Refinement of a Method for Identifying Probable Archaeological Sites from Remotely Sensed Data

James C. Tilton*[a], Douglas C. Comer[b], Carey E. Priebe[c], Daniel Sussman[c], Li Chen[c]

[a]NASA Goddard Space Flight Center, Mail Code 606.3, Greenbelt, MD, USA 20771;
[b]Cultural Site Research and Management, 2113 St. Paul St., Baltimore, MD, USA 21218;
[c]Dept. of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA 21218

ABSTRACT

To facilitate locating archaeological sites before they are compromised or destroyed, we are developing approaches for generating maps of probable archaeological sites, through detecting subtle anomalies in vegetative cover, soil chemistry, and soil moisture by analyzing remotely sensed data from multiple sources. We previously reported some success in this effort with a statistical analysis of slope, radar, and Ikonos data (including tasseled cap and NDVI transforms) with Student's t-test. We report here on new developments in our work, performing an analysis of 8-band multispectral Worldview-2 data. The Worldview-2 analysis begins by computing medians and median absolute deviations for the pixels in various annuli around each site of interest on the 28 band difference ratios. We then use principle components analysis followed by linear discriminant analysis to train a classifier which assigns a posterior probability that a location is an archaeological site. We tested the procedure using leave-one-out cross validation with a second leave-one-out step to choose parameters on a 9,859x23,000 subset of the WorldView-2 data over the western portion of Ft. Irwin, CA, USA. We used 100 known non-sites and trained one classifier for lithic sites (n=33) and one classifier for habitation sites (n=16). We then analyzed convex combinations of scores from the Archaeological Predictive Model (APM) and our scores. We found that that the combined scores had a higher area under the ROC curve than either individual method, indicating that including WorldView-2 data in analysis improved the predictive power of the provided APM.

Keywords: Image processing, classification, feature extraction

1. INTRODUCTION

Archaeological sites are being destroyed or compromised at a catastrophic rate in most regions of the world in this age of globalization. Populations are growing much more rapidly in developing countries than in developed countries. These populations are migrating to areas that in the past have been affected very little by human activities. As China, India, Brazil, and other countries aspire to levels of consumption enjoyed by the United States and Europe, large areas are cleared for agriculture and communities are springing up around mines, oil wells, and newly established industries. At the same time, tourism has been seized upon by developing countries because it holds the promise of jobs for which high levels of education and training are not necessary. If archaeological sites are not destroyed outright by construction of hotels, transportation systems, and other visitor amenities, the increasing numbers of people that inhabit or travel through heretofore isolated areas results in the discovery of archaeological sites not by archaeologists, but by indigenous populations. Site looters who are tied into the global, underground, and illicit trade of antiquities seek out members of indigenous communities with such knowledge and exploit them. For example, a Khmer statue that might have been obtained for a bowl of rice from a Cambodian living in Siem Reap might fetch thousands of dollars in one of the many antiquity boutiques in Bangkok. The only effective solution to all of the problems listed above is to find archaeological sites and landscapes before they are overtaken by agricultural and industrial development or found by looters, and then to take steps to protect them.

Detecting archaeological sites by the analysis of imagery acquired by airborne and spaceborne sensors is often possible because human activities hundreds or even thousands of years ago has frequently caused localized

---

*James.C.Tilton@nasa.gov; phone 1 301 286-9510; fax 1 301 286-1776; science.gsfc.nasa.gov/sed/

environmental changes that have persisted to the present. These can be the addition of humanly constructed features to the landscape that display a regular geometry and so are anomalous to their surroundings. Such features include houses, temples, causeways, roads, trails, agricultural terraces, reservoirs, canals, and administrative structures. Human activities also produce changes in soil chemistry that can be detected directly, or that cause subtle anomalies in the vegetative cover because certain plant species flourish in or are stressed by altered soils. Non-native vegetation introduced in ancient times for food or for aesthetic purposes sometimes persists for great periods of time.

Comer and Blom[1–3] developed a means of utilizing aerial and satellite remote sensing data to identify probable locations of archaeological sites over a wide area based on detecting the subtle anomalies in vegetative cover described above. The remote sensing data utilized included SAR (synthetic aperture radar) data collected by the GeoSAR (airborne synthetic aperture radar) sensor, 11-bit multiple band optical data from the IKONOS sensor (operated by GeoEye, Inc.). A DEM (digital elevation model) was derived from the SAR data from which slope data was computed. These data, along with features computed from the IKONOS data served as input image layers for the anomaly detection approach.

We report here on new developments in our work, performing linear discriminant analysis on 28 band difference ratios computed from 8-band multispectral Worldview-2 data. We first describe both the remotely sensed data and the archaeological site data that was utilized for training and testing our analysis approach. Then we elaborated on our new analysis approach, and the demonstrate the effectiveness of our approach over a test site over the western portion of Ft. Irwin, CA, USA and compare our results with an Archaeological Predictive Model (APM).

2. DESCRIPTION OF THE DATA

In this section we describe the remotely sensed data we utilized in our study, along with the archaeological site data that we utilize for training and testing our analysis approach.

2.1 The Remotely Sensed Data

The remotely sensed data employed in this study was multispectral WorldView-2 data. DigitalGlobe, Inc., Longmont, CO, USA (http://www.digitalglobe.com) contributed 14 swaths of WorldView-2 satellite imagery data to our study covering the entire land area of the China Lake and Ft. Irwin military reservations in California, USA. We report on our use of just one of these satellite imagery data swaths in this study, which covers a large part of the western portion of Ft. Irwin. This data was collected on December 30, 2010. While DigitalGlobe provided both panchromatic and multispectral data, we utilized only the multispectral data in our study. Table I shows the spectral bands provided by the WorldView-2 satellite. This data is orthorectified to 2-meter ground resolution and has 11-bit radiometric resolution (stored as 16-bit integers).

Table 1: WorldView-2 spectral bands (from DigitalGlobe Core Imagery Product Guide available on their website.

Band	Lower Band Edge (nm)	Center Wavelength (nm)	Upper Band Edge (nm)
Coastal Blue	396	427	458
Blue	442	478	515
Green	506	546	586
Yellow	584	608	632
Red	624	659	694
Red Edge	699	724	749
Near Infrared 1	765	833	901
Near Infrared 2	856	949	1043

DigitalGlobe provided this swath of data in 45 separate 4096x4096 pixel blocks of data, which we mosaicked together to form an eight spectral band data set with 9859 columns and 61098 rows. We used in our analysis a 9859 column by 23000 row subset of this data that covered the grounds of a western portion of Ft. Irwin. All locations outside of the Ft. Irwin land area boundary were masked out.

Based on the suggestion of G. Marchisio of DigitalGlobe,[4] we utilized all possible combinations of band difference ratios of the eight multispectral bands as features for our study. Band difference ratios offer some robustness to changes in lighting conditions at different times of data collection, which will help in our eventual extension of our analysis of this portion of Ft. Irwin to other portions of Ft. Irwin and to the nearby China Lake site. These band difference ratio features are defined as

$$B_{ij} = \frac{B_i - B_j}{B_i + B_j}, \quad \forall i > j. \tag{1}$$

2.2 The Archaeological Site Data

The archaeological site location data was stored in ESRI shapele (geospatial vector data) format with sites designated as Lithic Scatter or Habitation and Quarry. The tests in this study were performed using 62 Lithic Scatter sites and 22 Habitation sites. Precise for these sites were collected from October, 2010 to January, 2011 by revisiting and evaluating sites that had been recorded over the past 50 years at Fort Irwin, California, which is located in the Mojave Desert. Fieldwork was supervised by author Douglas C. Comer with a crew of professional archaeologists.

2.3 The Archaeological Predictive Model

The site detection protocols described in this paper are properly thought of as an approach that is different in kind from an approach that has been used in archaeology for several decades now to predict areas in which archaeological sites are likely to be found. Such archaeological predictive models (APMs) do not associate certain returns from aerial and satellite remote sensing devices with the locations of archaeological sites, as we do in this paper. Instead, they are intended to identify areas within which sites might be found, rather than finding the sites themselves. That is, APMs predict on the basis of possibilism instead of probablism. Being possibilistic, APMs deal only with how suitable an area is for a certain activity. It is important to note that not all suitable areas will be used. For this reason, APMs have had limited success in predicting the locations of archaeological sites.

Nonetheless, they have had some success, and are the standard against which new the results of new approaches can be compared. Also, in the future, some of the approaches and technologies that we have developed might very well be used to improve the performance of APMs. For this reason, we utilized an APM that had been developed for one of our test areas, Fort Irwin. It was produced by the University of Illinois at Urbana-Champaigns GIS and Geospatial Lab, through the U. S. Army Corps of Engineers Construction Engineering Research Laboratory (CERL) in Champaign, Illinois.[5]

Utilizing a variety of statistical techniques, Ruiz attempted to develop associations between the following environmental factors and the presence of archaeological sites:

- Distance to spring or lake

- Soil surface texture

- Soil water content

- Depth to bedrock

- Landform type

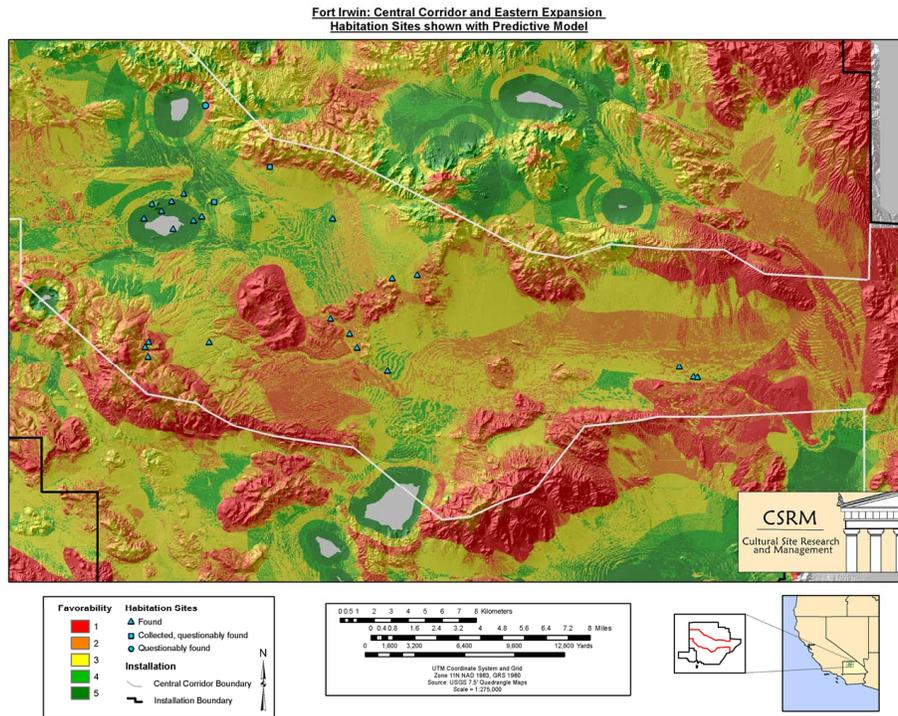- Elevation

- Slope

- Direction of slope (aspect)

Figure 1: The Archaeological Predictive Model for Habitation Sites

While the statistical treatments that were used to develop the APM for Fort Irwin are not well described, the performance of the resulting model was good as measured by the metric of gain, which has been devised for this purpose. The formula used to calculate the gain statistic is: $1 - \frac{\% \ of \ area}{\% \ of \ sites}$.[6]

The APM for Fort Irwin was actually composed of several APMs, each predicting areas in which different types of archaeological sites might be found. We will consider here only the APMs done for only two types of sites: habitation and lithic. An APM for these habitation sites is seen in Fig 1. Areas in green are those which are predicted to be most favorable for the presence of archaeological sites, red the least, and colors between those two in order of decreasing likelihood. For the highest class in the APMs (the area on the final favorability map that is colored dark green to indicate that the area is the most favorable for the type of site in question) the gain statistic is 0.86 for habitation sites. For the green area in the Lithic APM, a gain statistic of 0.67 was calculated. ROC curves presented elsewhere in this paper compare the performances of APMs and the site detection protocols that we have developed. It is significant that the two approaches can be combined to produce a better performance than could be obtained by either alone.

3. ANALYSIS APPROACH

In this work, we present a framework to classify whether a location is likely to be an archaeological site based on multi-spectral image analysis. Image processing is first used to form a quantitative representation of the location. Secondly, feature extraction is used on this image processed representation to form a low dimensional representation of the location. The final step is to classify this low dimensional representation to determine whether or not the location is suitable for an archaeological site. In order to build this classifier, appropriate

training data and classification techniques must be used. We validated this method by comparing with the archaeology predictive model (APM). By considering combined classifiers using the APM and the WorldView analysis we can demonstrate that our method provides additional discrimination power over the APM.

In this section, we present the statistical setting and a general framework for image processing, dimensionality reduction and classification and conclude with a discussion of our parameter selection and validation technique.

3.1 Statistical Setting

We consider a multi-spectral image as B random images. Formally, let w, h, and B be positive integers. Let $\mathcal{S} = \{1, 2, \ldots, w\} \times \{1, 2, \ldots, h\}$. Each element $s = (s_1, s_2) \in \mathcal{S}$ corresponds to a specific geographical location, such as latitude and longitude coordinates. A feature image $\mathcal{B} \in \mathbb{R}^{w \times h}$ is a random image which maps each location to an intensity value. A multispectral image is a B-tuple of bands $(\mathcal{B}^{(1)}, \mathcal{B}^{(2)}, \ldots, \mathcal{B}^{(B)}) \in (\mathbb{R}^{w \times h})^B$. The value of band $b \in \{1, 2, \ldots, B\}$ at location $s \in \mathcal{S}$ is denoted $\mathcal{B}_s^{(b)}$. We assume that the bands are fully observed. Note that the bands in the multi-spectral image need not be the original bands but could be derived or preprocessed bands, such as the band difference ratio features.

In addition to the bands, we suppose that each site location is given a class label in $\{0, 1, \ldots, K-1\}$ which is encoded in the random image $\mathcal{Y} \in \{0, 1, \ldots, K-1\}^{w \times h}$ for some positive integer K. We take $K = 2$ and denote the class label at location s by $\mathcal{Y}_s$. Class label $\mathcal{Y}_s = 1$ denotes that location s is considered to have archaeological significance. Class label $\mathcal{Y}_s = 0$ denotes that site s has no archaeological significance.

All together, we have a $(B+1)$-tuple $(\mathcal{B}^{(1)}, \mathcal{B}^{(2)}, \ldots, \mathcal{B}^{(B)}, \mathcal{Y})$ which is distributed according to some unknown joint distribution. The bands, the pixels within each band, and the class labels will all depend on each other. The unknown parameters for this distribution will be determined by the geographical formations, environmental and historical conditions, the sensor used to collect the image, as well as many other factors. Due to these complications, we do not attempt to estimate its parameters directly.

Finally, we suppose that there are n locations for which ground truth has been gathered.Formally, there are n locations $s^{(1)}, s^{(2)}, \ldots, s^{(n)} \in \mathcal{S}$ and for each location $s^{(i)}$ we have observed the class label $\mathcal{Y}_{s^{(i)}}$. These locations can be used for training our method and testing our classifier against the ground truth. All other class labels are un-observered.

We consider the classification algorithm as consisting of three components the image processor, feature extractor and classifier, determined by the functions f, g, and h:

$$f : \mathcal{S} \times (\mathbb{R}^{w \times h})^B \mapsto \mathbb{R}^{\tilde{d}}, \tag{2}$$

$$g : \mathbb{R}^{\tilde{d}} \to \mathbb{R}^d \ with \ d < \tilde{d} \tag{3}$$

$$h : \mathbb{R}^d \to \{0, 1\} \tag{4}$$

The classification algorithm is then given by the composition $h \circ g \circ f$. Each of the functions f, g, and h will be chosen from some class and may depend on the sites with observed class labels. Indeed, our paper considers using statistics on annuli centered at the site for image processing, PCA for feature extraction and LDA for classification. Figure 2 provides a diagrammatic representation of this protocol.

3.2 Image Processing

The image processing step transforms the bands into statistics about each location. We formally define an image processor as a function

$$f : \mathcal{S} \times (\mathbb{R}^{w \times h})^B \mapsto \mathbb{R}^{\tilde{d}}, \tag{5}$$

that takes a location and B image bands and returns a vector. This function returns local statistics about the image for each location. For location $s \in \mathcal{S}$ let $\tilde{\mathcal{X}}_s = f(s, \mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(B)})$ and for location $s^{(i)}$ let $\tilde{X}^{(i)} = f(s^{(i)}, \mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(B)})$.
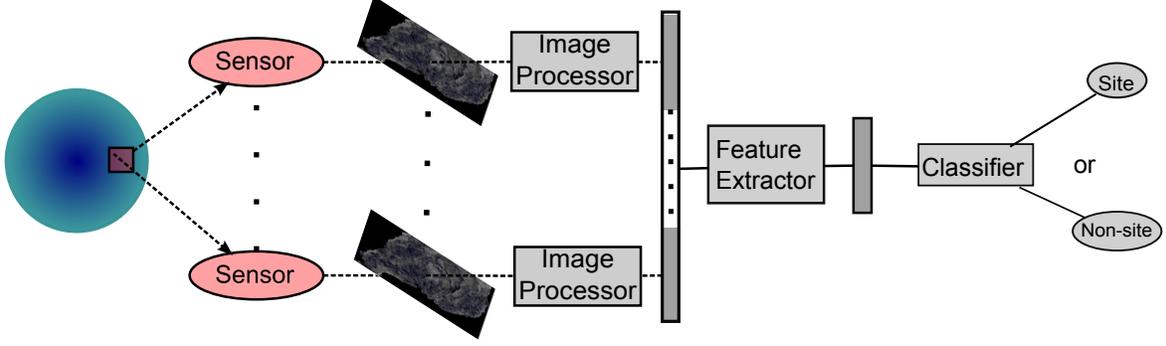
Figure 2: A diagrammatic representation of our protocol.

For this work, we suppose that

$$f(s, \mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(B)}) = \begin{pmatrix} f_1(s, \mathcal{B}^{(1)}) \\ f_2(s, \mathcal{B}^{(2)}) \\ \vdots \\ f_B(s, \mathcal{B}^{(B)}) \end{pmatrix} \in \mathbb{R}^{\tilde{d}}, \quad \forall s, \mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(B)} \tag{6}$$

where $f_b : \mathcal{S} \times \mathbb{R}^{w \times h} \mapsto \mathbb{R}^{\tilde{d}/B}$ depends only on band b for each b. (Clearly, we assume $\tilde{d}$ is an integer multiple of B.) Further we assume that all the functions f_b are identical, so that the same statistics are computed for each image. These assumptions may not be appropriate in some instances but is a useful simplification in the current setting.

In this paper, we use a particular imaging processing method based on statistics of pixels in annuli centered at each site. The image processor was motivated by the observation that the archaeological sites are typically circular and of relatively consistent sizes.

Let

$$A_s(r^{(in)}, r^{(out)}) = \{s' \in \mathcal{S} : r^{(in)} \le \|s - s'\| < r^{(out)}\} \tag{7}$$

be the set of coordinates in the annulus centered at $s \in \mathcal{S}$ with inner radius $r^{(in)}$ and outer radius $r^{(out)}$. We define $\nu : \mathcal{S} \times \mathbb{R}^{w \times h} \mapsto \mathbb{R}^{30}$ and $\delta : \mathcal{S} \times \mathbb{R}^{w \times h} \mapsto \mathbb{R}^{30}$ by

$$\nu_i(x, \mathcal{B}) = \text{median}\{\mathcal{B}_{x'} : x' \in A_x(r_i^{(in)}, r_i^{(out)})\} \tag{8}$$

$$\begin{aligned} \delta_i(x, \mathcal{B}) &= \text{MAD}\{\mathcal{B}_{x'} : x' \in A_x(r_i^{(in)}, r_i^{(out)})\} \\ &= \text{median}\{|\mathcal{B}_{x'} - \nu_i(x)| : x' \in A_x(r_i^{(in)}, r_i^{(out)})\} \end{aligned} \tag{9}$$

for $i \in \{1, 2, \ldots, 30\}$ where MAD denotes the sample median absolute deviation.

The inner and outer radii ranges we consider are given in this table:

i	1	2	3	...	10	11	12	13	...	20	21	22	23	...	30
$r^{(in)}$	0	3	6	...	27	0	5	10	...	45	0	7	14	...	63
$r^{(out)}$	2	5	8	...	29	4	9	14	...	49	6	13	20	...	69

Since the annuli overlap and cover a relatively large area, the statistics for these ranges will likely contain some redundant information so the feature extraction step will be useful to reduce this redundancy while maintaining the components that discriminate between classes.

We now define the image processing function $f_b : \mathcal{S} \times \mathbb{R}^{w \times h} \mapsto \mathbb{R}^{2 \cdot 30}$ by $f_b(x, \mathcal{B}) = [\nu(x, \mathcal{B})^T, \delta(x, \mathcal{B})^T]^T$. As in Eqn. 6, this induces the function $f : \mathcal{S} \times \mathbb{R}^{w \times h \times B} \mapsto \mathbb{R}^{2 \cdot 30 \cdot B}$.

3.3 Feature Extraction

Each location is now represented by a vector in $R^{\tilde{d}}$ (with $\tilde{d} = 60B$). The dimension $\tilde{d}$ is much larger than the sample size n. This leads to classification difficulty because not enough data is available to build a reasonable model. Such high dimensionality can also result in complicated computational issues and reduces classification accuracy. This problem is termed as "the curse of dimensionality".[7,8] Feature extraction for lowering the dimensions in feature space is thus necessary. It eliminates irrelevant features, reduces noise in the data set. Moreover it converts high dimensional feature space to a lower dimension that captures most variation of the original data. Mathematically, we define g as the following:

$$g : \mathbb{R}^{\tilde{d}} \to \mathbb{R} \ with \ d < \tilde{d}$$

Note that g depends on the labeled data $(X_1, Y_1), ..., (X_n, Y_n)$. We denote the feature extracted vectors for location $s \in S$ as $\mathcal{X}_s = g(\tilde{\mathcal{X}}_s) = (f \circ g)(s, \mathcal{B}^{(1)}, ..., \mathcal{B}^{(n)})$ and for location $s^{(i)}$ as $X^{(i)} = g(\tilde{X}^{(i)}) = (f \circ g)(s^{(i)}, \mathcal{B}^{(1)}, ..., \mathcal{B}^{(n)})$.

In this paper, principal components analysis (PCA) is used for feature extraction. Suppose the sample mean of the vectors $\tilde{X}_1, ..., \tilde{X}_n$ is zero; otherwise, we centralize the vectors by subtracting its mean. Let $\tilde{\Sigma} \in \mathbb{R}^{\tilde{d} \times \tilde{d}}$ be the sample covariance matrix corresponding to the data matrix $\mathbb{D}$ whose rows are data instances and columns are features. Note that the covariance matrix is positive semi-definite; thus it has all but negative eigenvalues and its eigenvectors form an orthonormal space. Let $U \in \mathbb{R}^{\tilde{d} \times d}$ be a matrix whose columns are orthonormal eigenvectors corresponding to the d largest eigenvalues of $\tilde{\Sigma}$. The feature extractor g of PCA is defined by $g(x) = \tilde{U}^T x \in \mathbb{R}^d$ for all $x \in \mathbb{R}^{\tilde{d}}$. Hence, the first principal component, a linear combination of the original features, is the dimension with the most variance in the original feature space; the second principal component, orthogonal to the first, captures second most variation of the data set and so on.... Usually, a small fraction of the total dimension is needed via PCA. To select the PCA dimension $\tilde{d}$ we use cross validation (see section 3.5).

3.4 Classification

We define a classifier as a function $h : \mathbb{R}^d \mapsto \{0, 1\}$. After image processing and feature extraction, h assigns a class label to each data vector on the lower dimension space.

The simple classifier we choose is linear discriminant analysis (LDA). LDA is the Bayes plug-in classifier under the assumption that the data is distributed according to class conditional multivariate normal distributions with the same covariance matrices. We estimate the means and covariance by

$$\hat{\mu}_y = \frac{1}{n_y} \sum_{i:Y^{(i)}=y} X^{(i)} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n-2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X^{(i)} - \hat{\mu}_{Y^{(i)}})(X^{(j)} - \hat{\mu}_{Y^{(j^T)}})^T, \tag{10}$$

respectively. We then have the estimate of the posterior probability that a site with feature extracted vector x is from class 1 defined by

$$\hat{\eta}(x) = \hat{\mathbb{P}}(Y = 1 | X = x) = \frac{\phi(x; \hat{\mu}_1, \hat{\Sigma})}{\frac{n_1}{n} \phi(x; \hat{\mu}_1, \hat{\Sigma}) + \frac{n_0}{n} \phi(x; \hat{\mu}_0, \hat{\Sigma})}, \tag{11}$$

where n_y is the number of training samples with class label y and $\phi(\cdot; \mu, \Sigma)$ is the density for a multivariate normal distribution with mean μ and covariance Σ.

The classifier h is defined as $h(x) = \mathbb{I}\{\hat{\eta}(x) > \tau\}$ for some threshold value $\tau \in \mathbb{R}$.

3.5 Parameter Selection and Validation

The key parameters we must select for this algorithm are the PCA dimension d and the threshold parameter τ. The threshold value τ will depend on how the classifier is being used so we do not attempt to explicitly select a threshold but consider a range of thresholds. We use leave-one-out cross validation to select the PCA dimension.

Leave-one-out cross validation is a technique to estimate the error of probability of a classifier that is trained on data. The procedure loops through each data sample, removing it and training on the remaining data. We then evaluate the resulting classifier on the removed sample and the estimated error rate is then the percent of the samples that were incorrectly classified. This results in a more accurate estimate of the generalization error of the procedure since we do not test on data that was trained on.

To select the PCA dimension d from among $\{1, 2, \ldots, n-1\}$ we perform a leave-one-out estimate of error for each d and then select the dimension which results in the smallest error. For these estimates we use a threshold value of $\tau = 0.5$. Finally, we train a final classifier on the whole dataset. This classifier then allows us to compute the posterior η for any new location.

To assess this procedure, we again use leave-one out cross validation. This means we have two loops of cross validation, an outer loop to train the classifier for each removed sample and compute a posterior and an inner loop select the PCA dimension. Additionally, we use receiver operating characteristic (ROC) curves to asses the the procedure. The ROC curve is parametrized by the the threshold τ and for each value of τ we plot the the percentage of false positives on the x-axis and the percentage of true positives on the y-axis. Depending on the application, one may be concerned with different aspects of the ROC curve. The area under the curve (AUC) is a summary statistic that computes the area under the ROC curve. This ranges from 0, reverse classification to 1, perfect classification.

4. RESULTS

We analyzed the above procedure using the 28 band difference ratio feature bands derived from the Worldview-2 data. For our true sites we consider two different classes, lithic sites ($n_1 = 33$) and habitation sites ($n_1 = 16$). We trained separate classifiers for the lithic and habitation classes. For the class zero training data we selected $n_0 = 100$ locations uniformly at random from the surveyed region of the image swath. We also ensured that the selected locations were at least 100 pixels from every known true site. Since the negatives are selected from the surveyed regions and are far from known sites, we can safely assume that these sites do not have archaeological significance.

The result of our analysis is a posterior probability for each site derived from classifier trained on the remaining sites. We denote this posterior by $\hat{\eta}_{WV}$ for the WorldView dervied posterior. Independent of the WorldView analysis, the archaeological predictive model (APM) provides a score which, when normalized to be in $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$, can be considered as an alternative posterior estimate which we denote $\hat{\eta}_{APM}$. We consider convex combinations of the APM posterior and the Worldview analysis posterior. For each value of $\gamma \in [0, 1]$ this gives a new posterior estimate:

$$\text{Combined}_\gamma = (1 - \gamma)\text{APM} + \gamma\text{WorldView} \tag{12}$$

$$\hat{\eta}_\gamma = (1 - \gamma)\hat{\eta}_{APM} + \gamma\hat{\eta}_{WV} \tag{13}$$

Figure 3 shows the AUC value for each $\gamma \in [0, 1]$ where 0 represents using only the APM and 1 represents using on the WorldView-2 analysis. For the habitation sites, Figure 3a, we see that the APM outperforms the WorldView but that a combined classifier with $\gamma \approx .3$ provides even better performance. Noticeably, there is a jump in the AUC at $\gamma = 0$. This is due to the fact that the APM gives only 6 possible values so for γ near zero the WorldView analysis acts as a tie breaker for sites that have the same APM score. Similar behavior is see for lithic sites, Figure 3b, but in this case the WorldView analysis outperforms the APM and that the a combined classifier with $\gamma \approx .65$ provides even better performance.

Figure 4 shows the receiver operating characteristic (ROC) curve of lithic sites and habitation sites corresponding to five classifiers: APM, Worldview-2, half of each ($\gamma = 0.5$), the optimal combined model, and the
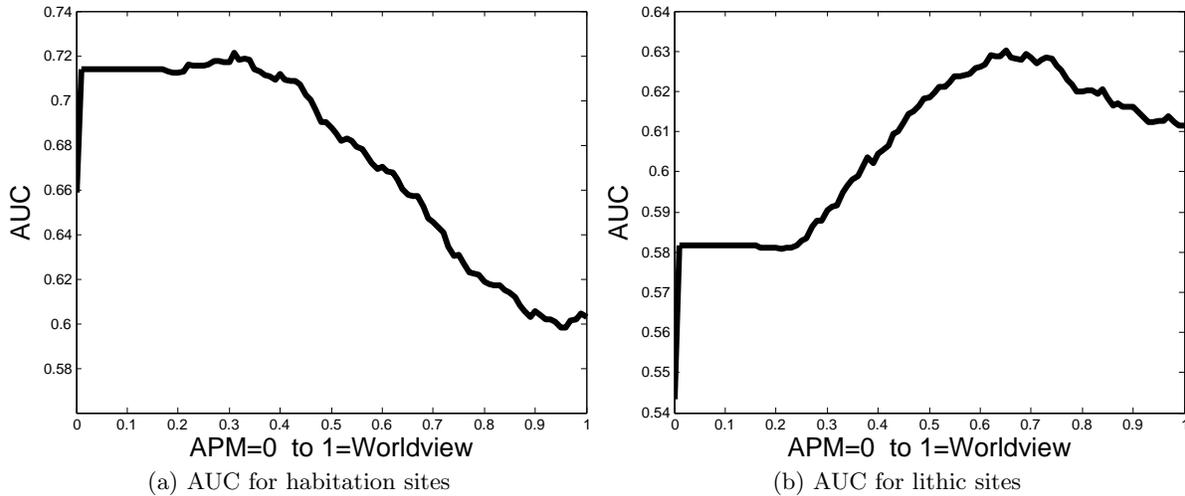
(a) AUC for habitation sites   (b) AUC for lithic sites

Figure 3: Area under the ROC curve for combined model with respect to γ.

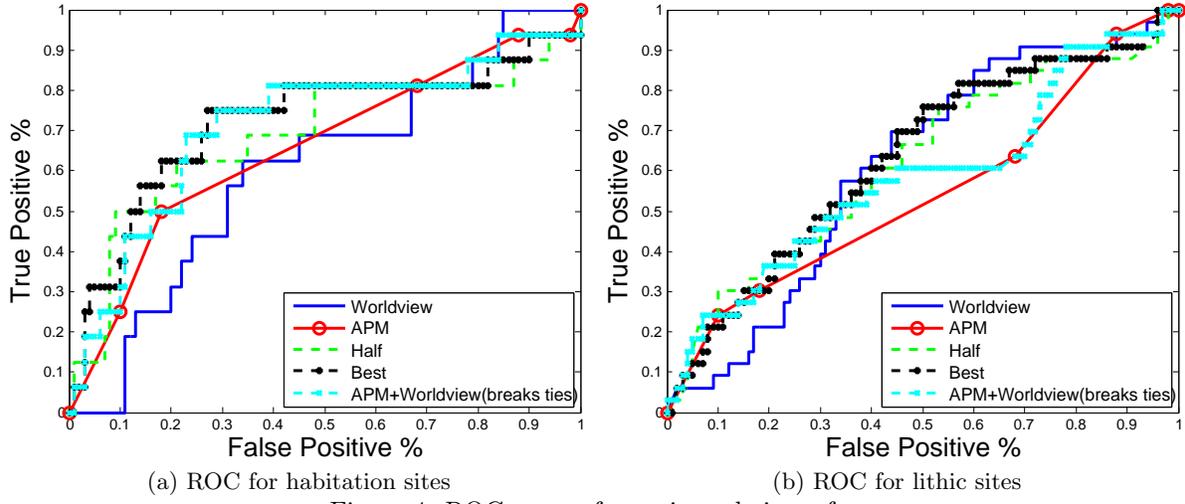(a) ROC for habitation sites   (b) ROC for lithic sites

Figure 4: ROC curves for various choices of γ.

tie-breaking model. For different choices of false positive percentage (or true positive percentage) we see that the combined model typically improves performance over using either the APM and the WorldView analysis alone. The consistent superior performance of the combined model to APM, indicate that adding-in our method increases prediction power over using the APM alone.

4.1 Discussion

In this paper we consider one particular example of a statistical approach to archaeological site discovery using multi-spectral imagery. In particular we consider the framework specified by three components: an image processor, feature extractor, and classifier. Specifically, we use a relatively simple image processor in the annuli technique and common feature extractors and classifiers in PCA and LDA.

We demonstrate that our method provides performance which is better than chance and which provides additional improved performance when used in conjunction with an archaeological predictive model. It gives strong evidence that including our method in archaeological prediction analysis improves classification accuracy. We do not claim that these represent optimal results but rather that this framework provides a valuable way to attach the problem of archaeological site discovery. Almost certainly there are other methods for each component which will give a further improvement in performance.

For example, the archaeological sites will have some known properties which will help in choosing an image processor. For example, it might be desirable to have f_b invariant to rotations, changes in scale, or changes in lighting. On the other hand, if we are attempting detect sites with a certain aspect, size, or brightness, then these invariance properties would be detrimental. There is a rich literature regarding useful image processors for remote sensing;[9] however, no image processor is optimal for every task so using prior knowledge about the sites and the imagery will be critical.

We used PCA for feature extraction and applied LDA on the reduced space. There is a wealth of empirical evidence that the composition of these two procedures often gives improved results over using LDA on the original feature space. One may also use a different feature extractor and predict the data with classifiers such as support vector machine, boosting, k-nearest neighbor. The literature on feature extraction and classification is vast and much improvement is likely possible within this basic framework.

In addition to changing the components, we also suspect that including additional feature bands, including especially slope data and possibly data related to aspect and other topographical properties, will improve performance. Future work will consider this framework with these additional features.

5. CONCLUSION

We have reported here on our latest work in generating maps of probable archaeological sites. This work follows up on a previous study that utilized the Student's t-test. We report here on our work with a new, more flexible approach utilizing principle components analysis followed by linear discriminant analysis. In a future study we hope to compare the effectiveness and ease of use of the two approaches, as well as how they might most effectively be combined. We compared our new approach to an archaeological predictive model (APM), which has been considered as the standard means for identifying areas that are likely to contain archaeological sites. While the APM had a better performance than our new method for habitation sites, our method outperformed the APM for lithic sites and a combination of our method with the APM outperformed either method by itself. Further, we have determined that the statistical treatments that we employed in this study can be used to greatly improve APMs. By combining improved APMs and the image analysis protocols that we have develop, we are optimistic that we can develop even more productive methods of finding archaeological sites in the future.

We have demonstrated the effectiveness of our new approach over a portion of the Ft. Irwin military reservation in California. We plan to extend our work to other portions of Ft. Irwin and also to the nearby China Lake military reservation, with the ultimate objective of establishing the capacity to conduct rapid, wide-areas surveys for archaeological sites with results as reliable, or nearly so, as those obtained from on-ground surveys for archaeological sites. We also will be working to gain acceptance of this survey approach by the Advisory Council on Historic Preservation and State Historic Preservation Offices as a substitute for such time-consuming and more expensive on ground surveys, or to direct on-ground efforts in ways that will great reduce the time and expense required for such surveys.

ACKNOWLEDGMENTS

REFERENCES

[1] Comer, D. C. and Blom, R. G., "Detection and identification of archaeological sites and features using synthetic aperture radar (sar) data collected from airborne platforms," in [*Remote Sensing in Archaeology*], Wiseman, J. R., El-Baz, F., ed., 103–136, Springer Science + Business Media, LLC, New York (2007).

[2] Comer, D. C. and Blom, R. G., "Remote sensing and archaeology: Tracking the course of human history from space,"

[3] Comer, D. C. and Blom, R., "Wide area inventory of archaeological sites using aerial and satellite data sets: Prologue to resource monitoring and preservation," *Proc. 32nd Int. Symp. on Remote Sensing of Environment* (June 2007).

[4] Marchisio, G., Padwick, C. and Pacifici, F., "Evidence of improved vegetation discrimination and urban mapping using worldview-2 multi-spectral imagery," *ASPRS Annual Conference* (May 2010).

[5] Ruiz, M. O., "The development and testing of an archaeological predictive model," tech. rep., U.S. Army Corps of Engineers Construction Engineering Research Laboratory, Champlain, Illinios (2003).

[6] Kvamme, K. L., "Development and testing of quantitative models," in [*Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*], Judge, W. J. and Sebastian, L., eds., 325–428 (1988).

[7] Duda, O. R., Hart E. P. and Stork, D. G., [*Pattern Classification*], Wiley, 2nd ed. (2001).

[8] Tan, P.-N., Steinbach, M. and Kumar, V., [*Introduction to Data Mining*], Addison Wisley (May 2005).

[9] Schowengerdt, R. A., [*Remote Sensing: Models and Methods for Image Processing*], 3rd ed. (September 2006).