

# Methodology For Evaluating Statistically Predicted Versus Measured Imagery

<sup>1</sup>Rob Kooper, <sup>1</sup>Peter Bajcsy and <sup>2</sup>Dennis Andersh

<sup>1</sup>National Center for Supercomputing Applications (NCSA)

University of Illinois at Urbana-Champaign, IL

and

<sup>2</sup> Scientific Applications International Corporation (SAIC), Champaign, IL

Email: [kooper@ncsa.uiuc.edu](mailto:kooper@ncsa.uiuc.edu), [pbajcsy@ncsa.uiuc.edu](mailto:pbajcsy@ncsa.uiuc.edu), [dennis.j.andersh@SAIC.COM](mailto:dennis.j.andersh@SAIC.COM)

## ABSTRACT

We present a novel methodology for evaluating statistically predicted versus measured multi-modal imagery, such as Synthetic Aperture Radar (SAR), Electro-Optical (EO), Multi-Spectral (MS) and Hyper-Spectral (HS) modalities. While several scene modeling approaches have been proposed in the past for multi-modal image predictions, the problem of evaluating synthetic and measured images has remained an open issue. Although analytical prediction models would be appropriate for accuracy evaluations of man-made objects, for example, SAR target modeling based on Xpatch, the analytical models cannot be applied to prediction evaluation of natural scenes because of their randomness and high geometrical complexity imaged by any of the aforementioned sensor modality. Thus, statistical prediction models are frequently chosen as more appropriate scene modeling approaches and there is a need to evaluate the accuracy of statistically predicted versus measured imagery. This problem poses challenges in terms of selecting quantitative and qualitative evaluation techniques, and establishing a methodology for systematic comparisons of synthetic and measured images. In this work, we demonstrate clutter accuracy evaluations for modified measured and predicted synthetic images with statistically modeled clutter. We show experimental results for color (red, green and blue) and HS imaging modalities, and for statistical clutter models using Johnson's family of probability distribution functions (PDFs). The methodology includes several evaluation techniques for comparing image samples and their similarity, image histograms, statistical central moments, and estimated probability distribution functions (PDFs). Particularly, we assess correlation, histogram, chi-squared, pixel and PDF parameter based error metrics quantitatively, and relate them to a human visual perception of predicted image quality. The work is directly applicable to multi-sensor phenomenology modeling for exploitation, recognition and identification.

**Keywords:** synthetic image evaluation, statistical multi-sensor phenomenology modeling, hyperspectral imagery.

## 1 INTRODUCTION

The problem of multi-sensor phenomenology modeling for exploitation, recognition and identification includes several fundamental theoretical, experimental and validation issues. In this paper, we focus only on validation issues in the context of multi-sensor phenomenology modeling. The objective of multi-sensor modeling, as defined in this work, is to predict image appearance (or pixel values) based on (a) previous knowledge about viewed scene and objects, (b) existing data (saved measurements), and (c) developed prediction models. The goal of a validation component in the process of multi-sensor modeling is to assess the goodness of image predictions based on modeling or application criteria.

In order to design a validation methodology, one has to understand other components of the multi-sensor phenomenology modeling, such as data dimensionality and type, sensor modality and prediction models. In our work, we consider imaging sensors that form a 2D raster image (or a grid of measurements). The imaging modalities could include several known sensors, such as Synthetic Aperture Radar (SAR), Electro-Optical (EO), Multi-Spectral (MS) and Hyper-Spectral (HS) modalities. We primarily focus on color (red, green and blue denoted as RGB) and HS imaging modalities.

While several scene modeling approaches have been proposed in the past for single-modal image predictions<sup>1, 3, 6, 7, 8</sup>, the problem of evaluating synthetic and measured images has remained an open issue. Although analytical prediction models would be appropriate for accuracy evaluations of object models, for example, SAR target modeling based on Xpatch<sup>2</sup> or thermal IR target modeling based on MuSES<sup>4</sup>, the analytical models cannot be applied to prediction

evaluation of scene models because of randomness and high geometrical complexity of natural scenes imaged by any of the multi-modal sensors<sup>7,8</sup>. It is very hard to provide accurate inputs for analytical methods about temporally varying geometry of each scene element, its surface properties, and derive boundary conditions for electro-magnetic waves interacting with scene elements. Thus, statistical prediction models are frequently chosen as scene modeling approaches<sup>30</sup>, and there is a need to evaluate the accuracy of statistically predicted and measured imagery. However, the use of statistical prediction models poses challenges in terms of selecting quantitative and qualitative evaluation techniques, and establishing a methodology for systematic comparisons of synthetic and measured images. The development of such an evaluation methodology became the motivation for our work.

One should also mention the motivation for developing systems based on statistical prediction models or hybrid, statistical and analytical, prediction models. The major advantages of statistical prediction models over other approaches are (a) low computational requirements (fast model building, small model storage, and fast model-based generation), (b) flexibility to cover a wide range of probability distribution functions (PDFs) and (c) applicability of statistical models to a wide range of spectral sensors<sup>30</sup>, e.g., SAR, color, multi-spectral or hyperspectral cameras. Furthermore, statistical models can easily re-use historical data, for example, color databases CUREt, Brodatz or VisTex<sup>27,28,29</sup>, or SAR measurements by Ulaby<sup>14</sup> modeled with Gaussian or Weibull PDF models<sup>6</sup>. In addition, statistical models can interpolate parameters between measurements and thus (a) decrease memory requirements for storing measurements at finer angular increments, and (b) generate image predictions for previously unseen viewpoints more efficiently than by re-computing predicted values with analytical models. Thus, a methodology for evaluating statistically generated and measured imagery becomes an essential part of a computationally efficient, real-time, sensor prediction model where the prediction accuracy has to be proven.

Applications of statistical modeling include real-time sensor scene predictions, as well as forensic image operations, statistical image restoration, efficient modeling for virtual reality environments, and accurate sensor-to-sensor predictions. For example, the accuracy of sensor-to-sensor prediction would have to be evaluated if one would predict scene clutter in color (red, green, blue) imagery using HS information as reported in<sup>31</sup>. One could also compare multiple techniques for image restoration and forensic studies when an object is removed from the scene and the removed pixels are generated using (1) statistical scene clutter model, (2) first principle, electro-magnetic (EM), simulations, or (3) interpolation methods based on Kriging or B-spline models.

In this work, we demonstrate clutter accuracy evaluations for modified measured and predicted synthetic images with statistically modeled clutter. We show experimental results for color (red, green and blue) and HS imaging modalities, and for statistical clutter models using Johnson's family of probability distribution functions (PDFs). The methodology includes several evaluation techniques for comparing image samples and their similarity, image histograms, statistical central moments, and estimated probability distribution functions (PDFs). Particularly, we assess correlation, histogram, chi-squared, pixel and PDF parameter based error metrics quantitatively, and relate them to a human visual perception of predicted image quality. The novelty of this work is in the proposed evaluation methodology and it is directly applicable to multi-sensor phenomenology modeling for exploitation, recognition and identification.

## 2 STATISTICAL MODELS FOR SCENE CLUTTER PREDICTION

**Statistical Models:** In order to build scene clutter prediction model, one has to characterize bi-directional texture functions (BTFs) for every clutter type, and illumination and viewing geometric configuration (azimuth, elevation and distance)<sup>27,15</sup>. In addition, clutter prediction has to incorporate illumination and camera type characteristics. If statistical models are employed then one has to estimate (1) what PDF model matches the measured image samples best, and (2) what parameters of the selected PDF fit the data best. To reduce the size of image data acquired for building scene clutter models, histograms have been proposed in the past<sup>7,8</sup>. Histograms as PDF approximations of measured data can take many different shapes and a single PDF model is usually not flexible enough to accurately model all occurring histogram shapes. One solution would be to use a set of distributions, for instance, Normal, Weibull, Exponential, Rayleigh, Lognormal, etc, assuming that one would know how to decide about the most appropriate PDF model<sup>16</sup>. Another way is to use a family of PDFs that can model multiple shapes. We chose a family of PDFs proposed by Johnson<sup>13</sup>. The Johnson's family of PDFs was explored for modeling HS imagery in<sup>30</sup> and was proven to outperform a set of aforementioned single distributions (see Bajcsy<sup>30</sup>). For mathematical definitions of the Johnson's family of PDFs, we refer the reader to Slipfker<sup>13</sup> and Bajcsy<sup>30</sup>.

**Special Cases of Statistical Distributions:** When dealing with digital data and performing statistical modeling, new issues arise due to the finite accuracy of digital measurements and the possible mismatch between a continuous statistical PDF model and a histogram distribution of discrete digital values. We address next (1) how to model clutters with a small number of distinct values that would not fit well any continuous PDF model and (2) how to treat data with a large percentage of identical values.

First, the case of clutters with a small number of distinct values occurs due to a variable dynamic range per wavelength. While most color cameras generate pre-processed RGB bands with equal dynamic range<sup>33</sup>, HS cameras, like the Opto-Knowledge HS camera<sup>18, 25</sup>, acquire data with a varying dynamic range as a function of the wavelength. For instance, a typical raw value in our HS experiments had a range of 120 to 170 in the lower bands (smaller wavelengths) and a range of 120 to 2100 in the higher bands (larger wavelengths) as shown in Figure 1. Thus, in order to perform accuracy evaluations with methods that require sample binning (e.g., histogram-based error metric) one has to decide on the number of bins, where the number of bins should be proportional to the dynamic range of the raw HS measurements. To be able to use any error metric using sample binning, the bin size has to match the smallest dynamic range found in the original data. In our case of HS imagery acquired by the Opto-Knowledge HS camera<sup>18</sup>, the smallest dynamic range was about 50 unique values resulting in setting the number of bins to 50. If we would have chosen larger number of bins than 50 then we would have observed predicted values in the bins that could not have been populated in the original image due to the dynamic range constraint (but would be populated based on a continuous PDF model).

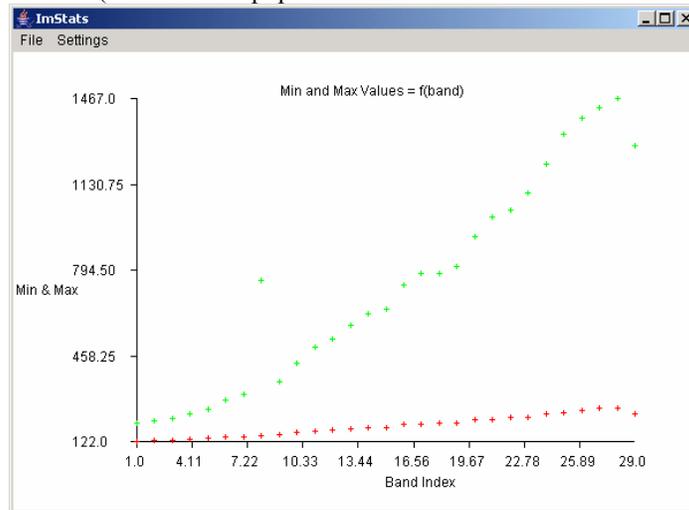


Figure 1: Wavelength dependent dynamic range (Min and Max raw values) measured by acquiring hyperspectral image of a white calibration board with reflectance approximately 97% at EL=45 and AZ=0. The band index one corresponds to 440nm and band index 29 to 720nm

Second, the case of data with a large percentage of identical values creates problems when the Johnson PDF parameters are being estimated based on finding four different values in the data at particular intervals (see Slipfker<sup>13</sup> for more details). If we obtain a dataset with a large number of similar values (i.e. about 40% the same) this parameter estimation method will fail, e.g., the case of black sand. In our implementation, we treat this case by splitting the dataset into subsets and modeling them separately. For example, all zero values in one array are modeled separately from the rest of samples in the case of black sand because the occurrence is very high. Now, we can simply calculate the percentage of values that should be zero vs. those that should be calculated using the Johnson distribution and the Johnson distribution parameters are calculated using the values other than zero.

### 3 EVALUATION METHODOLOGY FOR STATISTICAL IMAGE SIMULATIONS

Given a measured image of clutter, statistical probability distribution (PDF) model and synthetic image of the same clutter, we would like to evaluate the goodness of synthetic predictions with respect to the measured image or with respect to the PDF model that generated the synthetic data. We present an overview of the validation process for the statistical BTF modeling approach in Figure 2. First, we consider for comparisons continuous theoretical PDF models

and discrete samples following the theoretical PDF models. In other words, a comparison of (a) original and estimated PDF models, and (b) measured discrete samples viewed as realizations of a random process ( $X_1$ ), and generated samples according to a PDF model ( $X_2$ ). Second, we focus on comparisons of variables that can be (a) derived analytically, e.g., true central moments from PDF models ( $V_1, V_3$ ), or (b) estimated numerically, e.g., estimated central moments from random samples ( $V_2, V_4$ ). Third, we are interested in comparisons of binned data derived either from PDF models ( $Y_1, Y_3$ ) or from random samples ( $Y_2, Y_4$ ). Figure 2 also shows the types of tests (metrics) one can use to compare pairs of PDF models, ensembles of random samples, statistical descriptors or binned ensembles of random samples. Next, we define the comparative quality metric for the proposed BTF modeling approach as illustrated in Figure 2.

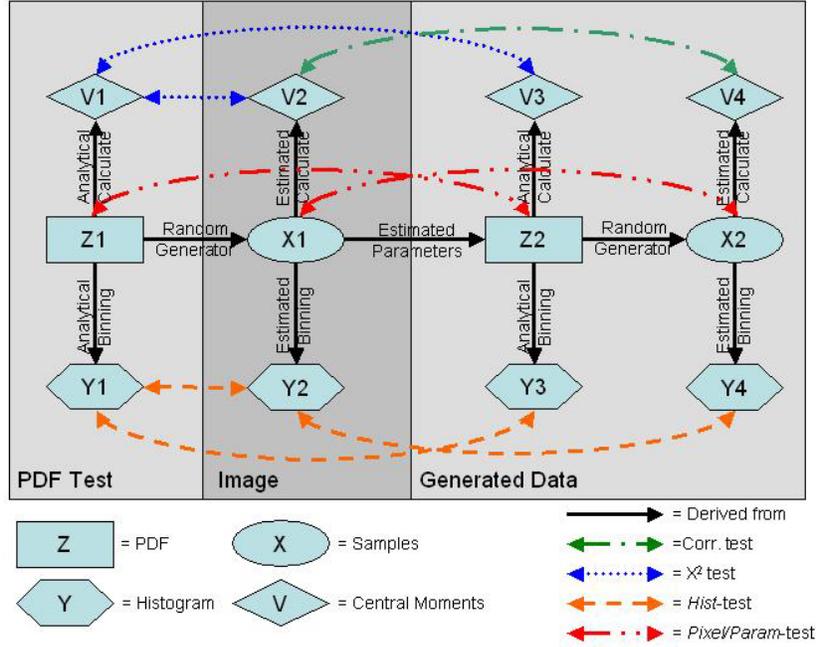


Figure 2: An overview of validation comparisons.

### 3.1 Correlation-Based Comparison

The metric labeled as “Corr.-test” corresponds to the correlation-based comparison. A linear correlation coefficient  $r$  is defined in Equation below, where  $x_i$  and  $y_i$  are original and synthetic (predicted) image values. It is computed over a mask and for each band separately in our experiments. The correlation coefficient error metric  $r$  is also known as the Pearson correlation.

$$r = \frac{\sum_i (x_i - x_{avg})(y_i - y_{avg})}{\sqrt{\sum_i (x_i - x_{avg})^2} \sqrt{\sum_i (y_i - y_{avg})^2}} \quad (1)$$

### 3.2 Histogram-Based Comparison

The metric labeled as “Hist-test” in Figure 2 corresponds to the histogram-based test. A histogram comparison error  $E_{ColorMis}^{Total}$  is the number of mismatched pixels in terms of their color between original and predicted images. The mathematical error definition is provided in Equation below. The metric  $E_{ColorMis}^{Total}$  does not reflect how far apart the

color predictions are from the original colors. Instead, it captures the fact that a pixel was assigned a wrong color. For example, there is no difference between a green pixel and a light blue pixel when the original pixel is blue (one would expect that the light blue color prediction is better than the green color prediction).

$$E_{ColorMis}^{Total} = \frac{1}{2} \frac{1}{N_{orig}} \sum_{i=1}^{numBins} \left\| h_{orig}(i) - h_{synth}(i) \frac{N_{orig}}{N_{synth}} \right\| \quad (2)$$

### 3.3 Chi-square Based Comparison

The metric labeled as “ $\chi^2 - test$ ” denotes the chi-squared-based comparison. The null hypothesis for the chi-squared-based comparison is that the compared distributions are identical. Nonetheless, there are two definitions of the chi-squared test,  $\chi^2$  and  $\chi_{SampleData}^2$ , depending on the type of comparison variables. If tests are performed with a PDF model and random samples then Equation (3) applies. If tests are conducted with two ensembles of random samples then Equation (4) applies.

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - np_i)^2}{np_i} \quad (3)$$

$n$  is the number of samples,  $p_i$  is the probability of a sample falling in the  $i$ -th bin and  $x_i$  is the number of observed samples in the  $i$ -th bin.

$$\chi_{SampleData}^2 = \sum_{i=1}^n \frac{(R_i - S_i)^2}{R_i + S_i} \quad (4)$$

$R_i$  is the number of samples in the  $i$ -th bin of the synthetic predictions and  $S_i$  is the number of samples in the  $i$ -th bin of the original measured data. There are other statistical metrics, such as Kolmogorov-Smirnov and Anderson-Darling tests<sup>24</sup>, or mutual information (Kullback-Leibler divergence), which could have been used as goodness-to-fit tests in addition to chi-squared test.

### 3.4 Pixel- and Parameter-Based Comparison

The metric labeled as “Pixel/Param-test” in Figure 2 refers to direct pixel-to-pixel comparison or to PDF parameter comparison. The pixel-to-pixel comparison is defined by Equation (5).

$$E_{Pixel}^{Total Avg} = \frac{1}{numBands} \sum_{band=1}^{numBands} E_{Pixel}^{Avg}(band) \quad (5)$$

where  $numBands$  is the total number of bands in the evaluated image and

$$E_{Pixel}^{Avg}(band) = \frac{1}{numPixels} \sum_{i=1}^{numPixels} \left\| I_{orig}(band, i) - I_{synth}(band, i) \right\| \quad (6)$$

The parameter-based comparison is based on a simple evaluation of all chosen parameters, for instance, central moments or PDF model parameters. The error is computed according to Equation (7).

$$E_{Param} = \begin{cases} pass; \text{ if } \|p_{orig}(i) - p_{estimated}(i)\| < Tolerance; \forall i, i \in [1, numParam] \\ fail; \text{ otherwise} \end{cases} \quad (7)$$

where  $p(i)$  is the  $i$ -th parameter, and  $\text{numParam}$  is the total number of parameters.

## 4 EXPERIMENTAL RESULTS

### 4.1 Test Data Preparation

**Test Data:** We conducted accuracy evaluation experiments with color (RGB) and hyperspectral images. The metrics described in Section 3 were applied to three types of test images. First, we obtained an aerial color photograph of a farm land in central Illinois acquired from a private airplane in an altitude about 250m (test data #1). This photograph was artificially modified with MS paint software to evaluate the magnitude of error deviations due to corrupted pixels measured by each metric. Figure 3 shows the two test images.



Figure 3: Color (RGB) aerial photograph (left) and its corrupted version (right) denoted as “scratch” image in our evaluations.

Second, we acquired in our laboratory a sequence of color images under multiple view angles of black and white sand using a visible spectrum camera, Sony SNC-RZ30N PTZ Pan/Tilt/Zoom Network Color Camera SNCRZ30N<sup>17</sup> (test data #2). Third, with the same scene setup in our laboratory as for the second test image set, we collected HS images using a hyperspectral camera manufactured by Opto-Knowledge Systems Inc.<sup>18</sup> and based on liquid crystal tunable filters<sup>19</sup> (test data #3). The angular orientation of the hyperspectral camera and the light source were measured by attaching an orientation tracker device, IS300Pro<sup>22</sup>, and reporting roll, yaw and pitch values (Euler angles). In order to compare unseen laboratory color and HS images with predicted ones, we created a test plate with half white and half black sand, separated in the middle as shown in Figure 4. We used our laboratory camera setup to take color and HS pictures of this test plate at the same angles as in the case of either white or black sand plates.

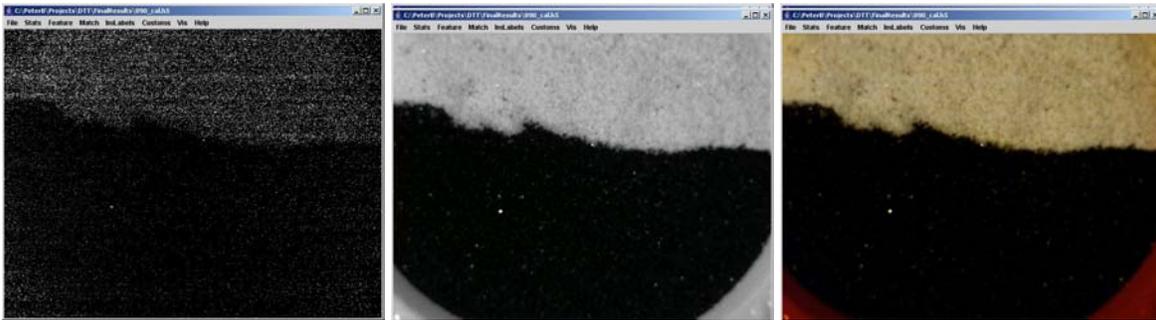


Figure 4: A measured HS image acquired at azimuth 270 and elevation 90 degrees. From left to right, the images correspond to band 1 (440 nm), band 29 (720nm), and to HS to RGB converted image.

**Image Calibration:** We did not perform any calibration of the test data sets #1 and #2 because there was no need to compensate for acquisition parameters. We compared measured images with predicted images that were derived directly from measured images by extracting model parameters and re-generating clutter values. We calibrated the test data set #3 by acquiring three images, denoted as black, white and measured images. The black image is acquired with lens cap on and corresponds to the camera’s spatial noise distribution. In other words, the black image defines the lowest value a pixel can have. The white image is acquired by viewing a calibration board with reflectance value of about 97% and it defines the highest value a pixel can have. The calibration formula is then defined in Equation (8). The outcome of the

calibration step is a reflectance value at each pixel from 0 to 1 with respect to black and white measurements (pixels with a value larger than white or smaller than black are clipped).

$$\text{calibrated\_pixel} = (\text{original\_pixel} - \text{black\_pixel}) / (\text{white\_pixel} - \text{black\_pixel}) \quad (8)$$

**Feature Space Transformation:** We explored the RGB and HSV (hue, saturation and value) color spaces for the test data sets #1 and #2 by transforming raw values in RGB space to HSV space. In the case of HS imagery (test data set #3), we evaluated raw values and transformed values using principal component analysis (PCA).

**Spatial Mask Creation:** In order to delineate clutter region in the data set #1 and the background from the sand regions in the data sets #2 and 3, we created spatial masks using manual, semi-automated and fully-automated methods. The manual creation utilized MS paint software while the semi-automatic creation was based on region-growing segmentation implemented in I2K<sup>32</sup>. We also used automated clustering (no constraint on labels to form a contiguous regions) and segmentation (spatially constrained to form contiguous regions) algorithms to create masks of the aerial image. Other techniques for texture and non-texture segmentation were explored to improve clutter separation<sup>9, 10, 11, 12</sup>.

**BTF Model Parameter Extraction:** For the test data sets #2 and #3, we estimated the PDF model from the family of Johnson’s PDFs and then extracted four parameters describing each PDF model. The parameters were stored in a database and indexed by clutter type (black and white clutter), and viewing azimuth and elevation angles. We did not vary other acquisition parameters. To be able to apply the histogram-based error metric to HS imagery, the number of bins was set to 50.

## 4.2 Experimental Results

In order to understand the significance of each computed metric value, we have investigated metrics described in Section 3 in terms of the impact on visual image appearance. Our reported results can be interpreted well if they are associated with a level of visual deviation. For example, the aerial photograph corrupted with a red curvilinear line in Figure 3 was created to investigate the impact of the “red scratch” on the four metric values. We then provide our evaluation results with variable (a) feature space (RGB, HSV, with PCA, without PCA), (b) azimuth and elevation angles of the camera view, (c) mask creation techniques, and (d) amounts of image corruption.

**Correlation-based error metric:** First, the correlation error metric was tested with the corrupted image from test data #1. Furthermore, we tested the correlation value changes due to multiple mask creation techniques and color space selection (RGB and HSV feature space). The correlation-based results are summarized in Table 1. We concluded that this metric led consistent observations with our visual perception of image quality. From a quantitative viewpoint, the decrease of the correlation value due to random curvilinear line in the “scratch test image” was about 2-8% of the maximum correlation value.

Table 1: Correlation coefficients computed from the original aerial photograph and generated synthetic images using multiple masking methods. Perfect – original image, scratch – image with inserted red lines, paint – manually created mask, rgb – RGB color space, hsv - HSV color space, iso5 – the mask created by using Isodata clustering algorithm with 5 labels (clusters), iso8 - the mask created by using Isodata clustering algorithm with 8 labels, isopaint - the mask created by using Isodata clustering algorithm and manually adjusted to form contiguous regions.

| Mask&Features\Band | Red      | Green    | Blue     |
|--------------------|----------|----------|----------|
| none & rgb         | 1        | 1        | 1        |
| scratch & rgb      | 0.972013 | 0.918532 | 0.988359 |
| paint & rgb        | 0.636607 | 0.588386 | 0.834411 |
| paint & hsv        | 0.661613 | 0.554613 | 0.833382 |
| iso5 & rgb         | 0.713721 | 0.732382 | 0.817106 |
| iso5 & hsv         | 0.73289  | 0.649264 | 0.827051 |
| iso8 & rgb         | 0.71428  | 0.434226 | 0.728956 |
| isopaint & rgb     | 0.68279  | 0.636866 | 0.847646 |

Second, we used the test data set #2 and compared the measured image of an unseen test plate with half white and half black sand with a predicted image using database representation of black and white clutter. This comparison resulted in a correlation coefficient equal to 0.97. We performed the same comparison with the test data set #3, and Figure 6 shows the comparative result of measured and synthetic images for black and white clutter types using the correlation-based error. The horizontal axis denotes all combinations of azimuth and elevation angles and the vertical axis refers to the absolute value of the correlation coefficient in zero to one interval. The higher the correlation value, the better match between measured and synthetic data. Given the metric, the white sand clutter prediction matches measured data almost perfectly (correlation equals to one) for all combinations of viewing angles, and regardless of feature selection (with PCA or without PCA). The black sand clutter prediction correlates with the measured data approximately at 0.55. We hypothesize that the black sand color values are more perturbed by the camera noise (also black) and therefore the prediction accuracy is worse than for the white sand clutter. One can perceptually evaluate an example prediction shown in Figure 5.

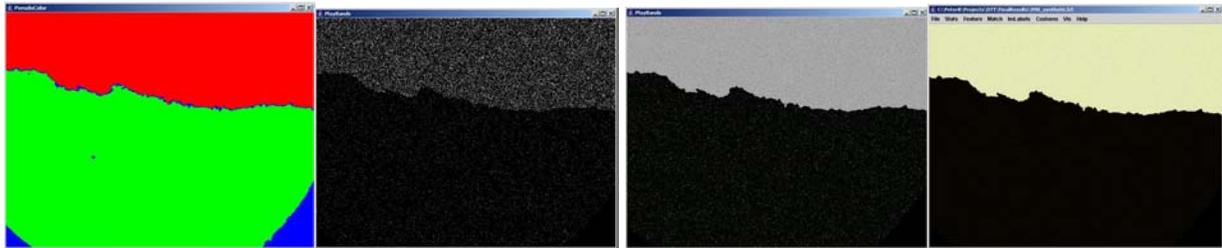


Figure 5: Left – the spatial mask corresponding to the HS image shown in Figure 4. Second left- predicted image for band 1 (440 nm). Second right – predicted image for band 29 (720nm). Right – predicted HS images converted to RGB color space.

### Correlation PCA vs Non PCA

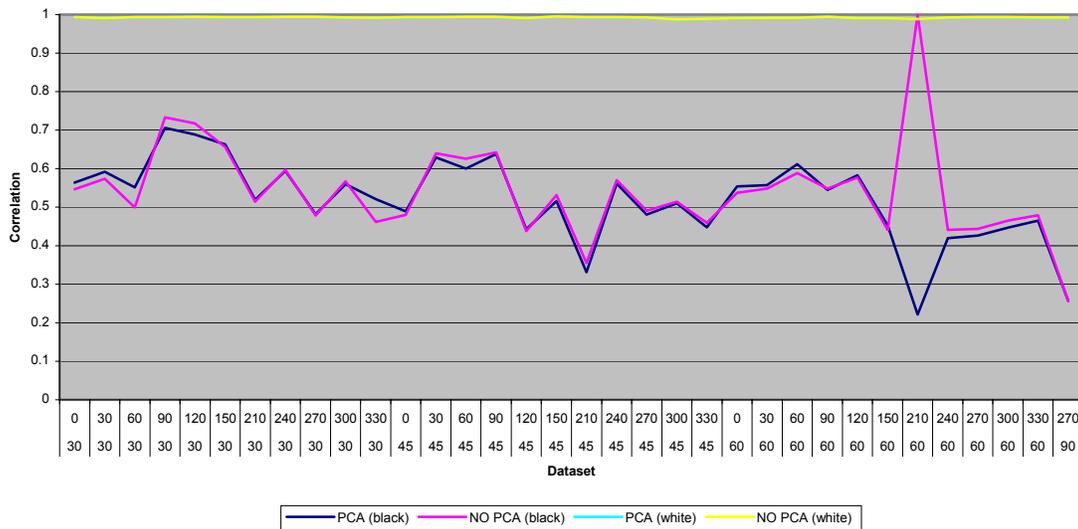


Figure 6: Linear correlation coefficient for black and white clutter types, and for de-correlated (PCA) and correlated (NO PCA) features. Horizontal axis represents the combination of azimuth (top) and elevation (bottom) angles.

**Histogram-based error metric:** To evaluate the sensitivity of the histogram-based error metric, we computed first the histogram error for the “scratch” image shown in Figure 3. The maximum error per band and per mask found was 3%. It occurs spatially in the clutter region of the left lower triangular that contains the longest scratch. Spectrally, the maximum error contribution comes from the green band because all green values were replaced by red color. The results for multiple clutter types and bands are reported in Table 2. The outcome of this experiment corresponds to our visual assessment of the image scratch.

Table 2: A summary of the histogram-based error for the scratched field image in Figure 3.

| Mask Label   | Band  | Error                |
|--------------|-------|----------------------|
| Green field  | Red   | 0.01625599175440584  |
| Green field  | Green | 0.0130440121920401   |
| Green field  | Blue  | 0.009944070283836481 |
| Road         | Red   | 0.0                  |
| Road         | Green | 0.0                  |
| Road         | Blue  | 0.0                  |
| Yellow field | Red   | 0.02462624728650553  |
| Yellow field | Green | 0.030217759162042057 |
| Yellow field | Blue  | 0.029872829582018312 |
| Brown field  | Red   | 0.0229080395802157   |
| Brown field  | Green | 0.020837441687035607 |
| Brown field  | Blue  | 0.015447317648067838 |

Next, we evaluated the test data set #3. Figure 7 shows the comparative result of measured and synthetic images for black and white clutter types using the histogram-based error. The horizontal axis denotes all combinations of azimuth and elevation angles and the vertical axis refers to the percentage mismatch measured by the error metric. The lower the error, the better match between measured and synthetic data. The overall error is about 8% with no significant difference between models built with or without PCA de-correlated features.

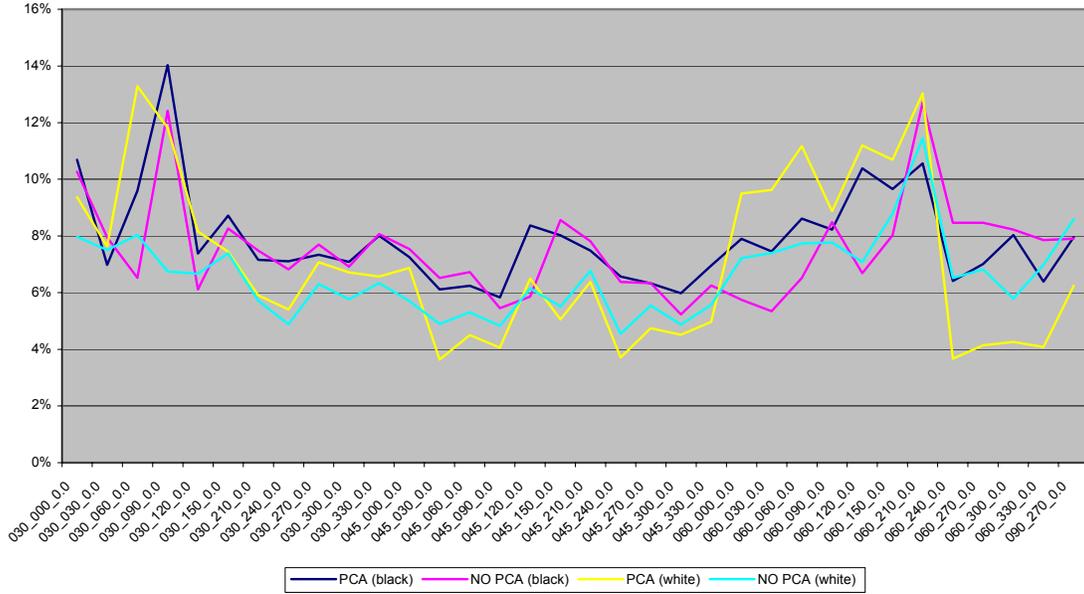


Figure 7: Total histogram error for black and white clutter types, and for de-correlated (PCA) and correlated (NO PCA) features. Horizontal axis represents the combination of elevation and azimuth angles.

**Chi-squared error metric:** According to Section 3.3, the chi-squared goodness of fit test compares the number of samples found with the expected number of samples per bin. A large chi-squared value indicates that the predicted values do not closely match the expected values. To assess the magnitude of changes using chi-squared metric, we varied the number of corrupted pixels equal to 5, 10 or 100 “wrong” pixels. According to the chi-squared value and its associated probability (cite Numerical recipes), the outcome results in declining the hypothesis that the original and the modified image with “wrong” pixels follow the same distribution represented by a histogram. These illustrative numbers shown in Table 3 were computed for a synthetic image of the size 100x100 pixels, single band, and all values were zero except for those labeled as “wrong” that were set to one. This is similar to corrupting the aerial photograph by the red line. Due to the fact that a corruption of 5 pixels among 10,000 samples is visually insignificant but would be reported by the chi-squared metric as a complete failure, the use of this metric should be carefully considered depending on the end application.

Table 3: An example of the chi-squared error values as a function of the number of modified image pixels.

| Number of Wrong Pixels | Chi-squared Error | Degrees of Freedom | Probability of Matching Distribution |
|------------------------|-------------------|--------------------|--------------------------------------|
| 5                      | 5.0012            | 1                  | 0.02533                              |
| 10                     | 10.0050           | 1                  | 0.00156                              |
| 100                    | 100.5025          | 1                  | $1.18247 * 10^{-23}$                 |

**Pixel- and parameter-based error metric:** The pixel-based error metric was applied to the test data set #3. Figure 8 reports the comparative result of measured and synthetic images for black and white clutter types using the pixel-based error defined in Section 3.4 and averaged over all bands. The horizontal axis denotes all combinations of azimuth and elevation angles and the vertical axis refers to the average color mismatch per pixel and per band. Clearly, this metric takes into account spatial locations of colors that are not model with our current approach. The higher the average pixel error the worse match between measured and synthetic data. Given this metric, the white sand clutter prediction matches measured data less accurately than the black sand predictions regardless of feature selection (with PCA or without PCA). The white sand clutter prediction corresponds to about 10% error in color prediction (400 out of 4000 values) and the black sand prediction corresponds to around 2.5% error (100 out of 4000 values). The parameter-based error metric was used during software development to test whether our generators and estimators are implemented correctly. We do not report any quantitative accuracy results.

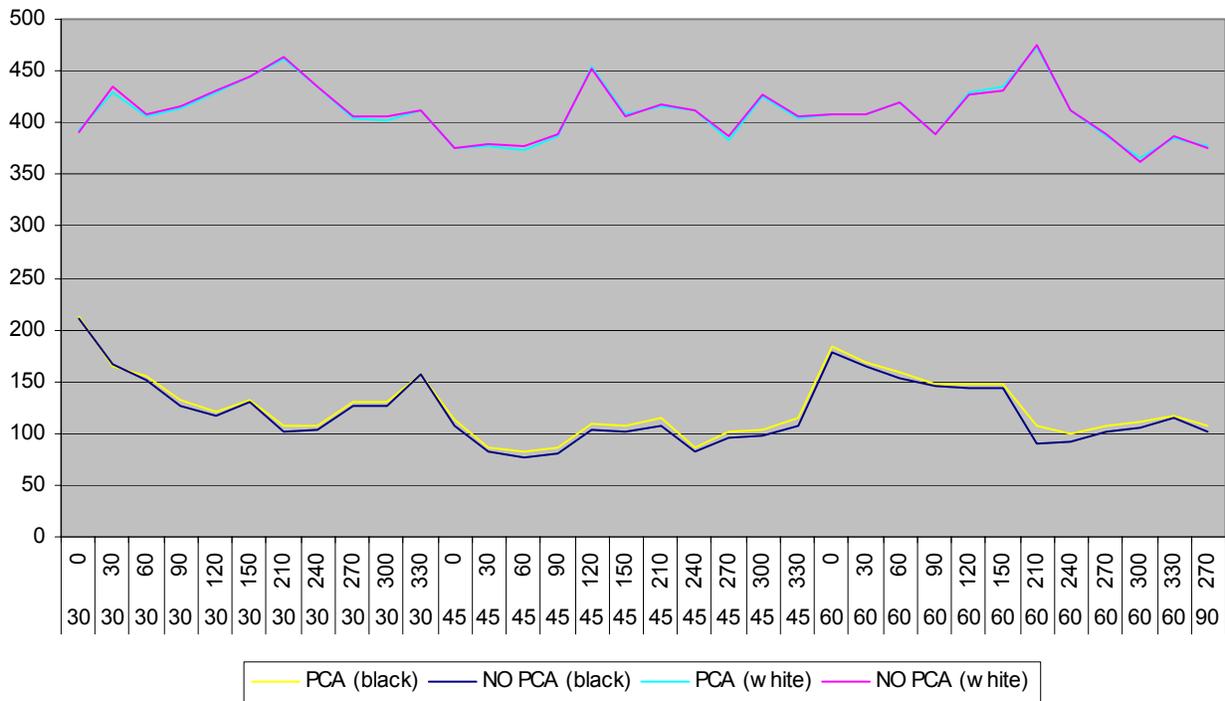


Figure 8: Average pixel error obtained by comparing original and synthetic images of black and white clutter types, and for decorrelated (PCA) and correlated (NO PCA) features. Horizontal axis represents the combination of azimuth (top) and elevation (bottom) angles.

## 5 CONCLUSIONS

We presented a novel methodology for evaluating statistically predicted versus measured multi-modal imagery. Our experimental results included color (red, green and blue) and HS imaging modalities, and Johnson's family of PDFs as the underlying statistical model. We concluded that correlation, histogram, and pixel based error metrics led to consistent observations with our visual perception of image quality. The chi-squared error metric is very sensitive to pixel level inaccuracy and should be carefully considered depending on the end application.

From a quantitative viewpoint, the correlation values for (a) corrupted aerial photograph (0.92-0.98) and (b) color predicted images of black and white sand clutter (0.97) were quite high. For the HS predicted images of black and white sand clutter, the correlation for white sand is almost 1.0 and for black sand approximately 0.55. We hypothesized that the black sand color values are more perturbed by the camera noise (also black) and therefore the prediction accuracy is worse than for the white sand clutter. Using the histogram-based metric on HS imagery, the overall error was about 8% with no significant difference between models built with or without PCA de-correlated features. Finally, the HS evaluations based on pixel-based error metric have shown prediction accuracy for black and white sand that is no worse than 90% regardless of feature selection (with PCA or without PCA). The white sand clutter prediction corresponds to about 10% error in color prediction (400 out of 4000 values) and the black sand prediction corresponds to around 2.5% error (100 out of 4000 values).

### Acknowledgements

This material is based upon work partially supported by the National Center for Advanced Secure Systems Research (NCASSR), and the National Laboratory for Advanced Data Research (NLADR).

### References:

- 1 MODTRAN, developed by AFRL/VSBT in collaboration with Spectral Sciences, Inc., MODerate spectral resolution atmospheric TRANSmittance algorithm and computer model for atmospheric spectral modeling, URL: <http://www.vs.afrl.af.mil/Division/VSBYB/modtran4.html>
- 2 Xpatch, developed by SAIC/Demaco, IL, synthetic aperture radar prediction of realistic far-field and near-field radar signatures for 3D target modeling, URL: <http://www.saic.com/products/software/xpatch/>
- 3 DIRSIG, developed by Digital Imaging and Remote Sensing Laboratory at the Rochester Institute of Technology, radiometric modeling of various multispectral and hyperspectral imaging sensors, URL: <http://www.cis.rit.edu/research/dirs/research/dirsig.html>
- 4 MuSES, RadTherm, and WinTherm, developed by ThermoAnalytics, Inc., thermal prediction, URL: <http://www.thermoanalytics.com/products/>
- 5 Glossary. Erosion and Sediment Control Management System. Lake Macquarie, Australia: 9 September 1999.
- 6 Bajcsy, P., D. Sullivan, and P. Ryan. "Estimating Weibull Distribution Parameters". in Intelligent Engineering Systems Through Artificial Neural Networks. 2000: ASME Press, New York.
- 7 Dana, K.J., et al., "Reflectance and texture of real-world surfaces". ACM Transactions on Graphics, 1999. **18**(1): p. 1-34.
- 8 Dana, K.J. and S.K. Nayar. "Correlation Model for 3D Texture". in International Conference Computer Vision. 1999.
- 9 Leung, T.K. and J. Malik. "Recognizing Surfaces using Three-Dimensional Textons". in International Conference Computer Vision. 1999.
- 10 Lu, S.Y. and K.S. Fu, "Stochastic Tree Grammar Inference for Texture Synthesis and Discrimination". Computer Graphics and Image Processing, 1979. **9**: p. 234-245.
- 11 Panjwani, D.K. and G. Healey, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images". IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995. **17**(10): p. 939-954.
- 12 Portilla, J. and E.P. Simoncelli, "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients". International Journal of Computer Vision, 2000. **40**(1): p. 49-70.
- 13 Slipfker, J.F. and S.S. Shapiro, "The Johnson System: Selection and Parameter Estimation". Technometric, 1980. **22**(2): p. 239-246.
- 14 Ulaby, F.T. and M.C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*. 1989, Norwood, MA: Artech House.

- 15 Varma, M. and A. Zisserman. "Classifying Images of Materials: Achieving Viewpoint and Illumination Independence". Proceedings of the 7th European Conference on Computer Vision in Copenhagen, Denmark, Vol 3. Springer-Verlag, May 2002, pp 255—271.
- 16 Zhu, S.C., Y. Wu, and D. Mumford, "Filters random fields and maximum entropy (FRAME): To a unified theory for texture modeling". International Journal of Computer Vision, 1998. 27(2): p. 1-20.
- 17 Sony camera. Available at <http://www.camerasuperstore.com/sonsnnetcolc.html>.
- 18 Official Opto-Knowledge Systems Inc. website. Available at <http://www.techexpo.com/WWW/opto-knowledge/index.html>
- 19 T. Chrien, C. Chovit and P. Miller. "Imaging Spectrometry Using Liquid Crystal Tunable Filters". Jet Propulsion Laboratory and Cambridge Research & Instrumentation Inc.
- 20 Duncan multispectral camera. Available at [http://www.duncantech.com/area\\_scan\\_cameras.htm](http://www.duncantech.com/area_scan_cameras.htm).
- 21 Omega thermal IR camera. Available at <http://indigosystems.com>
- 22 IS300 Pro motion tracker, Available at <http://www.isense.com/products/prec/is300/is300pro.htm>.
- 23 Karu Kalle, Jain Anil K., Bolle Ruud M. "Is there any texture in the image?", Pattern Recognition, Vol. 29, No. 9, pp. 1437-1446, 1996.
- 24 Cullen A.C. and H. C. Frey, *Probabilistic Techniques in Exposure Assessment*, Plenum Press, New York, 1999.
- 25 Gat, N. "Imaging Spectroscopy Using Tunable Filters: A Review". Proc. SPIE Vol. 4056, p. 50-64, Wavelet Applications VII, Harold H. Szu, Martin Vetterli; William J. Campbell, James R. Buss, Ed. (Invited paper). Publ. Date: 04/2000.
- 26 Cressie, Noel A.C. *Statistics for Spatial Data*. Wiley & Sons, Inc. 1993
- 27 CURet Database; URL: <http://www1.cs.columbia.edu/CAVE/curet/>
- 28 Brodatz P., *Textures: A Photographic Album for Artists and Designers*, Dover Publications, New York, 1966.
- 29 VisTex Databases; URL: <http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>
- 30 Bajcsy P., R. Kooper and D. Andersh, "Multisensor Scene Modeling Using Statistical Models for Bidirectional Texture Functions," Proceedings of SPIE 2004, Algorithms for SAR Imagery XI, April 12-15, 2004. Vol. 5427, [5427-30].
- 31 Bajcsy P. and R. Kooper, "Prediction Accuracy of Color Imagery from Hyperspectral Imagery," SPIE 2005, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI, 28 March - 1 April 2005, Orlando (Kissimmee), Florida USA.
- 32 Image to Knowledge (I2K), a library of software tools developed at NCSA/UIUC, web documentation at <http://alg.ncsa.uiuc.edu/tools/docs/i2k/manual/index.html>.
- 33 Ramanath R., W. E. Snyder, Y. Yoo and M.S. Drew, "Color Image Processing Pipeline," IEEE Signal Processing Magazine, Vol. 22, No. 1., January 2005, pp 34-43.