Image Analysis Techniques for Urban Land Use Classification. The Use of Kernel Based Approaches to Process Very High Resolution Satellite Imagery

Charalambos C. Kontoes
Institute of Ionospheric and Space Research
National Observatory of Athens, Lofos Nimfon, 11810, Thission, Athens, Greece.
e-mail: kontoes@creator.space.noa.gr

Summary. Information on land use classes of urban areas is very important for their management and planning. Satellite remote sensing data can provide regular and up to date information on urban areas. Although satellite technology has changed significantly, a robust methodology in exploiting fruitfully satellite imagery in urban areas is under investigation.

This paper gives a literature review of image analysis techniques used for the classification of medium and high resolution satellite imagery in urban areas. In addition, three kernel based classification techniques to process very high resolution imagery, are considered that integrate texture and spatial context properties for urban land use classes. The Evidence Based Interpretation Algorithm, a backpropagation neural network algorithm and the Kernel Reclassification algorithm. A summary of their use on Indian Remote Sensing (IRS) satellite imagery over the city of Athens is presented. Although some of these techniques return rather satisfactory results, the need for the development of more advanced interpretation methods, which integrate human reasoning in object identification is considered as indispensable, as the spatial resolution of the data is continuously increasing attaining the one of high altitude aerial photography (e.g. IKONOS system).

1. Introduction

The level of information that can be extracted from remotely sensed data is related to the spatial and temporal resolution of the acquired images. There is no doubt that the interpretation of the aerial photography typically provides a lot of the needed information for urban studies. However, this approach is costly and difficult to apply with sufficient frequency. On the other hand, satellite remote sensing can provide a method for acquiring regular and up to date information about urban areas, which may be particularly useful for monitoring changes within and on the fringes of urban development. Foster [7] makes an exhaustive presentation of the potentiality and advantages in using satellite technology for urban mapping and monitoring, but at the same time gives a detailed overview of the problems which arise when remote sensing imagery is treated in the frame of such studies.

During the last few years satellite technology has changed significantly and new sensors with higher spatial resolution and stereo-pair capabilities
are offered for wide use (e.g. IRS-1C, IRS-1D, Radarsat, IKONOS, SPOT 5, KVR, KFA). But the technological advances have not solved the problem to identify a robust, transferable and easily repeatable methodology in exploiting fruitfully the satellite imagery in urban areas and many research topics are still open and under investigation. The first experiments in integrating satellite imagery in urban studies were not encouraging in terms of land use class specificity ([3, 30, 32]). This was initially attributed to the coarse spatial resolution of the satellite sensors. The fact that the spectral responses of different land cover types (e.g. buildings, roads, trees, grass, etc), which coexist in the field of view of the sensor, are averaged and registered as one pixel value, led to produce broad composite signals making thus difficult to distinguish between different land categories. Examples in using Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) data for urban classification are reported in various studies ([12, 19]). The only possible distinction achieved was between urban and non-urban areas. Other studies ([16, 23]) investigate the use of multi-spectral SPOT imagery or merged Landsat TM and SPOT Panchromatic data for urban classification. These studies also showed that spectral and spatial resolution of the SPOT data, were also inadequate for defining land cover/land use classes with the required specificity. The use of merged data returned slightly improved classification results but still far from the required detail of specificity.

The use of higher resolution data from recently developed sensors, although still under investigation, has not returned the expected improvements and in some cases the returned accuracy is even worse compared to the one achieved by the use of coarser resolution satellite data. This was attributed, to the problem of "scene noise" ([12, 22]), as the spectral signatures of urban areas are much more varied, since are composed by the spectral responses of individual scene elements. This makes land use class description and identification problematic, especially when classification is treated on a per-pixel fashion. In some cases, as in [6], the problem of "scene noise" is considered as being so unresolvable, that it is suggested to remove textural information and interclass variability from the scene plane, by the use of specific smoothing algorithms, in order to achieve better classification results. An overview of the experience in exploiting satellite imagery for urban land use mapping follows.

2. Satellite Image Interpretation Techniques for Urban Classification

The continuous advancement of satellite technology resulting in the acquisition of higher quality images, is strongly required for urban monitoring and change detection studies. As Barsnley and Barr [1] state, the problem to identify specific land use classes in urban areas, should not be attributed only to
the spatial resolution of the data but mainly to the methods used to process and extract information from the scene. In many of the known studies the authors make use of pixel based approaches in classifying images of urban areas by the application of statistical cluster approaches. Although, this is a rather good technique for classifying large homogeneous agricultural fields, it may not return significant results within urban areas, especially because the spectral composition of urban land use classes, is violating the basic assumption for normal distribution ([11, 21]). This is because certain urban land use classes are composed by more land cover classes and with different percentage of their appearance within the boundaries of the land use class. Therefore, land use clusters for urban areas, will have multi-modal and considerably overlapping signatures which lead to significant miss-classifications. Therefore, pixel based approaches in urban areas, can suffer from being either too specific or too general, but far from the more abstract definition of land use classes.

The fundamental problem in producing accurate land use maps is that urban areas present a complex spatial arrangement of land cover types. Thus, the necessity for integrating techniques which account for the spatial arrangement of pixel values or alternatively land cover labels within a neighborhood was considered indispensible by a number of researchers. In many studies the spatial arrangement of pixel values is introduced by the use of texture measurements, which are integrated either in statistical classification models as additional layers or as input into advanced classification algorithms, through knowledge based systems, artificial neural network algorithms, image understanding approaches (especially the structural elements of texture). Texture has been used in the frame of many studies in the past, either in its structural or statistical form. Several investigators have used the most common gray-level co-occurrence matrices to measure entropy, angular second moment, variance, correlation, etc. ([5, 8, 13, 14, 17, 27, 33]). Ryherd and Woodcock [29] introduce the “local variance” texture measure, which has been computed within a $3 \times 3$ adaptively placed window is introduced for image segmentation. The method was tested on satellite data representing rural and urban areas. The experiments showed that the combined use of spectral with texture information returned in all the cases improved classification accuracy especially within urban areas exhibiting pronounced texture characteristics. Paola and Schowengerdt [26] have devised an interesting way to provide spatial texture to the classification approach. In this study a $3 \times 3$ geometric window of pixels is considered as input to a neural net work classification approach. The method was tested on a Landsat TM scene representing a complex urban area. The results were satisfactory since the incorporation of texture resulted in the reduction of the required time for network training and at the same time the returned thematic accuracy, was higher than the one given also by a neural net but based on a pixel-by-pixel approach instead of a $3 \times 3$ window.
Except for the statistical measurements of texture the structural methods have been also used in the frame of relevant image interpretation and analysis studies. Nagao and Matsuyama [25] introduce image understanding systems which is divided into two sub-systems: the low level which measures and assigns specific geometric properties to objects, which have been given by image segmentation. In the high level processing which follows, the use of specific knowledge, through an expert system, results in object recognition on the basis of their geometric properties and spatial relations with other neighboring objects. This detection system, was developed to recognise objects like, two types of crop fields, bare soil, crop field without plant, forest, grassland, road, river, car, building, house. In a very similar way Gahegan and Flack [9], have experimented with image understanding techniques integrating structural, spectral and spatial properties to identify, through the use of a knowledge based system, the labels of specific objects. Their system encompasses both pixel based but also feature based processing techniques to extract significant information for image interpretation.

However, the classification in urban land use classes is more than the identification of the exact label of classified objects on the image plane. Urban land use classes present a complex spatial assemblage of land cover types introduced by pixel labels or entire region labels, which are usually the result of the image classification and/or segmentation. A technique which accounts both for the frequency and spatial arrangement of class labels within a square kernel has been applied by Barsnley and Barr [1]. This technique was first suggested by Wharton [34] and divides the classification process into two stages: during the first stage each pixel is assigned a certain classification label through a standard per pixel classification approach. The second applies a spatial post classification process to measure the frequency and spatial arrangement of pixel labels within a user defined kernel to infer urban land use classes. The technique is using a statistical similarity criterion to assign the central pixel of the kernel to a land use class. In the study of Barsnley and Barr [1] initial classifications of urban areas in land cover types such as “small structure”, “large structure”, “tree”, “crop”, “grass”, and “soil” were combined spatially through the use of the kernel based reclassification technique, in order to infer land use classes like “low-density residential”, “medium density residential”, “commercial”, etc.

The integration of ancillary data in an attempt to refine and improve the classification in urban areas, is realised either during classification or in a post-classification stage of the process. There are two approaches to using the ancillary data during classification. The first introduces the ancillary layer as an additional layer forming an extended vector of values for each pixel ([31]). The second makes use of the ancillary layer to modify the apriori probabilities of the classes under consideration before the employment of the classification algorithm. The latter technique has been proved very demanding in terms of sampling efforts and the quality of the classification output.
is very scene depended. Moreover, the integration of the ancillary data as component of an extended vector usually results in the violation of the normal distribution of pixel values, required by many statistical classification approaches. These remarks have led many researchers to make use of ancillary data into a post-classification process. Kontoes et al. [20] has integrated ancillary thematic layers introducing geographic and spatial context information, through a knowledge based system to refine the first classification output. Another typical example is the work of Harris and Ventura [15]. In this study the integration of zoning and housing density data and their use in the frame of a knowledge based system, resulted in the reduction of confusion between classes and the increase of the number of identifiable land use classes within an urban area. The initial classification in 5 land use classes was further refined into 15 classes by integrating the two additional layers. Similarly Moller-Jensen [24] is integrating texture and context information into a knowledge-based classification approach to classify urban areas.

3. Kernel Based Classification Approaches for the Interpretation of IRS-1C Satellite Imagery

On the basis of the above mentioned studies, it becomes clear that the objective to define adequate image analysis techniques, using statistical and non-statistical approaches to extract reliable signatures for urban land use classes, requires the development of algorithms which measure the spatial distribution of the spectral and texture properties found on the pixel context and produce significant features to introduce into the classification procedure. Currently the author is involved into a pilot project, funded by the European Commission, aiming to exploit very high resolution imagery provided by the Indian Remote Sensing, IRS-1C, satellite system for urban classification of the city of Athens, according to the Eurostat's CLUSTERS\(^1\) nomenclature scheme. Three Kernel Based Classification Approaches have been considered, for the purposes of the study. The EBIS (Evidence Based Interpretation of Satellite Images), the supervised and unsupervised artificial neural network (ANN) and the kernel reclassification algorithms.

The application of an unsupervised neural network based on a topological network (Kohonen map) is suggested in order to decide about the optimal kernel size and input image information content, similarly to the work of Paola and Schowengerdt [26] and Wilkinson et al. [35]. The output of this investigation, is used to feed appropriately the back-propagation neural network, consisted of one input layer, two hidden layer(s) and an output layer. The number of hidden layers and their dimension is chosen heuristically, based on previous experience and studies.

\(^1\) Classification for Land Use Statistics: Eurostat Remote Sensing programme.
The “Evidence-Based Interpretation of Satellite Images” algorithm (EBIS) makes use of the reasoning scheme of the Shafer’s theory of evidence ([4, 10]). Therefore evidence functions are used to decide whether a pixel is assigned to a certain class or not. The parametric model used by the algorithm, is the multinomial distribution. Via evidence functions the pixel is assigned to a special class in case that the hypothesis, which is the local histogram of the training area, matches the histogram of the pixel observed in a specific window. The evidence function measures the degree to which a pixel’s feature match a class hypothesis. If the pixel’s feature match a histogram of a training area well, then this constitutes evidence in favour of the corresponding class hypothesis. Apart from the local histogram, EBIS also supports the common co-occurrence matrices introduced by Haralick et al. [13]. This is because, there are classes, which show the same histogram, but are clearly distinguishable by the human eye (e.g. an image with black and white stripes has the same histogram as an image with the same number of black and white pixels arranged as the pattern on a chess-board). In this case co-occurrence matrices, which record the relative frequencies of spatial co-occurrences of grey values, are useful. Horizontal, vertical, left-diagonal and right-diagonal co-occurrences can be defined.

The Kernel Based Reclassification Algorithm attempts to derive information on urban land use based on the frequency and the spatial arrangement of the land cover labels within a square kernel. The assumption underlying this approach is that individual categories of land use have characteristic spatial mixtures of spectrally distinct land cover types that enable their recognition in high spatial resolution images ([2, 34]).

The employment of neural networks ([18, 36]), evidential based approaches ([4, 10, 28, 33, 36]) and reclassification techniques, allows the integration of any type of sources of data classes to derive information (land use) classes, independently if the assumption of normal distribution of density functions is fulfilled or not. At the same time these techniques introduce texture and contextual information into the classification procedure.

4. Evaluation of the Use of the Kernel Based Algorithms to Classify IRS-1C Satellite Data over Athens

The application of the three kernel based classification algorithms on high resolution IRS-1C satellite data over Athens, provided useful information concerning the classification abilities of texture based classifiers in respect to the Eurostat’s CLUSTERS nomenclature scheme used for land use description. It should be noted that the CLUSTERS nomenclature scheme presents four hierarchical levels of detail in land use/land cover classes for urban, agricultural and forested areas. The experiments showed that some urban land use classes at Level VI and Level III could be derived with relatively good
accuracy by applying kernel based classifiers. However, this was the case only for the neural network (figure 4.2) and kernel reclassification algorithms (figure 4.3). The use of EBIS classifier did not produce meaningful map products. The EBIS algorithm returned the best results when the only input layer was the panchromatic scene of IRS-1C. In contrast multi-temporal and multi-spectral satellite data resulted in noisy classification products. The neural network and kernel reclassification algorithms produced more accurate maps at the fourth Level of the CLUSTERS nomenclature scheme and four residential classes which differ from one another in terms of housing density, were returned. These classes are “Continuous dense residential”, “Continuous residential with moderate density”, “Discontinuous residential with moderate density”, “Isolated building areas”. However these classes, are only a subset of the CLUSTERS classes, as many of the other classes in the nomenclature are functional classes, which can be derived solely from ancillary data and analog photo-interpretation. Although many experiments have been conducted by using various kernel sizes and image layer combinations, including multi-spectral and multi-temporal sets of enhanced (fused) data, it was not possible to separate class “Industrial or commercial activities”. This most likely happens because there does not exist a typical texture pattern for this class on the image plane.

The best classification results showed an accuracy of the order of 72–75% for the various classified urban classes. In general, the texture based classification approaches proved to be useful for the classification of high resolution data. In contrast, pixel-based approaches like e.g. the maximum likelihood classifier, classify most of these classes as a mixture of single classes, like vegetation, roads and houses of different kinds resulting in confused classifications with a lot of “salt and pepper” noise (figure 4.1). The results showed, that the classification accuracy is directly linked to the applied size of the kernel used, since heterogeneous classes need large window sizes to be correctly represented, whereas small objects demand the contrary. Therefore, a combination of both results is necessary in case homogeneous classes are needed, but without accepting a loss of all small structures.

Further research is required to study a set of pre-defined kernel sizes for data classification, which depend on the class or set of classes to be classified. For the purposes of this study, a trial and error approach was followed in order to decide about the best window size to be applied for the IRS-1C data classification. In all the experiments, the texture properties of the land use classes have been introduced by the use of enhanced (merged) satellite data, which is the result of the integration of multi-spectral images of the LISS III sensor with the high spatial resolution panchromatic image of the IRS-1C Panchromatic sensor.

The following sections describe in greater detail the classification results obtained from the maximum likelihood, neural network and kernel reclassification approaches.
4.1 Maximum Likelihood Classification

Five classification experiments have been conducted within the urban area of the city of Athens. Various combination of input layers, comprising multi-temporal and enhanced multi-spectral IRS-1C LISS III data with the addition of texture (variance layer) have been considered as input to the classification procedure. The aim was to evaluate the degree by which a per-pixel classification approach based on spectral and textural properties could return meaningful classifications. The experiments showed that only non-urban classes could be classified accurately. With the only exception of class "Dense Residential Areas" (A111) the rest of the urban land use classes (A) were classified wrongly and a lot of inter class confusion revealed. The classification accuracy for urban classes was ranging between 20 to 45%. Only at mixed levels of the nomenclature scheme the returned accuracy values are at higher levels. Figure 4.1 illustrates a subset of 514 lines × 319 pixels out of the total study area (40 × 40 km²).

Fig. 4.1. Maximum likelihood classification of a subset of the study area. Classification nomenclature: A = 'man-made areas', A111 = 'continuous dense residential', A32 = 'transport', A502 = 'sport facilities', A503 = 'green or leisure areas', B132 = 'fresh vegetation', B3 (B330) = 'permanent crops', C13 = 'conifers', D102 = 'bushes', D201 = 'herbaceous vegetation'.
4.2 Neural Network Classification

The back-propagation neural network classifier was applied on the combination of multi-temporal very high resolution and enhanced IRS-1C LISS III data. The neural network consisted of four layers: one input, two hidden and one output layer. The network applies image classification within a $7 \times 7$ kernel scanning the image. The experiment showed that the application of the back-propagation neural network resulted in accurate discrimination between urban land use classes as a function of housing density within the urban area of the city of Athens. The overall classification accuracy achieved was 72.02%. Figure 4.2 illustrates a subset of 514 lines $\times$ 319 pixels out of the total study area ($40 \times 40 \ km^2$).

Fig. 4.2. Neural network classification of the same subset of the study area as in figure 4.1. Classification nomenclature: As in figure 4.1 with the addition of, A112 = 'continuous residential of moderate density' A113 = 'discontinuous residential of moderate density', A12 = 'public services'.

4.3 Kernel Reclassification Algorithm

The kernel reclassification was applied on the first classification map, resulted by the employment of the ISODATA clustering algorithm on the enhanced IRS-1C LISS III satellite data. The unsupervised classification layer was introducing five land cover classes. The kernel reclassification algorithm inferred land use classes on the basis of the spatial arrangement of the five land cover labels within a $11 \times 11$ kernel. Similarly to the neural network, the experiment showed that this approach may result in accurate discrimination between urban land use classes as a function of housing density within the urban area of the city of Athens. The overall classification accuracy achieved was 72.04%. Figure 4.3 illustrates a subset of 514 lines $\times$ 319 pixels out of the total study area (40 $\times$ 40 km$^2$).

![Image of classified land use classes](image)

**Fig. 4.3.** Kernel based classification of the same subset of the study area. Classification nomenclature as in figure 4.2.
5. Conclusions

From the three classification experiments it was concluded that the back-propagation neural network classification returned a rather good classification, in terms of class specificity and accuracy. The kernel reclassification algorithm could be considered as producing the same level of accuracy for the most of the urban classes. The kernel reclassification algorithm presents the same potentiality for residential class discrimination as the neural network approach. The only exception is class “Industrial and/or commercial”. In general, this class was given with low classification accuracy, since there is no a typical texture pattern representing industrial and commercial zones. The rest of the classes are characterised either by their function (e.g., cultural sites, technical network infrastructure, manufacturing industry, heavy industry, marine transports, etc) or structure (e.g., airport, sport facilities, etc.). These land use classes may not be classified by automatic procedures integrating texture and spectral properties. There is a need of ancillary information and visual photo-interpretation or computer vision approaches to be applied. The analysis of image structure, which has been studied on aerial photographs by other scientists, integrating geometric properties, empirical and heuristic knowledge, semantic knowledge and computer vision techniques, should be considered thoroughly, for urban land use classification, especially when very high spatial resolution satellite imagery is used.

References


