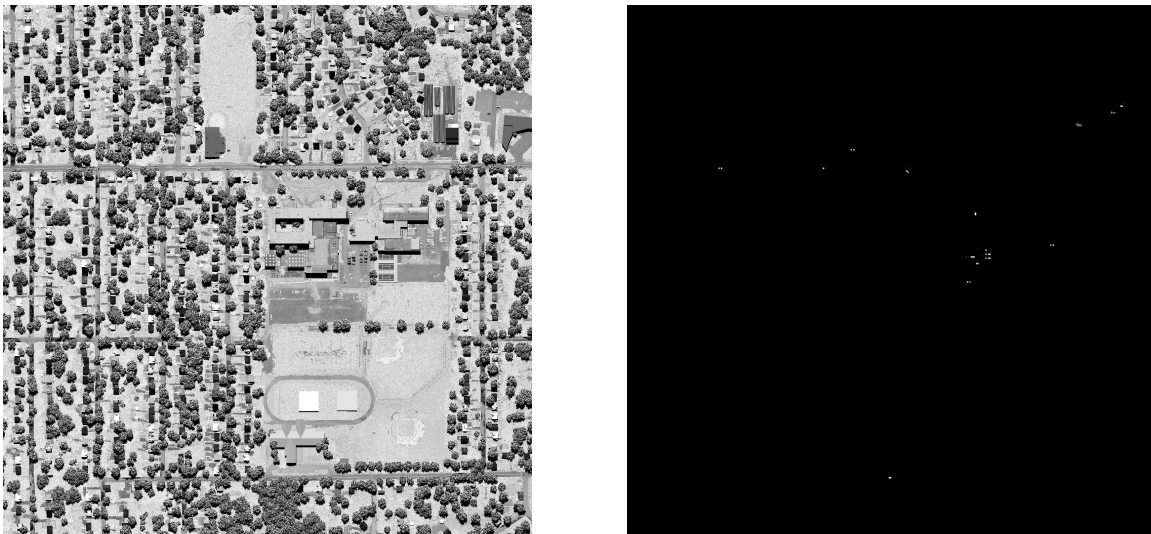


Figure 2. Example panchromatic and matched filter output image provided to the analysts.

To report the results in a uniform manner, the analysts were given a summary sheet to complete for each image pair. On the sheet they were asked to list the pixel location (in the spectral image) and their level of confidence (1 = lowest, 5 = highest) for each suspected white SUV. They were asked to list no more than twenty entries, and were not told how many vehicles were present, nor if there were any decoys in the imagery. The eighteen image pairs were analyzed by three volunteer analysts with six pairs assigned to each analyst.

5. ANALYST RESULTS AND COMPARISON TO METRICS

The analyst reports were compared to the known truth locations of the white SUV's in the imagery. If the analyst reported a target within one pixel of the known location it was scored as a correct decision. If there was no target within one pixel of the reported location it was scored as a false alarm only if the analyst assigned it a level of confidence higher than 1. Table 1 contains a summary of the results for each of the eighteen pairs of images. Note that the 2m spectral images contained $400 \times 400 = 160,000$ pixels, the 4m images contained $200 \times 200 = 40,000$ pixels and the 8m images contained $100 \times 100 = 10,000$ pixels.

Table 1. Volunteer analyst results for image pairs.

GSD (Pan/Spectral)	Signal-to-Noise Ratio	Spectral Image Type	Number of Correct Decisions (out of 11)	Number of False Alarms
0.5m/2m	1000	Hyperspectral	9	2
0.5m/2m	100	Hyperspectral	8	4
0.5m/2m	10	Hyperspectral	10	3
1m/4m	1000	Hyperspectral	3	3
1m/4m	100	Hyperspectral	3	0
1m/4m	10	Hyperspectral	3	0
2m/8m	1000	Hyperspectral	1	4
2m/8m	100	Hyperspectral	1	0
2m/8m	10	Hyperspectral	0	6
0.5m/2m	1000	Multispectral	6	1
0.5m/2m	100	Multispectral	8	5
0.5m/2m	10	Multispectral	9	2
1m/4m	1000	Multispectral	2	1
1m/4m	100	Multispectral	3	0
1m/4m	10	Multispectral	3	2
2m/8m	1000	Multispectral	0	7
2m/8m	100	Multispectral	1	5
2m/8m	10	Multispectral	0	6

Estimates of the detection probability (P_D) and the false alarm probability (P_{FA}) were made and are plotted for all image pairs in Figures 3 through 5 to show the sensitivity of these values to the various parameters studied. The only clear result was the sensitivity to spatial resolution with the results for the smaller GSD images clustering in the upper part of the P_D / P_{FA} plots, the middle GSD results in the middle, and the larger GSD results in the lower part.

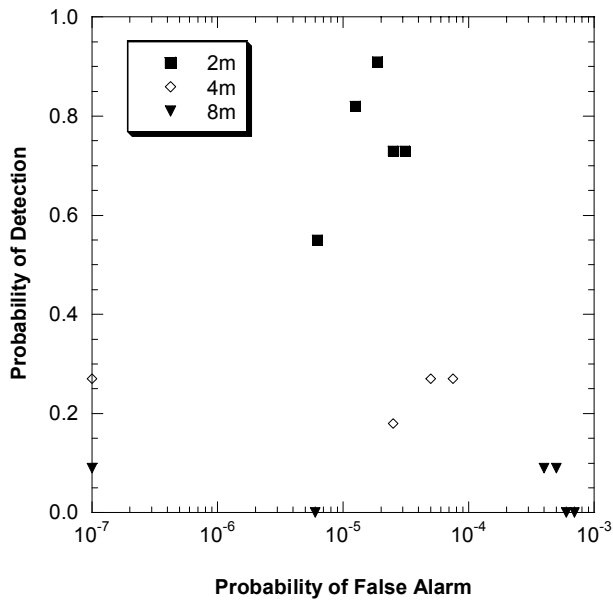


Figure 3. P_D/P_{FA} dependence on GSD.

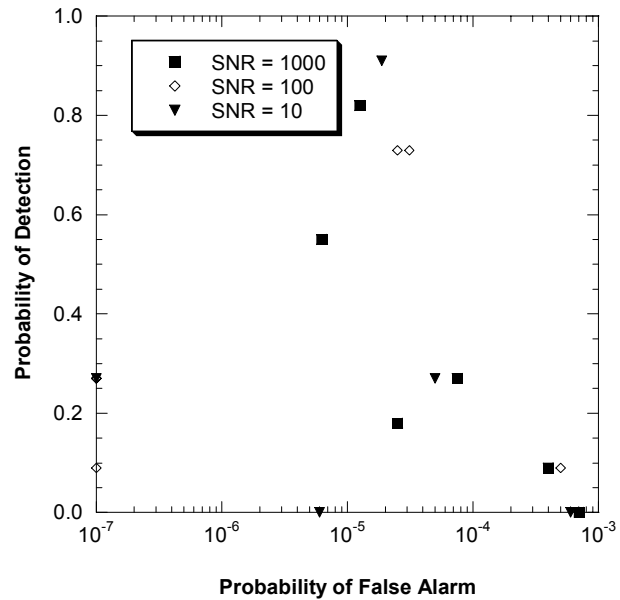


Figure 4. P_D/P_{FA} dependence on SNR.

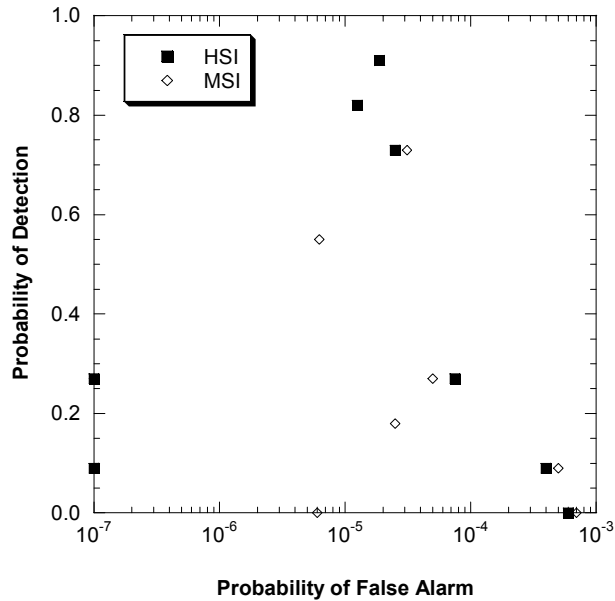


Figure 5. P_D/P_{FA} dependence on spectral image type.

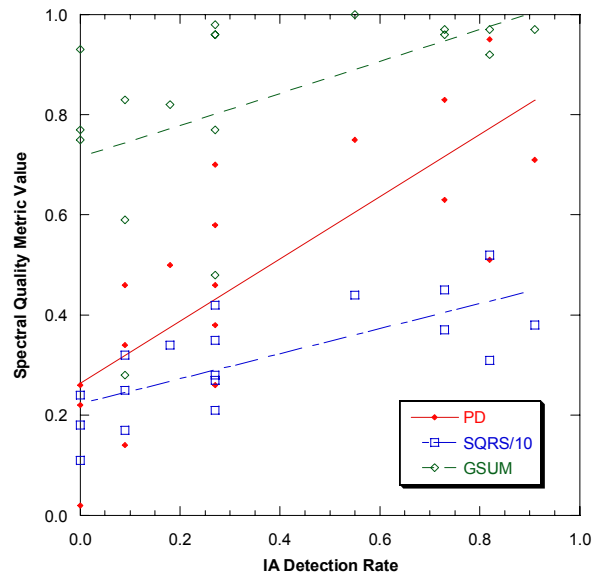


Figure 6. Relationship between IA results and spectral metrics.

The detection results of the volunteer IA's were compared with spectral quality metric predictions made from the three candidate metrics considered. These comparisons are plotted in Figure 6 as scatter plots with the predicted metric based on the spectral image parameters and equations in Section 2 paired with the detection probability obtained from analysis of the analyst reports. The PD parameter σ_{scene} value was relatively constant for the various images and a single value of 26 was used for all PD calculations. Also, for the particular parameter range considered in this study, we needed to modify the constant term in Eq. 2 from 6.25 to 4.47. This allowed the value of the PD metric to lie between 0 and 1. The spatial confidence portion of GSUM was calculated based on the target size and the panchromatic image GSD (with the recognition value for N_{50}) and the spectral confidence was calculated based on the normalized matched filter output (values divided by the maximum in the image) averaged for all declared detections, whether they were correct detections or false alarms. Note that this comparison ignores the false alarm rate differences between the IA results; given that only few false alarms were generated by the analysts, this assumption that the detection rate is relatively independent of the false alarm rate is not unreasonable. Also, since the analysts only provided target reports with no knowledge of the true targets, there was no data from which to adjust the detection rate for constant false alarm rates. This relatively minor inconsistency between the IA results and the SQM's is a limitation of this experimental approach.

Best fit lines were calculated to the scatter plots shown in Figure 6 with the resulting correlation coefficient R equal to 0.78 for the PD metric, 0.72 for the SQRS metric and 0.51 for the GSUM metric. The lower R value for the GSUM approach was mainly the result of high spectral confidence scores in the coarse resolution images resulting from false alarms. It is hard to draw any significant conclusions from this limited sample and the loose correlations, but the general trends between the metric predictions and the achieved analyst results are seen to be consistent.

6. SUMMARY AND FUTURE RESEARCH DIRECTIONS

An experiment was conducted to bring the cognitive ability of analysts into the empirical investigation of metrics for spectral image quality. Simulated images of an urban area containing unknown target vehicles were generated for a variety of spatial resolutions, spectral resolutions and signal-to-noise ratios. These spectral images were processed in a uniform manner and the results provided along with a higher resolution panchromatic context image to volunteer image analysts. These analysts then reported locations where they suspected the target vehicles were located. These reported locations were scored as correct detections or false alarms and these results were compared to several spectral quality metrics calculated from parameters describing the spectral images.

The results confirmed the dominant influence of spatial resolution on the quality of spectral imagery. While the trends were generally consistent between the IA results and the metric predictions, there was only a loose correlation (R values between 0.51 and 0.78) indicating the metrics are not capturing all of the necessary characteristics of the spectral imagery.

While the significance of these results is limited by the small scale of this experiment, the study does illustrate an approach to use image simulation to study the driving parameters of spectral image quality. In the future it would be desirable to repeat this type of experiment but with many different target configurations and a larger sensor parameter trade space.

ACKNOWLEDGMENT

The authors would like to thank the volunteer analysts for their contributions to this effort.

REFERENCES

1. J. Kerekes, A. Cisz and R. Simmons, "A Comparative Evaluation of Spectral Quality Metrics for Hyperspectral Imagery," *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, SPIE Vol. 5806, 2005.
2. R. Simmons, T. Elder, D. Stewart, E. Cincotta, C. Kennedy, and R. Craig Van Nostrand, "General Spectral Utility Metric for Spectral Imagery," *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, SPIE Vol. 5806, 2005.
3. 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.
4. 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.
5. J. Schott, S. Brown, R. Raqueño, H. Gross, and G. Robinson, "An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development", *Canadian Journal of Remote Sensing*, Vol. 25, No. 2, June 1999
6. A. Berk, L. Bernstein, G. Anderson, P. Acharya, D. Robertson, J. Chetwynd and S. Adler-Golden, "MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS," *Remote Sensing of Environment*, vol. 65, pp. 367-375, 1998.
7. E. Ientilucci and S. Brown, "Advances in wide-area hyperspectral image simulation," *Proceedings of Targets and Backgrounds IX: Characterization and Representation*, SPIE Vol. 5075, pp. 110-121, 2003.
8. USGS Landsat Thematic Mapper Data, http://edc.usgs.gov/guides/landsat_tm.html, 2006.
9. L. Rickard, R. Basedow, E. Zalewski, P. Silverglate, 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.