Using Maps to Automate the Classification of Remotely-Sensed Imagery

Mark J. Carlotto (markjc@mindspring.com)
Pacific-Sierra Research Corporation, 1400 Key Blvd., Suite 700, Arlington, VA 22209

Abstract

The accurate classification of remotely-sensed imagery usually requires some form of ground truth data. Maps are potentially a valuable source of ground truth but have several problems (e.g., they are usually out-dated, features are generalized, and thematic categories in the map often do not correspond to distinct clusters or segments in the imagery). We describe several methods for using maps to automate the classification of remotely-sensed data, specifically Landsat Thematic Mapper imagery. In each, map data are coregistered to all or a part of the image to be classified. A probability model relating spectral clusters derived from the imagery to thematic categories contained in the map is then estimated. This model is computed globally and adjusted locally based on context. By computing the probability model over a large area (e.g., the full Landsat scene) general relationships between spectral categories and clusters are captured even though there are differences between the image and the map. Then, by adjusting and applying the model locally, new features can be extracted from the image that are not contained in the map and, in certain cases, different classes can be assigned to the same cluster in different parts of the image based on context. Experimental results are presented for several Landsat scenes. Several of the methods produced results that were more accurate than the map. We show that these methods are able to enhance the spatial detail of features contained in the map, identify new features not present in the map, and fill in areas in which map coverage does not exist.

Key Words: Digital Cartographic Data, Maps, Multispectral Classification, Bayesian Inference, Contextual Information

1. Introduction

The accurate classification of remotely-sensed imagery usually requires some form of ground truth data. Maps are a potentially valuable source of ground truth data but have several problems. Unlike imagery which contains a great deal of spatial detail, maps are spatially generalized (i.e., small features eliminated, boundaries smoothed, positions of features altered to improve the appearance of the map product). In maps thematic categories often do not correspond to distinct clusters or segments in the imagery (e.g., built up areas which contain a mixture of man-made materials such as concrete, trees and other kinds of vegetation, etc.). Finally, maps are often old and thus do not contain all of the features that are present in the image.

We describe several methods for using maps to automate the classification of remotely-sensed data, specifically Landsat Thematic Mapper (TM) imagery. These methods use a map coregistered to all or a part of the image to be classified. A probability model relating spectral clusters derived from the imagery to thematic categories contained in the map is developed. This model is computed globally and adjusted locally based on context. By computing the probability model over a large area (e.g., the full Landsat scene) general relationships between spectral categories and clusters are captured even though there are local differences between the image and the map. Then, by adjusting and applying the model locally based on context, new features can be extracted from the image that are not contained in the map and, in certain cases, different classes can be assigned to the same cluster in different parts of the image.

Preprint from SPIE Conference 2758, April 8-12, Orlando, Florida, 1996.

2. Map-Based Methods

Mason et al (1988) survey the use of map data to guide the segmentation and classification of remotely-sensed imagery. Maps have been used to guide the segmentation process (McKeown et al 1985), to select training areas for classifier development (Catlow et al 1984), to partition imagery into similar types of regions prior to classification (Hutchinson 1982), and to estimate prior probabilities (Strahler 1980). Maps have also been used as additional sources of information used in classification (Anuta et al 1976).

We use maps as a source of ground truth information for classification, specifically for developing probability models that are used to assign classes in the map to clusters derived from the image. Unsupervised clustering converts an input image $\mathbf{x}(i, j)$ into a cluster image $k(i,j)$ consisting of K values each corresponding to a cluster. Map data are rasterized and the image and map are coregistered. The map image $m(i,j)$ is assumed to contain M classes.

The joint distribution $p(k,m)$ is the probability that the k-th image cluster occurs together with the m-th map class over the entire joint data set. The joint distribution is used to compute the conditional distribution (observation model)

$$p(k \mid m) = p(k,m) / p(m) \tag{1}$$

which gives the probability of observing the k-th cluster given the m-th class where

$$p(m) = \sum_{k=1}^{K} p(k,m) \tag{2}$$

is the marginal distribution of classes. The Bayes classifier assigns the class m' with the highest posterior probability to each cluster

$$\begin{aligned} m'(k) &= \arg\max_m \left\{ p(m \mid k) \right\} \\ &= \arg\max_m \left\{ p(k \mid m)\, \hat{p}(m) / p(k) \right\} = \arg\max_m \left\{ p(k \mid m) \hat{p}(m) \right\} \end{aligned} \tag{3}$$

where $\hat{p}(m)$ are the prior probabilities for the M classes and

$$p(k) = \sum_{m=1}^{M} p(k,m) \tag{4}$$

is the marginal distribution of clusters. It is noted that the observation model $p(k \mid m)$ and prior probabilities $\hat{p}(m)$ may be computed from different sources of information, or over different parts of the data set.

The Bayes classifier minimizes the probability of error

$$P(\varepsilon) = 1 - \sum_{k=1}^{K} p\big(m(k), k\big) \tag{5}$$

by assigning the class $m(k)$ with the largest posterior probability to each cluster k. If the relationship between classes and clusters is distinct, then the probability of error will be relatively small. If there is

overlap between classes and clusters (i.e., if image clusters correspond to multiple map classes and vice versa), the probability of error will increase. Several factors can potentially contribute to this overlap: changes that occur between the time the map was compiled and the image acquired (including seasonal changes), cartographic generalization (e.g., small features eliminated, boundaries smoothed, positions of features altered to improve the appearance of the map product), misregistration, and spectral variations within classes over the image.

3. Global vs. Local Methods for Computing the Joint Distribution

The joint distribution captures the relationship between image clusters and map classes. In order to assess the impact of the above factors, global and local methods of computing the joint distribution are considered. The global method computes the joint distribution over the full data set, and uses it to derive a single observation model and set of prior probabilities that are used to classify the full data set. We shall refer to this classifier as a GJGP classifier (global joint and global priors). The GJGP classifier is

$$\begin{aligned} m'\big(k(i,j)\big) &= \arg\max_m \Big\{ p\big(m \mid k(i,j)\big) \Big\} \\ &= \arg\max_m \Big\{ p\big(k(i,j) \mid m\big)\hat{p}(m) \Big\} \end{aligned} \tag{6}$$

where the output image class m' depends only on the cluster k.

The local method partitions the data set into N regions. Within each region, the joint distribution is computed and used to derive an observation model and set of prior probabilities that are used to classify that region. We shall refer to this classifier as a LJLP classifier (local joint and local priors). The posterior probability in the n-th region is

$$\begin{aligned} p(m \mid k,n) &= p(k,n \mid m)\, p(m) / p(k,n) = p(k,m,n) / p(k,n) \\ &= p(k \mid m,n)\, p(m,n) / p(k,n) = p(k \mid m,n)\, p(m \mid n) p(n) / p(k,n) \end{aligned} \tag{7}$$

where $p(k \mid m,n)$ is the observation model and $p(m \mid n)$ are the priors for the n-th region. The LJLP classifier is

$$m'\big(k(i,j), n(i,j)\big) = \arg\max_m \Big\{ p\big(k(i,j) \mid m, n(i,j)\big)\, \hat{p}\big(m \mid n(i,j)\big) \Big\} \tag{8}$$

where the output image class m' depends on the cluster k and the region n. In Section 5 we evaluate both classifiers and find that the global method of estimating the joint distribution to be superior to the local method for the data set considered.

4. Estimating the Priors

We now consider alternative methods for estimating the priors. All assume the joint distribution is computed globally over the full data set. The first two methods use a single set of priors over the full data set. They are referred to as global joint global priors (GJGP) classifiers. The last three methods compute the priors over subsets of the data. They are referred to as global joint local priors (GJLP) classifiers. All five methods are evaluated in Section 5.

4.1 Global Methods for Computing the Priors

If there is a distinct relationship between classes and clusters; i.e.,

$$p\big(k \mid m(k)\big) > p\big(k \mid m\big), m \neq m(k) \tag{9}$$

the choice of the prior probabilities is not critical and can, in lieu of other information, be assumed to uniformly distributed. The first classifier (GJGP-1) simply assumes a uniform distribution for the prior probabilities, i.e. $\hat{p}(m) = 1/M$.

In most situations there is overlap between classes and clusters. In order to minimize the probability of error it is necessary to bias the conditional distributions to favor those classes that are more likely to occur in the image. The second classifier (GJGP-2) assumes that the probability of the m-th class in the image is equal to the probability of the m-th class in the map, i.e., $\hat{p}(m) = p(m)$.

4.2 Local Methods for Computing the Priors

A problem is using a single set of priors for the entire data set is that the same cluster may correspond to different classes in different parts of the image. For example, the GJGP-2 classifier assigns the most probable class to each cluster. Large classes thus tend to dominate. Three methods have been developed to adjust the priors locally. All three use some form of contextual information. The general form of these global joint and local priors (GJLP) classifiers is

$$m'\big(k(i,j), n(i,j)\big) = \arg\max_m \Big\{ p\big(k(i,j) \mid m\big)\hat{p}\big(m \mid n(i,j)\big)\Big\}. \tag{10}$$

This equation is similar to Eq. 8 except that the joint distribution is computed globally and so the observation model does not depend on n. Depending on the classifier, the variable n will take on different meanings as discussed below.

Knowledge-Based Contextual Classifier - The first classifier (GJLP-1) adjusts the prior distribution of image classes $\hat{p}\big(m \mid n(i,j)\big)$ as a function of the map class $n(i,j)$. These priors can, in principle, be derived from knowledge about how the map was compiled and how the map and image differ, e.g., due to changes between the time the map was compiled and the image acquired. The GJLP-1 classifier is given by

$$m'\big(k(i,j), n(i,j)\big) = \arg\max_m \Big\{ p\big(k(i,j) \mid m\big)\hat{p}\big(m \mid n(i,j)\big)\Big\} \tag{11}$$

where the output image class m' depends on the cluster k and map class n. Lacking specific knowledge of map compilation and changes we assume the following distribution for the priors:

$$\hat{p}(m \mid n) = \begin{cases} d, & m = n \\ (1-d)/(M-1), & \text{otherwise} \end{cases} \tag{12}$$

where $1/M \le d \le 1$. At one extreme, when $d = 1$ the output image class is forced to equal the map class; thus $m'\big(k(i,j), n(i,j)\big) = n(i,j)$. At the other extreme when $d = 1/M$ the image classes are equally probable regardless of the map class and the result is the same as the GJGL-1 classifier. Between these limits there is a blending of map and image-derived information.

Window-Based Contextual Classifier - The second classifier (GJLP-2) estimates the prior probabilities from the local distribution of clusters within a window. The estimate of the distribution of classes is

$$\hat{p}(m \mid n) = \frac{1}{p(n)} \sum_{k=1}^{K} p(m \mid k,n)\hat{p}(k,n) = \sum_{k=1}^{K} p(m \mid k,n)\hat{p}(k \mid n) . \quad (13)$$

where $\hat{p}(k \mid n)$ is the distribution of clusters within the n-th window. If the conditional distribution does not depend on n, it can be estimated from the joint distribution, i.e., $p(m \mid k,n) = p(m \mid k) = p(k,m)/p(k)$. The GJLP-2 classifier is given by

$$m'\big(k(i,j), n(i,j)\big) = \arg\max_{m} \Big\{ p\big(k(i,j) \mid m\big)\hat{p}\big(m \mid n(i,j)\big) \Big\} \quad (14)$$

where the output image class m' depends on the cluster k and window n. The GJLP-2 classifier emphasizes those classes that are most likely given the local distribution of clusters within the window. In effect, the local distribution of clusters functions as a form of context to adjust the class priors. Thus the same cluster can be assigned different classes in different parts of the image depending on the context. As weshall see later in Section 5, if this method is applied within a rectangular window, as the window moves across the image, clusters that are assigned one class in one window may be assigned a different class in an adjoining window. The switching of classes for a cluster is particularly noticeable when the cluster crosses the boundary between regions.

Feature-Based Contextual Classifier - The third method of estimating priors locally (GJLP-3) uses the map to partition the image into N regions, one for each map class, and to estimate the distribution of image classes within each map class. It is well-known that in the compilation of maps, a single thematic category may be assigned to a map feature that is, in fact, composed of a mixture of categories (e.g., built-up areas which contain a mixture of man-made materials plus trees and other kinds of vegetation). The GJLP-1 classifier assumes that this mixture (prior distribution) is known. Because image clusters and map classes overlap, we can instead attempt to infer this mixture indirectly through the association of map classes and image clusters in the same way as was done in the GJLP-2 classifier using Eq. 13 where $\hat{p}(k \mid n)$ now represents the distribution of image clusters within the n-th map class. The GJLP-3 classifier is given by

$$m'\big(k(i,j), n(i,j)\big) = \arg\max_{m} \Big\{ p\big(k(i,j) \mid m\big)\hat{p}\big(m \mid n(i,j)\big) \Big\} \quad (15)$$

where the output image class m' depends on the cluster k and the region n defined by the n-th map class. The GJLP-3 classifier emphasizes those image classes that are most likely given the local distribution of clusters within a map class. As with GJLP-2, this distribution of clusters functions as a form of context to adjust the class priors so that the same cluster can be assigned different classes in different parts of the image depending on the context. In areas that do not contain map information, i.e., where $n = 0$, we can assume $\hat{p}(m \mid 0) = p(m)$.

5. Experimental Results

The map-based classification methods described in the previous sections have been evaluated over two full Landsat Thematic Mapper (TM) scenes. The first is over New Mexico (path 33, row 36) and was acquired on April 15, 1994 (Figure 1a). The second is over Virginia (path 15, row 34) and was acquired on October 23, 1993. USGS 1:250,000 scale Land Use/Land Cover (LU/LC) maps were rasterized, resampled, mosaicked, and coregistered to the Landsat data. The LU/LC map over the New Mexico scene is shown in Figure 1b. For the New Mexico scene, the LU/LC classes were merged into a reduced set of classes as summarized in Table 1. As shown in the table a two level classification scheme is used so that we can quantify the performance of the map-based classifier at different levels of thematic resolution. A similar classification scheme was used for the Virginia scene. A set of 300 random sample points were generated over each scene and classified visually by an image analyst. For the New Mexico scene, Landsat TM

imagery was used as truth. Landsat TM and 5 meter/pixel M7 aircraft multispectral imagery were used as truth in the Virginia scene. The sample points were used to estimate the overall accuracy (number of correct sample points divided by the total number of sample points) of each of the map-based classification methods.

Level 1 Classes	Level 2 Classes	LU/LC (Anderson) Classes
Built-Up	High Density	11
	Low Density	12-17
Grassland/ Agriculture		21-24
Rangeland		31-33
Forest	Deciduous	41
	Evergreen	42
	Mixed	43
Water		51-54
Wetland		61-62
Barren	Sand	71-73
	Rock	74-75
	Soil	76-77
Tundra		81-85
Snow/Ice		91-92

Table 1 Definition of land cover classes

Both Landsat TM images were clustered using a spectral shape clustering technique (Carlotto 1995) that is based on representing the spectral data in terms of the relative values between bands. The technique automatically determines the number of clusters and found 257 clusters in the New Mexico scene and 166 clusters in the Virginia scene.

In order to assess the local and global methods of computing the joint distribution, the overall accuracy of the Bayes classifier was measured as a function of the size of the region classified (Eq. 8). A 4000x4000 pixel region in the center of the New Mexico scene was classified as a single, 4000x4000 pixel region, as four 2000x2000 pixel regions, and as sixteen 1000x1000 pixel regions. In each classification the joint distribution and prior probabilities are computing over the same area where

$$\hat{p}(m \mid n) = \sum_{k=1}^{K} p(k, m \mid n) \tag{16}$$

where $p(k, m \mid n)$ is the joint distribution in the $n-$th region. Level 2 classification results are summarized in Table 2. As the size of the region decreases, the average accuracy decreases and the variation in individual accuracies within regions increases; i.e., individual accuracies can be much higher or much lower than the average. This suggests that it is better to estimate the joint distribution over large areas and so favors global joint (GJxx) over local joint (LJxx) classifiers.

Area of Region(s)	*% Accuracy within Region(s)*	*Average*	*Standard Deviation*
4000	72	72	
2000	83 70 69 64	71	11

1000	63 81	66 69	100 57	50 85	90 25	100 28	71 66	75 50	67	22

Table 2 Classification accuracy as a function of the size of the region classified

We then compared the overall Level 2 accuracies of five GJxx classification methods:

1) Bayes classifier using a uniform distribution of image class priors (GJGP-1)
2) Bayes classifier where the distribution of image class priors is assumed to equal to distribution of map classes (GJGP-2)
3) Knowledge-based contextual classifier where the distribution of image class priors for each map class is assumed known (GJLP-1)
4) Window-based contextual classifier where the distribution of image class priors is estimated from the distribution of clusters within non-overlapping windows for different window sizes (GJLP-2)
5) Feature-based contextual classifier where the distribution of image class priors is estimated from distribution of clusters within each map classes (GJLP-3)

The overall classification accuracy of each method over the full New Mexico scene is shown in Table 3. Except for the first and third methods (GJGP-1 and GJLP-1), the accuracies are comparable (for 300 points the 95% confidence interval is about ±5%). For the third method, the measured accuracies were between that of the map (57%) and the first method (46%).

Method		*% Accuracy*
Bayes Classifier - Uniform Priors (GJGP-1)		46
Bayes Classifier - Priors Estimated from Map (GJGP-2)		63
Knowledge-Based Contextual Classifier (GJLP-1)		57-46
Window-Based Contextual Classifier (GJLP-2)	1600 x 1600	67
	800 x 800	66
	400 x 400	66
	200 x 200	65
	100 x 100	66
Feature-Based Contextual Classifier (GJLP-3)		67
Accuracy of LU/LC Map		57

Table 3 Comparison of map-based classification methods over New Mexico scene

Figure 1c shows the full scene GJLP-3 classification result. (The images in Figure 1 are reduced 40x in resolution. The classes have been merged and assigned shades of gray: water and wetlands are darkest, followed by forested areas, grassland/agriculture, rangeland, and barren; built-up areas are the brightest.) For all classifiers the distribution of errors were randomly distributed across the scene (Figure 1d shows the distribution of GJLP-3 errors).

Figure 2 shows the LU/LC map along with the results for the GJGP-2, GJLP-2 and GJLP-3 classifiers over Albuquerque. (The images in this figure are reduced 5x in resolution.) Although the overall accuracies of these three methods are comparable over the full scene, the GJLP-2 classifier identified only a small part of the imagery of Albuquerque as built-up (Figure 2b). An edge effect noted earlier can be seen in the GJLP-2 classification image in the upper right corner of Figure 2c. In addition, a large number of pixels were classified as built-up along the west slope of the Sandia Mountains east of Albuquerque. In the GJLP-3 classification result (Figure 2d) these errors are not present.

Similar experiments were performed for the Virginia scene and comparable results obtained. Table 4 summarizes the accuracy of the GJLP-3 classifier and the LU/LC maps for the two scenes at Levels 1 and 2. Based on these results, it appears that by using maps in conjunction with imagery we are able to significantly improve the overall thematic accuracy over that provided by maps alone.

Data Source(s)	*% Accuracy*			
	New Mexico Scene		*Virginia Scene*	
	Level 1	*Level 2*	*Level 1*	*Level 2*
LU/LC Map	58	57	77	55
LU/LC Map + Landsat TM Imagery	67	67	81	66
Absolute Improvement	9%	10%	4%	11%
Relative Improvement	15%	18%	5%	20%

Table 4 Summary of Results for Two Scenes for GJLP-3 Classifier

6. Discussion

In Section 5, the overall accuracy was found to decrease as the size of the region used to compute the joint distribution decreased. The probability of error depends on the overlap between classes and clusters. Four potential factors that can contribute to this overlap are: changes between the map and the image, cartographic generalization, misregistration, and spectral variations within classes over the image. Cartographic generalization and misregistration will tend to affect all parts of the data set and would thus not favor one method over the other. (We note that since the Landsat imagery is acquired close to nadir, misregistration is not strongly dependent on terrain relief and will tend to be distributed uniformly over the entire scene. For off-nadir geometries the effect would be expected to be more severe over mountainous areas.) Changes between the map and the image will, in most cases, tend to be localized and thus affect some parts of the data set more than others. The results do in fact show that some regions are classified much more accurately as the window size decreases. Unfortunately more regions are classified less accurately because of changes and so the overall accuracy decreased. Spectral variations within classes did not appear to be a significant factor for the clustering method used since the overall accuracy did not decrease as the size of the region increased.

Based on these preliminary observations we tend to favor the use of larger windows for estimating the joint distribution. Except for the use of uniform priors and assumed prior distributions in the GJGP-1 and GJLP-1 classifier, the other three methods for estimating the priors (GJGP-2, GJLP-2, and GJLP-3) did not result in significantly different overall accuracies. However the context-based classifiers were able to extract more built-up areas over Albuquerque than the Bayes classifier and seemed qualitatively to give better results.

Perhaps our most interesting finding is that the accuracy of a map-based classification can be better than the accuracy of the map itself. By selecting the class with the largest posterior probability (a non-linear operation), the classifier is somewhat tolerant of small errors between map classes, image clusters and ground truth. An important area of future work is to better understand the relationship between the clustering and map errors on classification accuracy.

7. Summary

The use of maps to automate the classification of remotely-sensed imagery was described. Local and global methods for estimating the joint distribution, which captures the relationship between image clusters and map classes, were evaluated. For the scene considered, the overall accuracy was found to decrease as the size of the region used to compute the joint distribution decreased, thus favoring global estimates of the joint distribution over local ones. Five methods of estimating the prior probabilities were then described and evaluated. Two involve choosing a single set of values that are used to classify the entire data. However because a cluster can correspond to different classes in different parts of the scene, three additional methods were developed to estimate the prior probabilities as a function of context. The first assumes that the distribution of image classes for each map class is known. The second method estimates the prior probabilities as a function of the distribution of clusters within fixed sized windows that cover the image. The third estimates the distribution of image classes from the distribution of clusters within map feature classes. Three of these five methods produced results that were more accurate than the map. We show that these methods are able to enhance the spatial detail of features contained in the map, identify new features not present in the map, and fill in areas in which map coverage does not exist.

Acknowledgments

The author wishes to credit Mark Lazaroff and Ronald DeWitt for suggesting the use of Bayesian methods for combining maps with image-derived information, and to thank Greg Heberle for assisting in the accuracy assessment.experiments.

References

D.C. Mason, D.G. Corr, A. Cross, D.C. Hoggs, D.H. Lawrence, M. Petrou, and A.M. Tailor, "The use of digital map data in the segmentation and classification of remotely-sensed images," *International Journal of Geographical Information Systems*, Vol. 2, No. 3, 1988.

P.E. Anuta, H.Huaska, and D.W. Levandowski, "Analysis of geophysical remote sensing data using multivariate pattern recognition techniques," *Proceedings 1976 Symposium of Machine Processing of Remotely Sensed Data*, Purdue University, 1976.

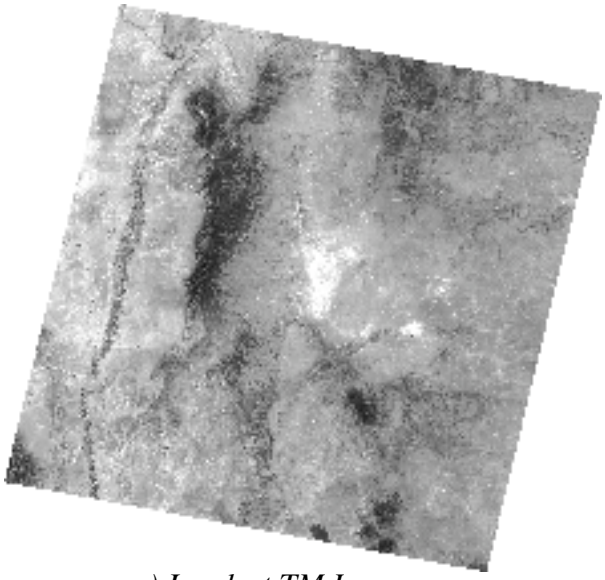
D.R. Catlow, R.J. Parsell, and B.K. Wyatt, "The integrated use of digital cartographic data and remotely-sensed imagery," *Proceedings of the EARSeL Conference on Integrated Approaches in Remote Sensing*, Guildford, England, 1984.

C.F. Hutchinson, "Techniques for combining Landsat and ancillary data for digital classification improvement," *Photogrammetric Engineering and Remote Sensing*, Vol. 48, No. 2, 1982.

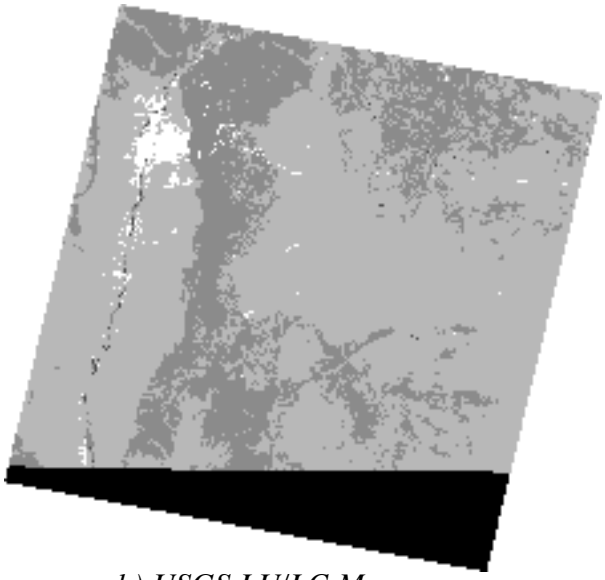
D.M. McKeown and J.L. Denlinger, "Map-guided feature extraction from aerial imagery," *Proceedings IEEE Workshop on Computational Vision Representation and Control*, Pittsburgh, PA, 1984.

A.M. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sensing of Environment,* Vol. 10, pp 135-163, 1980.

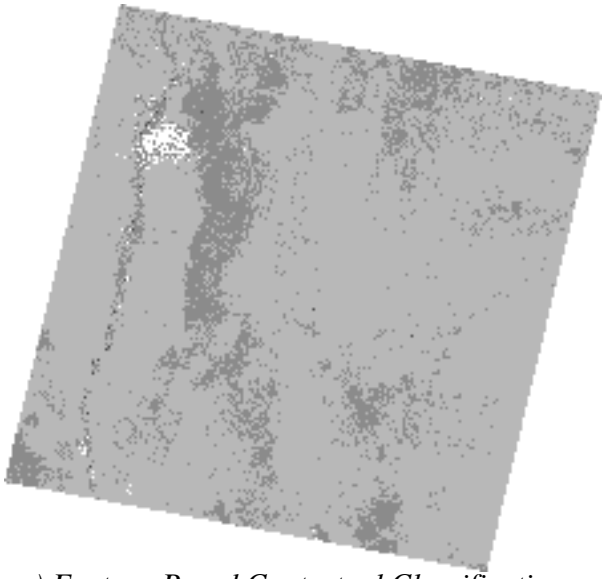
Mark J. Carlotto, "Spectral shape classification system for Landsat Thematic Mapper," *Proceedings SPIE*,, Vol. 2758, Orlando, Florida, 1996.



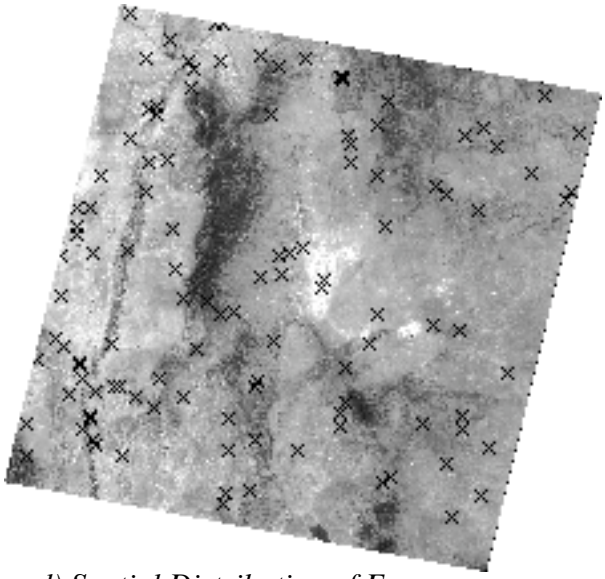
a) Landsat TM Image



b) USGS LU/LC Map

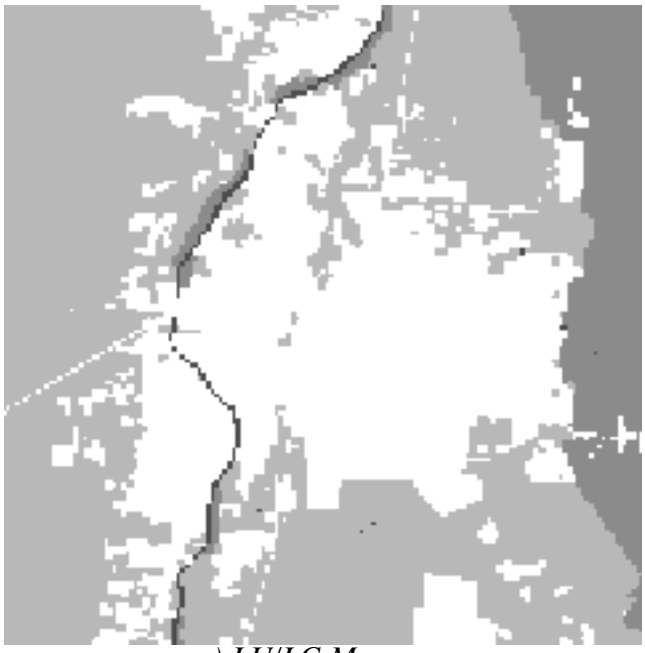


c) Feature-Based Contextual Classification

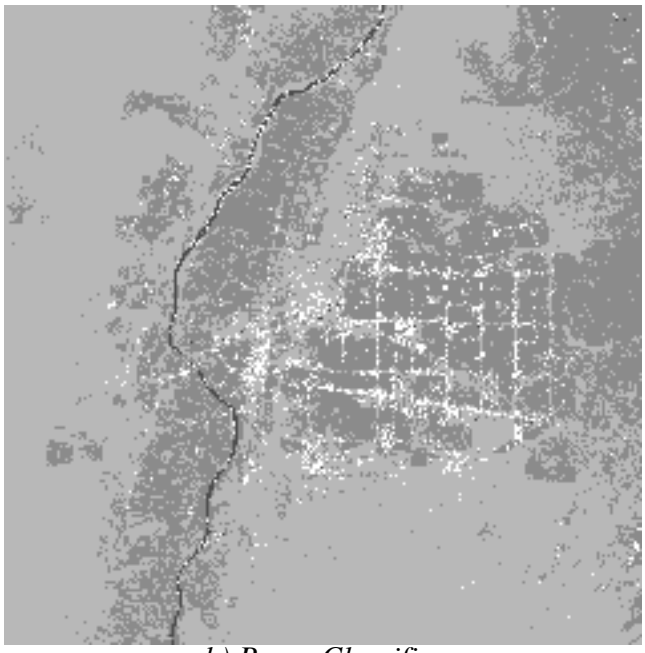


d) Spatial Distribution of Errors

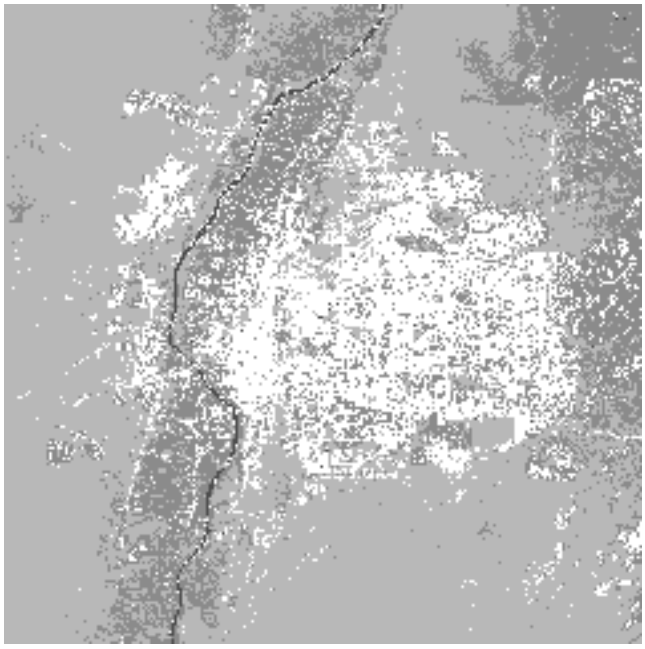
Figure 1 Results for New Mexico Scene



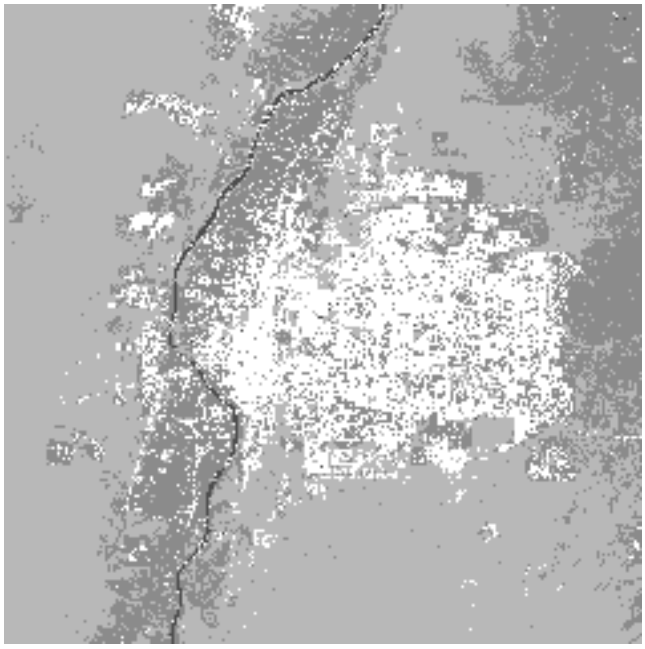
a) LU/LC Map



b) Bayes Classifier



c) Window-Based Contextual Classifier (1600x1600 window)



d) Feature-Based Contextual Classifier

Figure 2 Classification Results over Albuquerque, New Mexico