# A SELF-SUPERVISED APPROACH FOR FULLY AUTOMATED URBAN LAND COVER CLASSIFICATION OF HIGH-RESOLUTION SATELLITE IMAGERY

A. K. Shackelford [a], C. H. Davis [b]

[a] Radar Division, Naval Research Laboratory, Washington, DC – shackelford@radar.nrl.navy.mil
[b] Center for Geospatial Intelligence, University of Missouri-Columbia, Columbia, MO – DavisCH@missouri.edu

**KEY WORDS:** Self-supervised classification, high-resolution satellite imagery, urban land cover, feature extraction

**ABSTRACT:**

Commercially available high-resolution satellite imagery from sensors such as IKONOS and QuickBird are important data sources for a variety of urban area applications including infrastructure feature extraction and land cover mapping. Land cover maps from medium and high-resolution imagery are typically generated through supervised spectral classification of multispectral imagery. Supervised classification algorithms require training data as input and are thus semi-automated approaches. However, by automating the generation of training data, these supervised classifiers can be utilized in a fully automated, or self-supervised fashion to perform urban land cover classification. In this paper, we present a self-supervised approach for fully automated urban land cover classification of high-resolution satellite imagery. Automated feature extraction techniques are utilized to generate training data that are then input into supervised classification algorithms, thereby producing a self-supervised urban land cover classifier. These feature extraction techniques do not seek to extract all features present in the imagery. Instead, they are used to identify very high confidence instances of the different urban land cover classes. In this way, we limit the amount of incorrect training data that is input into the classifier. Because labeled training data is generated internally by the system, this classification approach is referred to as self-supervised. Self-supervised classification systems differ from unsupervised classifiers in that unsupervised classifiers output an unlabeled classification, requiring further analysis to determine the class labels, whereas the output of a self-supervised classifier is a labeled classification. Initial test results indicate that the overall accuracy of the self-supervised classification is 87-93%. There is only a 2% increase in overall accuracy when manually supervised classification is performed on the same test site.

## 1. INTRODUCTION

High-resolution satellite imagery became commercially available in late 1999 with the launch of Space Imaging's IKONOS satellite. In subsequent years, several other high-resolution commercial satellites were launched (DigitalGlobe's QuickBird and ORBIMAGE's OrbView-3). The spatial resolution and spectral information provided by these sensors make them well-suited for urban area applications. In particular, the high spatial resolution (0.6 – 1 m) allows for the delineation of fine-scale features in the urban environment, such as individual roads and buildings, which is not possible when utilizing imagery from medium resolution sensors (e.g. Landsat). The large volume of data collected by these sensors exceeds the human capacity of training image specialists to analyze. Currently, there are several additional high-resolution satellite sensors in the developmental stage, and when they become operational the problem will be further exacerbated. Automated upstream processing is needed to exploit the vast quantities of high-resolution commercial satellite imagery available from current and next generation sensors.

The generation of urban land cover maps from remote sensing imagery is typically accomplished through the use of supervised classification techniques, such as maximum likelihood. Supervised classification techniques require human generated training data and are thus only semi-automated. However, if the generation of training data is automated, supervised classifiers can be utilized in an unsupervised, or self-supervised fashion to perform urban land cover classification. In this paper, a fully automated approach for classification of urban land cover is presented. Feature extraction techniques are utilized to generate training data that are then input into the supervised classification algorithms, thereby producing a self-supervised urban land cover classifier. These feature extraction techniques do not seek to extract all features present in the imagery. Instead, they are used to identify very high confidence instances of the different urban land cover classes so as to minimize the amount of incorrect training data input into the classifier.

## 2. FULLY AUTOMATED TRAINING DATA GENERATION

Utilization of classification algorithms in an unsupervised or self-supervised fashion requires that the training data be generated automatically. Fully automated feature extraction techniques are used for this purpose. Because our goal here is to generate training data, not produce a complete extraction of the features of interest, the correctness of the extracted features is much more important than the completeness of the features. If incorrectly extracted features are used as training data, the errors will propagate through the classification process and lead to poor classification accuracies. The only concern in terms of the completeness of the extracted features is that a representative sample of the different spectral and spatial characteristics of the feature classes are obtained from the extraction. The strategy adopted here for generating training data for each urban land cover class is to output a fuzzy membership value for each extracted feature. The membership value represents a confidence level that the extracted feature is a valid member of the particular land cover class. Using these membership values, features with high confidence are selected and used as training data for each land cover class.

The urban land cover classes used in this study are: *Road*, *Building*, *Grass*, *Tree*, *Water*, and *Shadow*. To generate training data for the *Road*, *Building*, and *Shadow* classes, the automated feature extraction algorithms described in (Shackelford and Davis, 2003a; Shackelford and Davis, 2004) have been modified to output confidence values for each of the extracted features. The training data for the vegetation classes are generated by first identifying vegetation areas in the image with the NDVI (Jenson, 1996), followed by texture analysis to discriminate between *Grass* and *Tree* land cover classes. Confidence values are then assigned to the identified vegetation pixels. Training data for the *Water* class are generated through analysis of the DMP and the NDVI. The training data generation for each of the urban land cover classes is discussed in greater detail in the following subsections.

### 2.1 Road Training Data Generation

The fully automated road network extraction algorithm presented in (Shackelford and Davis, 2003a) was modified to output a fuzzy membership value for each extracted road, indicating the level of confidence in the validity of the extracted road. The road network extraction algorithm is an iterative process that first identifies and then grows road segments using several features extracted from the imagery and knowledge of the general characteristics and topology of a road network. Roads are initially identified as long linear segments of non-vegetation pixels, with longer segments having a higher confidence as being part of the road network than segments with short length. The algorithm begins by examining the longest length line segment present in the imagery, progressing to smaller length line segments as it iterates. Once a potential road segment has been identified, the algorithm examines the endpoints of the line segment and attempts to track the segment through small gaps and around curves in the road network. As road segments are iteratively added to the road network, a buffer is set up around them to exclude any line segments that are similar in angle and close to previously identified road network segments. This helps avoid overestimation of the road network and also eliminates multiple responses originating from a single road. The algorithm continues to iterate, adding new line segments to the road network until no line segments larger than a minimum length can be found.

The fuzzy membership confidence values are based on the length of the initial line segment detected for each road and the percentage of non-vegetative pixels present in the extracted road, as measured by the NDVI. Roads consisting of long line segments and low percentages of vegetation receive high confidence values.

### 2.2 Building and Shadow Training Data Generation

The fully automated 2-D building footprint extraction algorithm presented in (Shackelford and Davis, 2004) has been modified to output confidence values for the extracted building footprints and shadows. The building extraction algorithm is based on a multi-detector fusion strategy where buildings and their shadows are extracted from the Differential Morphological Profile (DMP) of panchromatic imagery and a segmentation of the pan-sharpened multispectral imagery. The DMP is a multi-scale image analysis technique where a morphological profile of the image is constructed through the use of morphological opening and closing by reconstruction operations while varying the size of the structuring element (SE) (Pesaresi and Benediktsson, 2001; Vincent, 1993). The DMP provides information about both the size and contrast of multi-scale structures in the image, with bright structures having a strong response in the opening portion of the DMP and dark structures having a strong response in the closing portion of the DMP. A multi-detector fusion approach is utilized for building extraction to accommodate the spatial and spectral variability in the appearance of urban buildings in high-resolution imagery. Buildings with a bright spectral response are extracted from the opening portion of the differential profile, while shadows are extracted from the closing portion of the differential profile. The extracted shadows are used to define search areas where the presence of buildings is likely. The search areas are then overlaid onto a segmentation of the multispectal imagery to identify building objects.

Fuzzy membership confidence values are computed for each extracted building footprint and shadow based on the geometric properties of the extracted objects. Building objects with rectangular shape and area similar to typical urban area buildings receive high fuzzy confidence values.

### 2.3 Vegetation Training Data Generation

Training data for the *Grass* and *Tree* classes are generated using the NDVI statistic in conjunction with the first order entropy texture measure (Gonzalez and Woods, 2002). First, the NDVI is used to identify pixels with vegetative land cover. Then, the entropy texture measure is utilized to identify high confidence instances of *Grass* and *Tree* from within the pixels identified as containing vegetative land cover. The first order entropy texture measure is calculated using an 11x11 pixel window. Pixels with both high NDVI values and high entropy values receive large membership values in the *Tree* class, whereas pixels with high NDVI values and low entropy values receive large membership values in the *Grass* class.

### 2.4 Water Training Data Generation

While the appearance of different bodies of water such as rivers, streams, lakes, and ponds varies significantly in high-resolution imagery, if one of these types of water bodies can be automatically identified, the extracted pixels can be used to train the classifier. Of the above-mentioned water body types, small lakes and ponds have the least variability in appearance, typically appearing as large, dark compact objects with no vegetation present. Objects fitting this profile are easily identified in the closing differential profile. Confidence values are computed based on the strength of the DMP response, the amount of vegetation present in the object, and the area of the object.

### 2.5 Training Data Generation

The fully automated feature extraction algorithms described above output a fuzzy confidence value for each extracted feature. Extracted features with high confidence values, indicating valid features, are utilized as training data, and the rest of the extracted features are discarded. This is accomplished by thresholding the fuzzy confidence value of each extracted feature. The thresholds are chosen such that they produce training data that is both accurate and representative of the variability within each land cover class. The training data generated for the *Road* and *Building* classes in a dense urban area, as well as the fuzzy confidence values for the extracted features, are shown in Fig. 1. The training data generated for the *Grass* and *Tree* classes for an area with suburban land cover, as well as the fuzzy confidence values for the extracted features, are shown in Fig. 2.

## 3. SELF-SUPERVISED URBAN LAND COVER CLASSIFICATION

Once high confidence instances of each urban land cover class have been identified, this data can then be utilized to train a supervised classification system. Because labeled training data is generated internally by the system, systems of this type can be referred to as self-supervised. Self-supervised classification systems differ from unsupervised classifiers in that unsupervised classifiers output an unlabeled classification, requiring further analysis to determine the class labels, whereas self-supervised classifiers output a labeled classification. The supervised classification scheme utilized here follows that described in (Shackelford and Davis, 2003b, Shackelford and Davis, 2003c), where the data is passed sequentially through three classifiers: a maximum likelihood classifier, followed by a pixel-based fuzzy classifier, and finally an object-based fuzzy classifier. A brief summary of the fuzzy classification approaches is provided below.

### 3.1 Supervised Fuzzy Classification

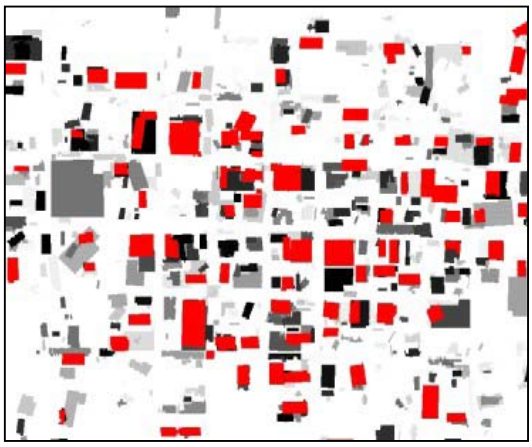
Due to the large numbers of spectrally similar land cover types present in the urban environment, traditional classification approaches such as maximum likelihood often result in significant numbers of misclassifications, especially between the *Road* and *Building* classes, and the *Grass* and *Tree* classes.

By utilizing spatial features in addition to the spectral information, the fuzzy pixel-based classifier is able to more accurately classify high-resolution imagery of urban areas. This classifier uses the results of an initial maximum likelihood classification of the imagery to group the classes where significant misclassifications occur together into sets. Subsequent processing using spatial features is then performed to differentiate between the spectrally similar classes. This approach allows for different groups of classes to be classified using the features best suited for discrimination between those classes. This alleviates the problem of features simultaneously decreasing the confusion between one set of classes and increasing it for another set.

The fuzzy pixel-based classification technique is significantly more accurate than maximum likelihood classification. However, more detail is needed to accurately represent the land cover types present in dense urban areas. A non-road, non-building *Impervious Surface* class is also needed to represent features such as parking lots, concrete plazas, etc. To distinguish between these urban land cover classes, an object-based classification approach is used to examine features such as object shape and context (neighborhood) and then classify the image objects using a fuzzy logic rule base. To facilitate object classification, the imagery is first segmented with a region merging segmentation technique. Several features are extracted from the image objects and used by the object-based classifier along with the fuzzy pixel-based classification. These features are the class labels of each segment's constituent pixels, shape information from the image objects, neighborhood analysis, and spectral statistics of the object. A shape model for the *Building* class, based on the skeleton of the image objects, is constructed using fuzzy membership functions, and the neighborhood analysis consists of examining the relationship between *Building* and *Shadow* segments. The image objects are then classified by a fuzzy logic rule base.

a)

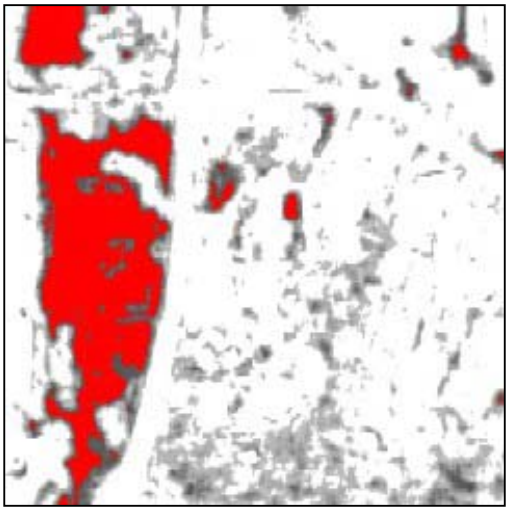
b)



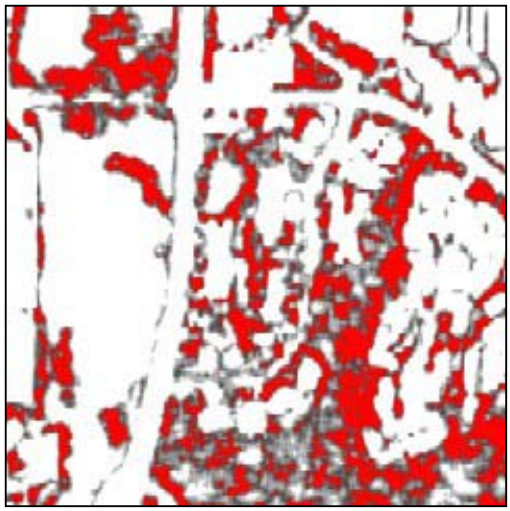
c)

**Figure 1.** Automatically generated training data from a) dense urban area for b) *Road* class, and c) *Building* class. Training data shown in red, high confidence features shown in dark gray, and low confidence features shown in light gray.

a)



b)



c)

**Figure 2.** Automatically generated training data from a) residential area for b) *Grass* class, and c) *Tree* class. Training data shown in red, high confidence features shown in dark gray, and low confidence features shown in light gray.

### 3.2 Unsupervised Clustering

Because of spectral variation within individual land cover classes, it is necessary to train the supervised classifier on multiple sub-classes within each urban land cover class. Following classification, the sub-classes from each land cover class are combined. To accommodate within class spectral variation, the automatically generated training data from each class is divided into sub-classes via unsupervised clustering before maximum likelihood classification is performed. After maximum likelihood classification, the sub-classes are combined into the urban land cover classes of interest and the classification process continues through the fuzzy pixel-based and object-based approaches. Unsupervised clustering is performed utilizing the standard *k*-means clustering algorithm (Theodoridis and Koutroumbas, 1999).

## 4. TEST RESULTS

The fully automated self-supervised classification scheme was applied to an IKONOS image of Columbia, Missouri. The test site is shown in Fig. 3 and consists primarily of dense urban land cover. An accuracy assessment of the resulting classifications was performed making use of reference pixel datasets. Accuracy assessments were performed for each of the classification outputs (maximum likelihood, fuzzy pixel, and fuzzy object) produced by the self-supervised classification scheme. The individual class accuracies and the overall classification accuracy for each output, as well as the corresponding results from the semi-automated supervised classifiers requiring human input of the training data, are displayed in Tables 1 through 3.

**Table 1**
Accuracies of Maximum Likelihood Classifications

| | **Supervised (%)** | **Self-Supervised (%)** |
|---|---|---|
| *Road* | 84.4 | 80.2 |
| *Building* | 83.1 | 63.7 |
| *Grass* | 92.8 | 93.8 |
| *Tree* | 83.5 | 91.9 |
| *Overall Accuracy* | **85.9** | **82.4** |

**Table 2**
Accuracies of Fuzzy Pixel-Based Classifications

| | **Supervised (%)** | **Self-Supervised (%)** |
|---|---|---|
| *Road* | 97.9 | 97.2 |
| *Building* | 90.7 | 93.2 |
| *Grass* | 94.5 | 100.0 |
| *Tree* | 96.2 | 79.6 |
| *Overall Accuracy* | **94.8** | **92.5** |

**Table 3**
Accuracies for Fuzzy Object-Based Classifications

| | **Supervised (%)** | **Self-Supervised (%)** |
|---|---|---|
| *Road* | 99.2 | 95.0 |
| *Building* | 76.1 | 70.1 |
| *Imp. Surf.* | 81.0 | 72.2 |
| *Grass* | 91.3 | 100.0 |
| *Tree* | 99.9 | 99.1 |
| *Overall Accuracy* | **89.5** | **87.4** |

For each of the classifier outputs, the overall accuracy of the fully automated self-supervised classification is only 2-3% lower than that of the semi-automated supervised classification. There is a significant decrease in the *Building* class accuracy of approximately 20% between the self-supervised and manually supervised maximum likelihood classification, as can be seen in Table 1. This is due to over classification of the *Road* class. However, the problem is solved by the hierarchical fuzzy pixel-based classification stage, where the average accuracies of the *Road* and *Building* classes exceed that of the manually supervised classification. There is a 17% decrease in the *Tree* class between the manual and self-supervised fuzzy pixel-based classification due to over classification of the *Grass* class. This error is unexpected and believed to be due to the fact that all of the automatically generated training data for the *Tree* class is extracted from highly textured areas. While appropriate for the majority of this class, there are areas within the *Tree* land cover class that are not highly textured, possibly due to trees with very large and homogeneous crowns. Because the texture of these instances of *Tree* land cover matches that of *Grass*, they are misclassified. The problem of misclassification of *Tree* reference pixels as *Grass* is solved in the object-based stage of the self-supervised classification. As seen in Table 3, the self-supervised classifier produces a classification with virtually no errors in the reference data of the vegetative classes. The object-based classifier is able to correct this problem because the proportions of each class present in the object are used as features, resulting in a majority filtering type operation within the object. The regions where *Tree* land cover are incorrectly identified as *Grass* in the self-supervised pixel-based classification are all quite small and are removed by the majority filtering effect of the object-based classifier.

There are 4%, 6%, and 9% decreases in the accuracies of the *Road*, *Building*, and *Impervious Surface* class accuracies, respectively, between the semi-automated and self-supervised object-based classifiers. It is believed that the decrease in the accuracy of the *Building* and *Impervious Surface* classes is partially due to errors in the classification of the *Shadow* class, which is used in the identification of the *Building* class. The self-supervised fuzzy object-based classification of the urban test site is shown in Fig. 4.

## 5. CONCLUSION

A fully automated self-supervised classification approach for urban land cover classification of high-resolution multispectral satellite imagery is presented in this paper. The classifier is based on supervised classification approaches presented in (Shackelford and Davis, 2003b, Shackelford and Davis, 2003c). However the training data is automatically generated using feature extraction techniques that identify high confidence instances of the urban land cover features. The automated road network and 2-D building footprint extraction algorithms described in (Shackelford and Davis, 2003a; Shackelford and Davis, 2004) have been modified to output a fuzzy confidence value for each extracted feature. Other spatially and spectrally based feature extraction algorithms have been developed to identify training data for the other urban land cover classes. After feature extraction and selection of high accuracy training data, the extracted features are subdivided into spectrally coherent sub-classes by unsupervised spectral clustering. The training data are then used to train a maximum likelihood classifier, followed by the hierarchical fuzzy pixel-based classifier, and finally the object-based classifier. Test results indicate that the self-supervised classification approach is able to produce urban land cover maps with overall accuracies that are only 2-3% less than that of the semi-automated supervised classifiers that require human input of training data.

## REFERENCES


Gonzalez, R. C. and R. E. Woods, 2002, *Digital Image Processing*, 2nd ed., Prentice Hall: New Jersey.

Jenson, J. R., 1996, *Introductory Digital Image Processing*, 2nd ed., Prentice Hall: New Jersey.

Pesaresi, M., and J.A. Benediktsson, 2001, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing,* Vol. 39, No. 2, pp. 309-320.

Shackelford, A. K. and C. H. Davis, 2003a, "Fully automated road network extraction from high-resolution satellite multispectral imagery," *Proceedings of International Geoscience and Remote Sensing Symposium,* Vol. 1, pp. 461-463, Toulouse, France, 21-25 July, 2003.

Shackelford, A. K. and C. H. Davis, 2003b, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sensing,* Vol. 41, No. 9, pp. 1920-1932.

Shackelford, A. K. and C. H. Davis, 2003c, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sensing,* Vol. 41, No. 10, pp. 2354-2363.

Shackelford, A. K., C. H. Davis, and X. Wang, 2004, "Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery," *Proceedings of International Geoscience and Remote Sensing Symposium,* Vol. 3, pp. 1996-1999, Anchorage, Alaska, 20-24 September, 2004.

Theodoridis, S. and K. Koutroumbas, 1998, *Pattern Recognition*, Academic Press: California.

Vincent, L., 1993, "Morphological greyscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Processing,* Vol. 2, pp. 176-201.

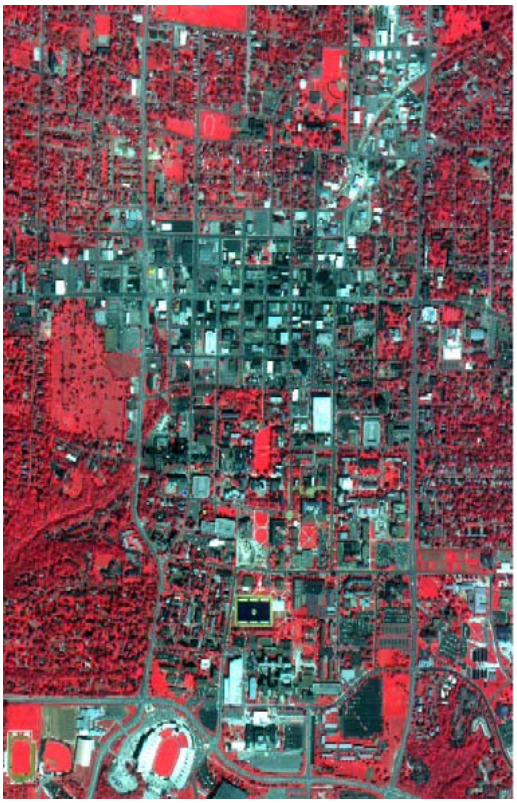


**Figure 3.** Pan-sharpened multispectral IKONOS image of dense urban area.

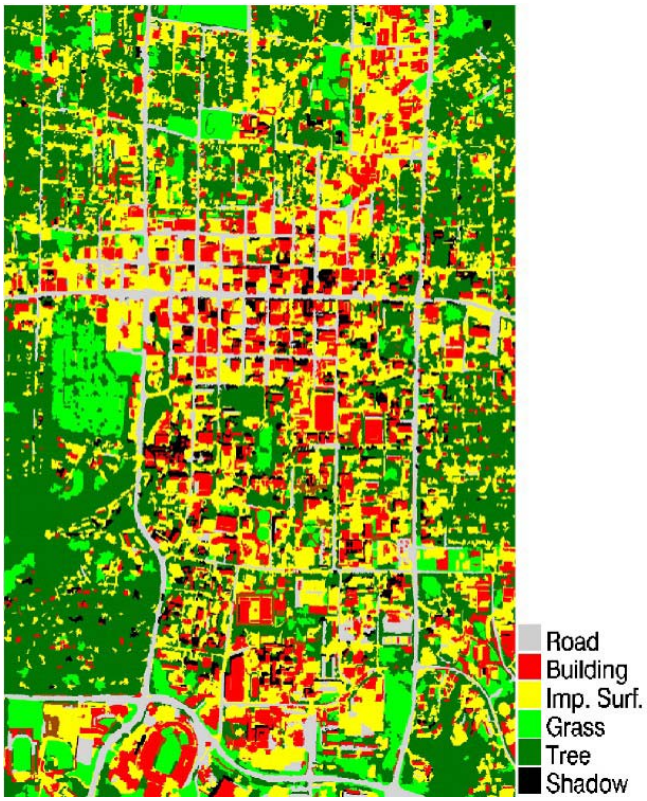



**Figure 4.** Self-supervised object-based classification of dense urban area test site.