Histogram Curve Matching Approaches for Object-based Image Classification of Land Cover and Land Use

Sory I. Toure, Douglas A. Stow, John R. Weeks, and Sunil Kumar

Abstract

The classification of image-objects is usually done using parametric statistical measures of central tendency and/or dispersion (e.g., mean or standard deviation). The objectives of this study were to analyze digital number histograms of image objects and evaluate classifications measures exploiting characteristic signatures of such histograms. Two histograms matching classifiers were evaluated and compared to the standard nearest neighbor to mean classifier. An ADS40 airborne multispectral image of San Diego, California was used for assessing the utility of curve matching classifiers in a geographic object-based image analysis (GEOBIA) approach. The classifications were performed with data sets having 0.5 m, 2.5 m, and 5 m spatial resolutions. Results show that histograms are reliable features for characterizing classes. Also, both histogram matching classifiers consistently performed better than the one based on the standard nearest neighbor to mean rule. The highest classification accuracies were produced with images having 2.5 m spatial resolution.

Introduction

Remotely sensed images have been most commonly classified using pixel-based approaches to generate maps of land-use and land-cover (LULC). With such approaches, classification is primarily based on the spectral characteristics of individual pixels. In a supervised classification, a pixel is identified or labeled as the class to which it has the most similar spectral signature relative to training data signatures (Mitchell, 2008). Pixel-based approaches to image classification normally do not incorporate information about spatial location and geographic association, and output classification products commonly contain spurious pixels that are misclassified and/or smaller than the minimum mapping unit. Moreover, there is an implicit assumption that the spatial resolution of the image data is similar to or coarser than the classification features of interest (Goodchild, 1994). However, this is not true for most high spatial resolution imagery (Castilla *et al.*, 2008). When classifying high spatial resolution imagery for urbanizing environments, Cleve *et al.* (2008) found that a pixel-based approach was 41.74 percent less accurate than an object-based approach for the built area category.

Object-based classification is a useful alternative to traditional pixel-based approaches, as groups of pixels or image objects are classified as an ensemble rather than individually (Blaschke, 2009). As a component of geographic object-based image analysis (GEOBIA), adjacent pixels that meet some similarity criteria are grouped to form image segments or objects (Haralick and Shapiro, 1985). Image objects are subsequently classified after they are delineated *vis-à-vis* a segmentation routine. Spectral as well as spatial and contextual information are available for the classification of objects relative to single pixels (Blaschke, 2009).

Classification of segments is usually based on statistical measures of central tendency (e.g., mean) and dispersion (e.g., variance or standard deviation) associated with normally distributed (i.e., Gaussian) data. Walter (2003) utilized the mean grey value of the blue, green, red, and near infrared (NIR) bands, their variances, the mean grey value texture from the blue band, the mean grey value and variance vegetation index, the variance texture, and four others (sixteen in total) as input features while using an object-based classification approach of remote sensing data for the purpose of change detection. High spatial resolution images tend to be spatially heterogeneous, particularly for built landscapes that are composed of heterogeneous mixtures of surface material and land-cover types. The frequency distributions of such data often yield within-segment pixel groups having complex and non-normal frequency distributions. Thus, classifying segments having heterogeneous pixel data with traditional classification methods that assume a normal distribution of the data may not be statistically appropriate or accurate (Stow *et al.*, 2012). Wu *et al.* (2007) classified urban land uses using a per-field approach based on parcel boundary data to delineate image objects (rather than through image segmentation) and tested a variety of classification algorithms. They found that decision tree and other classifiers that do not require assumptions regarding the statistical properties of the input data, such as parallelepiped and neural network, performed better compared to Mahalanobis distance and maximum likelihood approaches, algorithms that are based on normality assumptions. Thus, alternative classification approaches should be developed and tested for the mapping of heterogeneous scenes such as urban areas when classifying high spatial resolution data. Curve matching techniques may be appropriate

Sory I. Toure, Douglas A. Stow, and John R. Weeks are with the Department of Geography, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493 (stoure@rohan.sdsu.edu).

Sunil Kumar is with the Department of Electrical and Computer Engineering, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493.

Photogrammetric Engineering & Remote Sensing
Vol. 79, No. 5, May 2013, pp. 433–440.

0099-1112/13/7905–433/$3.00/0

for the classification of within segment pixel frequency distributions, especially when based on high spatial resolution data of built environments (Stow *et al.*, 2012).

In remote sensing, curve matching techniques have been used for classification of hyperspectral image data, by measuring curve similarity between a known spectrum from a spectral library or training pixels with the spectra of individual pixels within an image (van der Meer, 2006). Mineralogies and lithologies are commonly mapped by quantitatively comparing surface reflectance curves from imaging spectrometer data set with known diagnostic reflectance spectra of minerals and rock units using spectral matching techniques (e.g., Kruse *et al.*, 1993). Histogram matching is also applied to entire images as an approach to radiometric normalization of multitemporal image data sets (Coulter *et al.*, unpublished data, 2009).

Curve matching techniques for the purpose of classification have been used in fields other than remote sensing. In computer vision, color is one of the most popular and effective low-level cues for content-based image retrieval (CBIR) (Qui *et al.*, 2004). Most often, color is represented through diverse forms of histograms. Swain and Ballard (1991) proposed a color indexing algorithm to identify color images. Histograms are generated for the input and target images in a database. Images are then matched using the histogram intersection method. To query image and video databases, Saykol *et al.* (2005) generated histograms of the color and shape of objects that needed to be retrieved. The two histograms were then integrated and a similarity measure (histogram intersection) was computed between the query object and a database object. Histogram matching can be implemented through template matching procedures by searching for image objects similar to training templates (Brunelli and Milch, 2001). Thus, histogram shapes can be used as the basis of similarity measures between training and target objects and for object classification.

As a precursor to this study, curve matching techniques were tested by Stow *et al.* (2012) for the classification of LULC of the city of Accra, Ghana. They examined the within and between class variability of frequency distributions of QuickBird multispectral digital numbers (DN) values of within-segment pixels to determine if within-class signatures are similar and between-class curves are separable. Stow *et al.* (2012) evaluated a quantitative measure and classification approach that exploits characteristic frequency distributions (i.e., histogram signatures) of within-segment pixels. Generally, LULC classes exhibited characteristic histograms and the curve matching approach performed slightly better than a nearest neighbor classifier. The nature of the Stow *et al.* (2012) study was exploratory. The authors used a limited sample size with general LULC classes. Only a single curve matching algorithm was analyzed. The objective of this study is to further examine the between- and within-class variability of image object histogram and evaluate object-based classification techniques, primarily with histogram matching classifiers, using very high spatial resolution image data. Airborne multispectral data of the City of San Diego, California was evaluated in the context of mapping five LULC classes. The general objective of the study was to exhaustively test the viability of histogram matching techniques as effective classifiers within GEOBIA.

The following questions were examined as part of this study:

1. Do LULC objects derived from high spatial resolution multispectral images have non-normal digital number frequency distributions?
2. Are the histograms of LULC objects consistent and uniquely characteristic?
3. Does the classification of image objects representing LULC objects by histograms matching algorithms show potential to improve classification accuracy relative to a conventional nearest neighbor classifier?
4. Is the histogram matching classification approach sensitive to differences in spatial resolution of image data?
5. Do different quantization levels influence histogram matching classification results?
6. Which feature inputs improve classification accuracy?

Study Area and Data

The full extent of the study area covers about 152 km^2 and encompasses the communities of La Jolla, Clairemont, and Mira Mesa, within the City of San Diego, as well as Descanso, all within San Diego County, California. Figure 1 depicts the study area and the location of the study sites. The geographic coordinates of the rectangular study area are 32°46'13" to 32°56'38" N, and 116°46'15" to 117°12'44" W. It contains residential, light industry, schools, office parks, and other open space land-use (parks and golf courses). The variety of LULC in this area is representative of San Diego County in general. Subsets of the aerial imagery contained in Figure 1 were generated and utilized for testing per-object classification approaches.

The physical geography and built environments of the study area are typical of those found throughout southern California. Topographic features consist of mesas, canyons, and coastal plains. A Mediterranean-type climate is prevalent, with rainy winters and mild temperatures year round. The terrain and climate type have facilitated the settlement of a large and diverse human population. The vegetation is composed mainly of chaparral and coastal sage shrub vegetation communities.

An ADS40 airborne multispectral image of San Diego, California was utilized for assessing the utility of curve matching classifiers applied to image objects in the context of LULC classification. Multispectral image data were captured by EarthData (now called Fugro EarthData, Inc.) and made

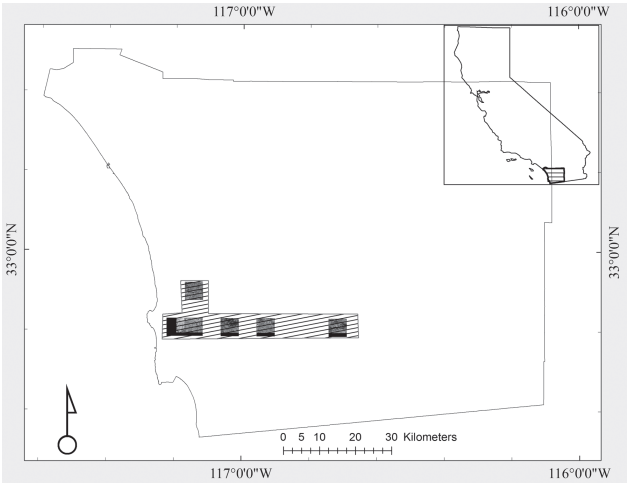

Figure 1. Study area map within San Diego County in California. The shaded rectangular areas delineate the extent of coverage of the ADS40 imagery, and the individual frames depict locations of the specific image frames used for object-based classification trials.

available *pro bono* to public institutions in San Diego County. The Leica ADS40 sensor is based on a linear array radiometer and yields digital orthoimagery through a softcopy photogrammetry workflow. Image data were acquired in October and November 2003. The imagery covers the areas burned in the Cedar and Paradise fires as well as a substantial amount of peripheral unburned areas, which was used for this study. The ADS40 data were originally provided in UTM projection with a spatial resolution of 0.5 m. The data were reprojected to State Plane coordinates with a ground sampling distance of about 0.5 m. The multispectral data set includes four broad wavebands in the blue, green, red, and near infrared (NIR) portion of the electromagnetic spectrum and has an original radiometric resolution of 11 bits. Only the red and NIR bands were used in the study. The data set had been compressed from 11 to 8 bits.

A geographic information system (GIS) layer depicting LULC distributions in vector format and representing 2004 conditions had been generated by the San Diego Association of Governments (SANDAG) and was utilized to extract training data for analysis of LULC histogram signatures and for LULC classifier tests. The most commonly occurring LULC classes within the study area are: (a) Single Family Residential/Mobile home parks (SFR), (b) Multi-Family Residential (MFR), (c) Parks/Golf courses, (d) Schools, and (e) Non-residential Urban. The Non-residential Urban class consists of three subclasses: Commercial, Industrial park, and Office High Rise (OHR).

Methods

Since the study emphasizes the classification phase of GEOBIA, sample image objects or segments were selected and delineated manually rather than through semi-automated segmentation. The Interactive Data Language (IDL) programming tool within ENVI 4.8, an image processing and analysis software package, was used to develop and test histogram curve matching classifiers. Figure 2 describes the processing flow. Training and testing objects or regions of interest (ROI) were interactively (manually) delineated using ENVI tools. Two histogram matching algorithms: Histogram Matching Root Sum Squared Differential Area (HMRSSDA) and Histogram Angle Mapper (HAM), as well as the conventional nearest neighbor (or minimum distance in feature space) classifiers, were implemented with IDL.

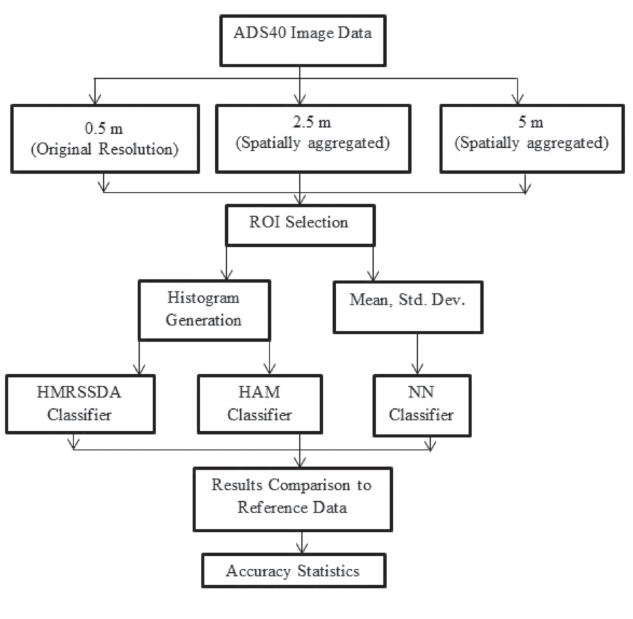

Figure 2. Processing flow - HMRSSDA: Histogram Matching Root Sum Squared Differential Area (HAM: Histogram Angle Mapping; NN: Nearest Neighbor).

Image objects derived from the 2003 ADS40 image were classified into one of the seven (when considering each Non-residential Urban class individually) LULC classes defined above. The image has an original spatial resolution of 0.5 m and was spatially aggregated using simple pixel block averaging to create coarser images having 2.5 m and 5.0 m spatial resolution (representing a range spatial resolutions for commercial satellite multispectral data), to examine sensitivity of histograms and curve matching classification to spatial resolution. During the conversion procedures, the new pixel value represents the average of the spatially aggregated pixel values.

Generation of LULC Objects, their Histograms and Statistics

Regions of interest (ROI) representing the seven LULC classes were interactively interpreted and digitized from the ADS40 images with reference to the SANDAG vector file. The LULC polygons of a particular class were overlaid on the images and ROIs for that class were manually digitized with the Region of Interest tool in ENVI. Many of the polygons representing park/golf courses and single family residential (SFR) classes were large, in which case ROIs were digitized from within the polygons. For the remaining classes, ROIs were delineated using the original polygon boundaries of the SANDAG LULC vector layer. Table 1 shows the number of ROIs and their size for each class. Between (13) and (50) ROIs were selected for each LULC types. The total number of ROI amounts to 183. The number of ROIs varied for each class because of their relative occurrence within the study area. SFR was the most prevalent class and School the least frequently occurring. The same ROIs polygons were used to extract pixel groups for all three spatial resolution image sets.

Once ROIs were delineated, DN values for the red and NIR bands were extracted for all pixels belonging to a ROI and statistics were generated. The statistics included the mean and standard deviation. Histogram curves for each ROI were normalized by dividing the frequency value of each DN by the total number of pixels contained within that ROI. Thus, the frequency distribution of DNs for each ROI varies between 0 and 1. An analyst visually analyzed the curves shapes and normality of distributions, and checked for histogram characteristics and similarities within a class and dissimilarities between classes.

After generating histograms and statistics for all ROIs, the ROIs were separated into two categories: one group for training and the other for testing. The training histogram curve of a particular class was selected after evaluating all histograms for that class. Following Stow *et al.* (2012), three ROI samples were averaged to generate training curves, by computing the average of the three DN values for each bin. The three ROIs were chosen that were most similar and representative of a given class. The Park/Golf Courses class had important variability in its histograms. Although they were bell-shaped, the peaks of the curves generally occurred between two separate and distinct intervals. Two means histograms that represented the two trends (subclasses) for the Park/Golf Courses class were therefore generated. The remaining classes were represented with one mean histogram. The mean ROIs were kept the same for the red and NIR bands, as well as for the 0.5 m, 2.5 m, and 5 m spatial resolutions. After selecting ROIs to be used as training curves, the remaining ROIs were utilized for testing purposes. Table 1 presents the number and size of the ROIs as well as a summary of their utilization for each class.

Table 1. Number and Size of Test and Training Regions of Interest (ROI) per Class; Total Areal Extent of the Study Area is 152 km²

Land-Cover / Land-Use Classes	No. Training ROIs	Area Covered by Training ROIs (km²)	No. Testing ROIs	Area Covered by Test ROIs (km²)
Commercial	3	0.04	22	0.18
School	3	0.27	10	0.51
Industrial Park	3	0.12	27	1.49
Multiple Family Residential	3	0.07	30	1.41
Office High Rise	3	0.07	27	0.49
Park	6	0.19	20	0.61
Single Family Residential	3	0.04	47	0.95

Classification Algorithms

Two curve matching classification measures were utilized to classify histograms of the test LULC ROIs for the Red and NIR bands: Histogram Matching Root Sum Squared Differential Area (HMRSSDA) and Histogram Angle Mapper (HAM). The formulae for these measures are:

$$\text{HMRSSDA: } \sqrt{\sum_{i=0}^{i=q}(FR_{i^2} - FR_{i^2})^2} \tag{1}$$

$$\text{HAM: } \alpha = \cos^{-1}\left[\frac{\sum_{i=0}^{i=255} FS_i * FR_i}{\left(\sum_{i=0}^{i=255} FS_{i^2}\right)^{\frac{1}{2}}\left(\sum_{i=0}^{i=255} FS_{i^2}\right)^{\frac{1}{2}}}\right] \tag{2}$$

where: FS_i = frequency of subject histogram at bin i = DN, FR_i = frequency of reference histogram at bin i = DN, and q = quantization level (i.e., number of DN bins).

The HAM routine represents a modification of the original formula for the spectral angle mapper algorithm used in classification of hyperspectral image data, in which *FS* is an unknown histogram curve and FR is a reference/training curve (Park *et al.*, 2007). The combination of feature inputs can potentially improve classification results. Therefore, the curve matching scores for the Red and NIR bands combined were computed and compared using the arithmetic, geometric, and Pythagorean means as follows:

$$\mu_{\text{Arithmetic}} = (\text{Red} + \text{NIR})/2 \tag{3}$$

$$\mu_{\text{Geometric}} = (\text{Red} \times \text{NIR})^{1/2} \tag{4}$$

$$\mu_{\text{Pythagorean}} = (\text{Red}^2 + \text{NIR}^2)^{1/2} \tag{5}$$

where Red and NIR = histogram matching or Nearest Neighbor score derived for the red and NIR waveband data, respectively.

The test image object data were also classified with the nearest neighbor algorithm, a traditional classifier. Objects means and standard deviations were used as inputs. We measured similarity by the Euclidean distance (d(fs,fr)):

$$d(fs, fr) = \sqrt{\left(FS^2 - FR^2\right)}$$

where *FS* = mean and/or standard deviation of the subject , and *FR* = reference mean or standard deviation. Objects mean (μ) and standard deviation (σ) values were also combined to perform this classification (NN). The formulae for computing mean scores from multiple band inputs were:

$$\text{Quadratic Mean: } \sqrt{\frac{\mu_{red}^2 + \mu_{nir}^2 + \sigma_{red}^2 + \sigma_{nir}^2}{4}} \tag{6}$$

$$\text{Power Mean: } \sqrt[4]{\frac{\mu_{red}^4 + \mu_{nir}^4 + \sigma_{red}^4 + \sigma_{nir}^4}{4}} \tag{7}$$

$$\text{Geometric Mean: } \sqrt{\frac{\mu_{red} * \mu_{nir} * \sigma_{red} * \sigma_{nir}}{4}}\,. \tag{8}$$

In addition to testing the performance of the three classification measures, two other variables were analyzed, image spatial resolution, and histogram quantization level. Images with 0.5 m (original data), 2.5 m and 5.0 m (aggregated or coarsened) spatial resolutions were evaluated. The ADS40 data used in this study have a radiometric resolution of 8-bits, such that DN values have a potential range from 0 to 255, or 256 bins. In order to analyze the effect of the apparent radiometric precision (i.e., number of histogram bins) and the utility of bin aggregation on reducing noise effects on classification results, the original quantization levels of ROI data values were reduced to produce data sets having reduced quantization levels of 128, 64, and 32. The reduction of quantization level or number of histogram bins was achieved through bin aggregation using a script written in IDL.

Classification Analyses

Test ROIs were subjected to the two histograms curve matching procedures and to the nearest neighbor classifiers based on LULC templates derived from the reference/training object curves. With multiple spatial resolutions (three), quantization levels (four), classifiers (three) and approaches to combining multiple band (three) and statistical measures (three), 159 accuracy/error matrices were generated through the classification trials. For each error matrix, the overall accuracy, producer's and user's accuracy of the individual classes, KHAT statistic and its variance, as well as conditional kappa for each class and their variances were generated. Conditional Kappa is used to assess the agreement for an individual class within an error matrix (Congalton and Green, 2009).

A Z-test was used to test for the significance of a classification. If the Z-value is greater than 1.96 at the 0.05 significance level, the classification product is more accurate than a random classification (Congalton and Green, 2009). A Z-test was also performed to test if any two error matrices were significantly different from each other. Within an analysis subsection, only the largest difference between error matrices was tested for significance when the test result showed that they were not significantly different. This was done because the non-significance of the biggest difference between accuracy results implies that smaller differences are also not significant. Two classification results are significantly different from each other if the Z-value is also greater than 1.96 (Congalton and Green, 2009). The objective was to determine which spatial resolution, quantization level, and classification algorithm yielded the most accurate classification results. The same statistical tests were run for each class using their conditional kappa coefficients and variances to examine individual class accuracies.

Results

Histogram Curve Characteristics

Figures 3a, 3b, and 3c present the mean value histogram curves for the red and NIR wavebands from training data for all seven classes and associated subclasses. The SFR and Park classes are normally distributed while the Commercial, Industrial Park, MFR, School, and OHR classes have non-normal histograms. MFR histograms are broad and positively skewed. School histograms are highly skewed to the left, broad and dispersed. Their curves have a staircase like shape with a main peak occurring at about DN value of 180.

Although both the Park and SFR classes have reasonably symmetric histograms, their DN value distributions are different. The SFR distribution is platy-kurtic, whereas the Park distributions are meso- or even lepto-kurtic. The histograms of the Park class exhibited considerable variability in their shape. The Park class was therefore represented by two sub-classes with different training curves: Park 1 and Park 2. Both histograms are normally distributed, with peaks occurring at different digital number values, approximately 80 for Park1 and 140 for Park 2.

Among the classes with non-normal histograms, the Industrial Park and OHR exhibit similar shapes. Their histograms are platy-kurtic and tend to be bimodal, although some have more than two modes. The Commercial class histograms, however, have three general peaks, are broad and highly dispersed. Histogram curve shapes are consistent between the Red and NIR band for all classes.

Overall Classification Results

The ranges of overall classification accuracy results varied between 57.0 percent to 70.0 percent for HMRSSDA, 58.0 percent to 71.6 percent for HAM, and 27.0 percent to 73.0 percent for NN. The median scores are 67.21 percent, 67.76 percent, and 47.0 percent for the HMRSSDA, HAM, and NN algorithms, respectively. The histogram matching algorithms yielded higher accuracies than the conventional nearest neighbor to mean classifier. Seventy-five percent of all nearest neighbor results are below the minimum scores of the histogram matching algorithms, and about 95.0 percent (37/39) of them have overall accuracy estimates of less than 65.0 percent. The combination of the Red and NIR bands produced higher accuracy results as compared either single band, and histograms with fewer bin sizes generally yielded higher accuracies. The spatial resolution of 2.5 m also tended to yield the highest scores.

Classifier Influence on Classification Accuracy

A primary objective of this study was to assess how well histogram matching classifiers performed relative to the standard nearest neighbor to mean classifier. As shown in Figure 4, both histogram matching algorithms performed better than the Nearest Neighbor classifier when spatial resolutions, bin size, and input band combinations were the same. The mean kappa accuracy for all resolutions and bin sizes are 56.6 percent for the HMRSSDA, 58.9 percent for the HAM and 31.1 percent for NN. However, the single highest overall accuracy value of 72.7 percent was produced with the Nearest Neighbor classifier, with the Quadratic waveband mean combination formula at a spatial resolution of 2.5 m.

Of the two histogram matching algorithms tested, the HAM outperformed the HMRSSDA 53/60, or 88.0 percent of the time. The greatest difference of 8.4 percent in accuracy between the two algorithms occurred at 5 m, with the bin size of 128, and the

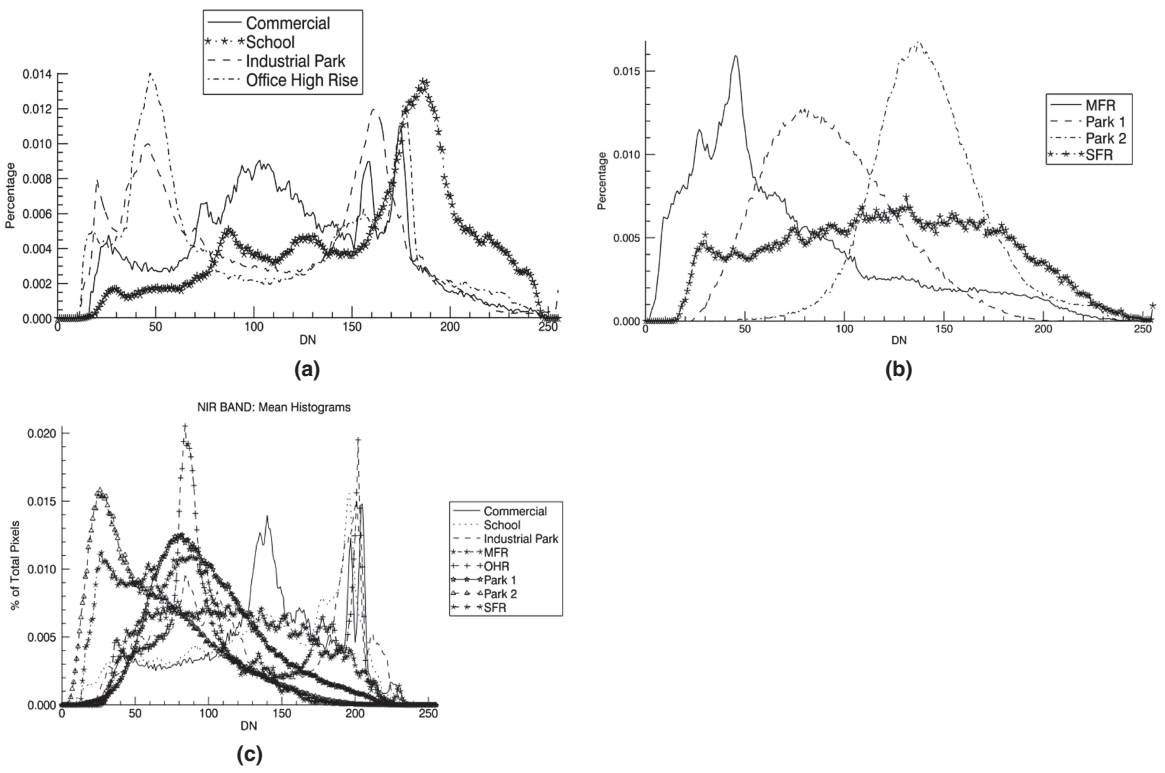

Figure 3. (a) and (b) red waveband histograms for training objects, and (c) NIR waveband histograms for training objects.

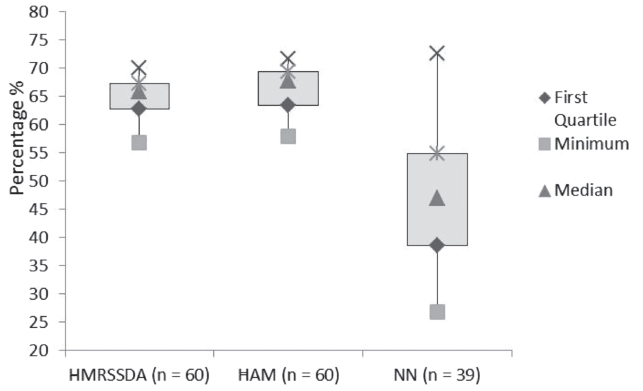

Figure 4. Comparison of frequency distribution of overall accuracy figures for the three classification measures.

arithmetic combination formula. With a sample size of 183, the difference is not significant based on the z-test ($z = 1.33 < 1.96$). Thus, the smallest difference is also not significant. Table 2 compares the performance of the histogram matching algorithms relative to the three spatial resolutions. The biggest difference of 3.7 percent occurred with 5 m resolution images where histograms are coarser. This suggests that the HAM algorithm is less sensitive to noise present in histograms.

Spatial Resolution Influence on Classification Accuracy

The smoothness of histograms increases with an increase in spatial resolution; histograms are smoother at 0.5 m than at 5 m due to the larger sample sizes associated with a greater number of pixels per ROI. It was hypothesized that higher spatial resolutions would yield higher accuracy results for histogram matching algorithms. Object classifications based on 2.5 m spatial resolution images yielded the highest accuracies 77.3 percent (41/53) of the time as compared to those from the highest (0.5 m) resolution images. The greatest difference of 33.6 percent in accuracy due to different spatial resolution inputs occurred with the nearest neighbor algorithm based on the power mean combination formula. This difference is significant based on the z-test at the 5.0 percent significant level ($z = 5.19 > 1.96$). Moreover, the 2.5 m outperformed the 5 m spatial resolutions 71.7 percent (38/53) of the time. The greatest difference of 25.3 percent was also significant based on the z-test ($z = 3.69 > 1.96$). It occurred with the nearest neighbor algorithm based on the power mean combination formula. Finally, classification results for the 5 m spatial resolution products were higher than those in the 0.5 m 58.5 percent (31/53) of the time. The greatest difference of 14.1 percent in accuracy results between these two spatial resolutions (5 m and 0.5 m) occurred with the nearest neighbor algorithm based on the quadratic combination formula. The results were significantly different based on the kappa z-test at the 0.05 level of significance with a sample size of 183 ($z = 2.04 > 1.96$).

Table 2. Mean Overall Accuracy Results for Histogram Matching Classifiers

Spatial Resolution	HMRSSDA	HAM
0.5 m	64.5%	65.6%
2.5 m	65.5%	66.5%
5 m	63.9%	66.6%

Quantization-level Influence on Classification Accuracy

Classifications based on histogram matching routines were performed with four bin sizes: 256, 128, 64, and 32. The highest scores were achieved at 5 bit quantization (32 bins) 19 out of 30 (63.3 percent) times and was highest or shared the highest scores 80.0 percent (24/30) of the time. At 6-bit quantization (64 bins) the highest accuracy was achieved 26.7 percent (8/30 - four singles and four ties) of the time. The greatest difference of 5.2 percent in result due to bin size only occurred at 5 m, with the HMRSSDA matching algorithm and the geometric combination formula. The results are not significantly different based on the z-test at the .05 level of significance ($z = 0.83 < 1.96$). While the highest classification accuracy of 65.3 percent of the study was achieved for histograms with 64 and 128 bins, 32 bin histograms yielded a majority of the highest classification accuracies. When considering individual classes, the bin sizes did not generally affect results. The reduction of bin sizes smoothes out spikes in histograms which are then efficiently classified by the algorithms.

Waveband Influence on Classification

The classification of image object histograms was performed with the individual red and NIR bands and a combination of the two. Figure 5 depicts the results of the comparison of these band inputs. When holding spatial resolution, quantization level, and classifier constant and considering the Non-residential urban class as a single class, classification results for the red band consistently ranked higher than that for the NIR. The red band yielded higher accuracies 93.3 percent (28/30) of the time than the NIR band. The greatest difference of 19.1 percent between the results occurred at 2.5 m, with the nearest neighbor classifier based on the standard deviation. With a sample size of 183, this difference was significant at the 0.05 level based on the KHAT z-test ($z = 3.16 < 1.96$).

All three different formulae for combining red and NIR bands outperformed either single band. The arithmetic mean outperformed the geometric mean formula 60.0 percent (18/30) of the time, while the Pythagorean outperformed both the arithmetic and geometric 70.0 percent (21/30) and 86.7 percent (26/30) of the time, respectively. The products based on the Pythagorean combination formula yielded most of the highest classification accuracies based on the KHAT statistics, followed by the arithmetic and geometric formula, respectively.

Individual Class Accuracy and Separability

The Commercial and OHR classes have high conditional kappa scores with the HMRSSDA and HAM algorithms while their scores are generally low when classifying with the standard nearest neighbor measure. Fifty percent of all Industrial Park scores range between 40 to 50.0 percent for both histogram matching measures while they are below 30.0 percent for the nearest neighbor classifier. The Park class was accurately classified with all three algorithms. It was uniquely separable in most cases. Results for the school class were consistent between histogram matching algorithms: 50.0 percent of all results were between 47.0 percent and 57.0 percent. Its conditional kappa scores ranged from below 0.0 percent to above 80.0 percent when classifying with the nearest neighbor. The MFR and SFR classes also exhibited consistent and higher accuracies with the histogram matching classifiers compared to the nearest neighbor measure.

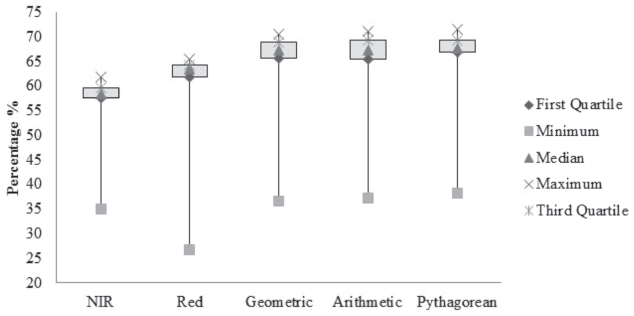

Figure 5. Comparison of overall accuracy for various band inputs and modes of combining band inputs.

Discussion and Conclusions

The purpose of this study was to analyze histograms of image objects corresponding to LULC polygons and evaluate classification approaches exploiting histogram curve matching. It was found for the specific study area context and imagery type that LULC classes generally exhibited unique and characteristic histogram signatures that could be exploited for object-based image classification purposes. This study extends the findings of a previous study conducted by Stow *et al.* (2012). The purpose of that study was to examine the within- and between-class variability of digital number histograms extracted from ROIs representing known LULC types to determine if within-class signatures were similar and between-class curves were separable, and to evaluate a quantitative measure and classification approach that exploits histogram of within-segment pixels. The only histogram matching algorithm considered in the Stow *et al.* (2012) study was the HMRSSDA. The data inputs were the normalized difference vegetation index (NDVI), normalized difference pavement index (NDPI), and a combination of these spectral indices, NDVI + NDPI. The authors found that ROI histograms could be consistent and characteristic for LULC classes. The curve complexity was dependent on landscape characteristics, classification objectives, and the spatial resolution of the input data. The overall classification results of the histogram matching classifiers were slightly better or identical to a nearest neighbor classifier. With more training and testing of a dataset, this study expanded on the Stow *et al.* (2012) study by incorporating another histogram matching classifier (HAM) and two more variants of the nearest neighbor classifier, three different spatial resolutions, and four different quantization levels.

The first two questions addressed in this study were related to the normality and characteristic shapes of image-object histograms. After analyzing the histograms of the seven LULC classes considered in this study, it was found that histograms were reliable features for characterizing classes. The shapes of the histograms depend primarily on the land-cover composition within land-use units. Vegetation and land-cover histograms tended to be normally distributed while land-use categories presented more complex histograms. Land-use classes that have some amount of vegetation present curves that are unimodal but skewed. Non-residential Urban classes exhibit bimodal, non-normal histograms while the histograms of the Park and SFR classes are normally distributed. Schools and MFR histogram curves are weakly Gaussian and exhibit various degrees of skewness.

The next research question dealt with the performance of the histogram matching classifiers relative to the nearest neighbor classifier. While the Stow *et al.* (2012) study showed that curve matching classifiers could yield higher classification accuracies than a nearest neighbor classifier for the classification of classes with complexly-shaped histograms, it was found in this study that both histogram matching measures that were tested consistently yielded more accurate classification results than a standard nearest neighbor with results that were significantly different based on the kappa statistics. Between the histogram matching algorithms, the HAM yielded the highest results although the differences between the two algorithms were not significantly different based on the kappa statistics.

Next, a question on the effects of spatial resolutions on classification results was also addressed. Results show that images with spatial resolution of 2.5 m yielded classifications with the highest accuracies. The 5 m spatial resolution produced the next highest accuracies followed by 0.5 m. Further research could incorporate data sets with spatial resolutions higher than 0.5 m and lower than 5 m to further evaluate the effects of spatial resolution on accuracy results of curve matching classifiers and nearest neighbor.

Moreover, histogram curve matching classifications were tested at four quantization levels: 256, 128, 64, and 32. The 5-bit quantization (32 bins) generally yielded the highest classification accuracies. In general, histograms with fewer bins yielded higher classification accuracies. This is probably due to the generalization (i.e., smoothing) of the curves that occur when bins are aggregated.

A combination of band inputs produced higher accuracy results than single bands. This suggests that the use of more information channels is preferable. The choice of band combination formula is also important as some consistently generate higher results that others. Finally, we found that classes with complex histogram shapes such as commercial or OHR were better classified with histogram matching classifiers as compared to the nearest neighbor.

Training objects were selected visually and interactively and finalized upon analyzing the histograms curves of all ROIs belonging to a class. In a normal classification setting, only a sample of the data is analyzed, and the characteristics of that sample are applied to the classification of the whole data set. A similar approach could be used in the selection of training histograms. An interpreter could select few ROIs for a particular class based on interpretation of the study area and use them to create mean curve(s) for that class. In case it appears that subclasses are needed, histograms could be grouped based on a quantitative measure or tool. A dendrogram could be such a tool.

Histogram matching algorithms were evaluated by comparing kappa accuracy results. While comparing the effectiveness of spectral similarity measures for the analysis of hyperspectral imagery; van der Meer (2006) used the probability of spectral discrimination (PSD) and the power of spectral discrimination (PWSD) to assess the effectiveness of the spectral similarity measures. More robust measures such as PSD and PWSD should be tested in future studies that compare histogram matching algorithms. Also, the use of the kappa coefficient as a measure of assessing the accuracy of remotely sensed data classification is not universally accepted because the degree of chance agreement may be overestimated (Foody, 1992; Pontius and Millones, 2011). However, one of the advantages of the kappa coefficient is its ability to test for significance between two error matrices.

References

Blashke, T., 2009. Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65:2–16.

Brunelli, R., and O. Milch, 2001. Histograms analysis for image retrieval, *Pattern Recognition*, 34: 1625–1637.

Castilla, G., G.J. Hay, and J.R. Ruiz, 2008. Size-controlled region merging (SCRM): An automated delineation tool for assisted photointerpretation, *Photogrammetric Engineering & Remote Sensing*, 74(4):409–419.

Cleve, C., M. Kelly, F.R. Kearns, and M. Moritz, 2008. Classification of the wildland-urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography, *Computers, Environments and Urban Systems*, 32:317–326.

Congalton, R.G., and K. Green, 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press, Boca Raton, Florida, 192.

Coulter, L., D. Stow, J. Weeks, A. Hope, M. Tsou, and P. Jankowski. 2009. *A Border Security Spatial Decision Support System Driven by Remotely Sensed Data Inputs*, Hampton, Virginia: A NASA Research, Education, and Applications Solution Network (REASoN) Project Final Report Publication.

Foody, G., 1992. On the compensation for chance agreement in image classification accuracy assessment, *Photogrammetric Engineering & Remote Sensing*, 47(3):343–351.

Goodchild, M.F., 1994. Integrating GIS and remote sensing for vegetation analysis and modeling: Methodological issues, *Journal of Vegetation Science*, 5:615–626.

Haralick, R.M., and L.G. Shapiro, 1985. Image segmentation techniques, *Computer Vision Graphics and Image Processing*, 29:100–132.

Kruse, F.A., A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P. . Barloon, and A.F.H. Goetz, 1993. The spectral image processing system (SIPS) - Interactive visualization and analysis of imaging spectrometer data, *Remote Sensing of Environment*, 44:145–163.

Mitchell, S.W., 2008. Distance to second cluster as a measure of classification confidence, *Remote Sensing of Environment*, 112(5):2615–2626.

Pontius, R.G., Jr., and M. Millones, 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment, *International Journal of Remote Sensing,* 32(15):4407–4429.

Qui, G., X. Feng, and J. Fang, 2004. Compression histogram representations for automatic colour photo categorization, *Pattern Recognition*, 37:2177–2193.

Saykol, E., U. Gudukbay, and O. Ulusoy, 2005. A histogram-based approach for object-based query-by-shape-and-color in image and video databases, *Image and Vision Computing* 23:1170–1180.

Stow, D., S.I. Touré, C.D. Lippitt, C.L. Lippitt, and C. Lee, 2012. Frequency distribution signatures and classification of within-object pixels, *International Journal of Applied Earth Observation and Geoinformation,* 15:49–56.

Swain, M.J., and D.H. Ballard, 1991. Color indexing, *International Journal of Computer Vision*, 7:11–32.

van der Meer, F., 2006. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery, *International Journal of Applied Earth Observation and Geoinformation,* 8:3–17.

Walter, F., 2004. Object-based classification of remote sensing data for change detection, *Photogrammetric Engineering & Remote Sensing*, 58(2):225–238.

Wu, S., J. Silvan-Cardenas, and L. Wang, 2007. Per-field urban land use classification based on tax parcel boundaries, *International Journal of Remote Sensing,* 28(12):2777–2801.

(Received 10 May 2012; accepted 09 August 2012; final version 17 December 2012)