



Conference Proceedings Paper

# Feature Investigation for Large Scale Urban Detection Using Landsat Imagery

Fathalrahman Adam \*, Thomas Esch and Mihai Datcu

Earth Observation Center, German Aerospace Center(DLR) ; Münchner Str. 20, Weßling, Germany;  
{fathalrahman.adam, thomas.esch, mihai.datcu}@dlr.de

\* Correspondence: fathalrahman.adam@dlr.de; Tel.: +49-8-153281532

Received: date; Accepted: date; Published: date

**Abstract:** Many works dealing with the problem of urban detection at large scale have been published, but very little attention has been paid to the investigation of the features' relative importance. Feature selection is known to be an NP-hard problem, with many heuristics suggested to approximate the solution. In this paper, a survey of the features used for large scale urban detection is presented, then the question of finding the best subset of features is investigated. Using Landsat scenes of five urban areas, most common features were extracted to represent the full feature set. Employing mutual information based ranking methods, Support Vector Machine (SVM) and Random Forest feature ranking, an importance score was assigned to each feature by each method. To aggregate the individual rankings of features, a two stage voting scheme was implemented to choose a subset of size  $N$  as the most relevant features. The most important features for all five cities taken together were listed.

**Keywords:** Urban detection, large scale classification, feature selection, Landsat

## 1. Introduction

The freely available and easily accessible Landsat archive of more than 40 years, of consistent high resolution multispectral satellite data, offers an unparalleled opportunity for large scale analysis of phenomena observable at Landsat temporal and spatial resolutions. The urban growth is one of these phenomena, with great importance in fields like urban planning, risk and vulnerability analysis, telecommunication and other socio-economical analysis.

Several studies have shown that, Earth Observation can provide valuable data and information to monitor and analyse urbanization, and map the extent of settlements. However, a major issue with any classifier design, regardless of its complexity or robustness, is the feature selection step. This is a specially difficult task in the case of global or semi-global analysis. As the degree of generalization of the classification task increases, larger are to be considered, which results in increase in the inter- and intra-class variability, as more natural samples are added to the data pool. In the particular case of urban detection, the look, the distribution, and even the definition of what comprises urban areas differ across the globe in a significant way. For instance, whether roads are considered urban or not, this changes from one part of the world to another, according to the local definition of urban areas. The semantic disparity adds to the natural variation of the look and distribution of roads across the globe, which is substantial as it is.

Another major example is the buildings class. The appearance of buildings is heavily influenced by many cultural, environmental, economical and social factors, typically vary wildly on a global scale.

There are not many published works in the field of remote sensing doing exhaustive analysis of the feature space. This is first due to data availability, which greatly reduces the choice available

for any research study, and secondly to the field expertise, which tempts researchers to use the typical features for the task in hand, rather than performing a very time consuming feature selection routine. There are, nevertheless, some surveys in the literature for the feature space. Torija and Ruiz [1] have implemented three different methods for feature selection. First method is a correlation based feature-subset selection algorithm, the second one is a wrapper method, using three different induction algorithms. The third method is using PCA as an embedded dimensionality reduction tool, in order to project the features to a lower dimensional space. They did the study to select among the 32 variables relevant in the noise-pollution in the environment, before applying three different classifiers. Tokarczyk et al. [2] did an investigation of features for the case of image segmentation into four classes. PCA and Deep Believe Networks were used as feature reduction techniques. Chan et al. [3] have investigated texture features at 250m resolution. 12 texture parameters were calculated for the infrared band. Different window sizes for the calculation of these features were used. The resulting 60 features were investigated using decision trees as an example of wrapper method.

The purpose of this work is to do an investigation of feature space for the large scale urban detection problem using Landsat data. The aim is to select the feature sub-set which maximizes the classification accuracy, for a reasonable computational cost. This feature selection step will precede and be independent of classifier choice, so the chosen feature set will perform consistently better than any other set, regardless of the used classification algorithm.

## 2. Landsat features for urban classification

Many features have been suggested for the classification of Landsat data, with different achieved accuracy. The most popular features include:

- Raw Landsat pixels [4], [5].
- Spectral indices [6], [7].
- Raw + Indices [8], [9].
- Texture [10], [11].

### 2.1. Raw pixels of Landsat scene

This is a straightforward feature set, just using the spectral signature composed by the raw pixels in the various bands of Landsat to classify the land cover. This can be considered the baseline, with large room for improvement by extracting more specialized features. An obvious problem facing any method employing the raw pixels is the occlusion caused by clouds. There are easy ways to overcome this problem, given enough data and somewhat relaxed constraints for the classification problem. Simple averaging of the pixels over some time period (a year for instance) is ought to reduce the effect of clouds greatly.

### 2.2. Spectral indices

The indices calculated from the Landsat bands emphasize different properties of land cover, according to the wavelength of the bands involved in the index computation. The general formula for these indices is

$$I_{i,j} = \frac{B_i - B_j}{B_i + B_j} \quad (1)$$

where  $I_{i,j}$  is the index between the two bands  $i$  and  $j$ .

$B_i$  is band number  $i$  in the scene.

These indices represent band-pass filters, sensitive for the corresponding wavelengths of the bands involved. As can be seen in the formula, these indices range from -1 to 1. There are generally six bands of the same resolution in Landsat 4,5,7. These six bands are the remaining ones after excluding the panchromatic and the thermal bands, as both have different resolutions compared to the other 30m bands. If an algorithm is to be compatible with old Landsat data as well as the new one, the new

bands introduced in Landsat 8 should be avoided, restricting the used bands to the common ones. Using these six bands, it is possible to calculate 15 different indices with the formula above.

### 2.3. Texture

High texture value is a distinct feature of urban areas, as urban areas typically appear as dense texture against a rather smooth background (e.g. vegetation, bare-land, water). There are different types of texture descriptors in use for the problem of urban delineation, the most common one is the Gray Level Coocurance Matrix (GLCM) [10], [11]. Other more sophisticated texture descriptors are Gabor and Haar textures.

## 3. Method

The goal here is not to use a generic feature ranking method, in order to derive an absolute feature importance ranking; the goal is rather to rank the features we have for the urban detection task, which is a binary classification problem. This condition makes the issue of feature selection much more accessible. The feature selection problem is described as follows

Let the feature vector be  $V_f = \{f_1, f_2, \dots, f_M\}$ , if the classification accuracy of an ideal classifier  $C$  using all  $M$  features is  $\rho(l = C|V_M)$ , for label  $l$ , then we would like to find the indices  $s$  of a subset  $V(s)$  of size  $N$  which maximizes

$$s = \operatorname{argmax}_s (\rho(l = C|V_f(s)))$$

where  $V_N = V(s) \subset V_f$ .

The typical curve of the classification accuracy of any classifier monotonically increases as a function of the number of features in the set  $V_s$ , until it comes to a plateau after all important features have been added. The absolute theoretical maximum is achieved when  $V_s = V_N$ . The order of the problem of feature ranking is  $O(N!)$ , where  $N!$  is the factorial of  $N$ .

We used the two common methods of features selection in this work, these are wrappers and filters.

### 3.1. Filter methods

In this class of methods the features are investigated separately, which implies the assumption of independence between them. This is in fact not the case in most of remote sensing data, as the features typical originate from the same source and therefore correlate. But these are still useful to evaluate the relative importance of features, and it is possible to combine them with other methods, in order to make the feature selection method more efficient.

#### 3.1.1. Information theoretic based methods

All these methods rely on the concept of mutual information, and they calculate the feature importance based on this value. The mutual information (MI) between the two sets  $x$  and  $y$  is defined as

$$MI(x, y) = -H(x, y) + H(x) + H(y) \quad (2)$$

where  $H$  is the entropy, found as

$$H(x) = - \sum_{i=1}^N \rho(x_i) \log_{10} \rho(x_i). \quad (3)$$

Few measures have been used in this category [12].

- Asymmetric Dependency Coefficient (ADC)

$$ADC(x, y) = \frac{MI(x, y)}{H(x)} \quad (4)$$

- Normalized Gain Ratio ( $U_s$ )

$$U_s(x, y) = \frac{MI(x, y)}{H(y)} \quad (5)$$

- Symmetric Gain Ratio ( $S_u$ )

$$S_u = \frac{2 \cdot ADC(x, y)}{H(x) + H(y)} \quad (6)$$

### 3.2. Wrapper methods

Wrapper methods or embedded methods are a family of algorithms which use the classifier itself for feature ranking, what is referred to as *induction algorithm*. This can be demonstrated by Random Forest [13] and SVM [14], both of which can assign an importance index to each feature as part of their supervised training.

### 3.3. Voting scheme

To combine the sub-sets suggested by all above methods, we implemented a two-tire voting scheme to aggregate the ranking of all methods. In the first round the different categories of methods each votes the best  $N$  features according to its measure. This is done by simply sorting the features according to importance, then choosing the  $N$  most important features. This pool of features is further reduced by a second round of voting, the size now is  $5 * N$  features, as the 5 categories of measures will vote  $N$  features each. In this round, the  $N$  statistical modes are taken to be the aggregated decision of all measures. Table ?? lists the voting result for  $N = 4$ .

## 4. Results

### 4.1. Data

To evaluate the methods, five Landsat 8 scenes in Europe were used. The chosen scenes are: Path/Row (183/33) Athens, (193/23) Berlin, (183/29) Bucharest, (193/26) Munich and (199/26) Paris. As a first step towards a global urban feature design, this region in Europe was taken, with scenes from eastern, central and western Europe. The socio-economics of this region is comparable to some degree, and the landscape and urban structures are rather similar.

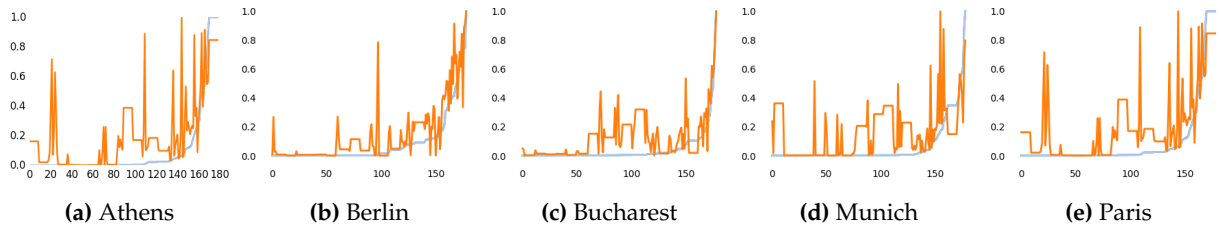
The scene with the least cloud-coverage from the year 2014 for each path/row was chosen, then the cloud masking was performed. The used ground truth was extracted from the Global Urban Footprint (GUF) [15]. GUF is a binary layer which delineates the urban areas on a global scale based on TanDEM-X data.

For the subsequent experiments, 50,000 points were randomly chosen uniformly over each scene, with a constraint that at least 20% of the samples be of urban class. This is to adjust the class ratio, as the urban class covers only 3% of the pixels on average in all scenes, rendering the classification problem highly unbalanced.

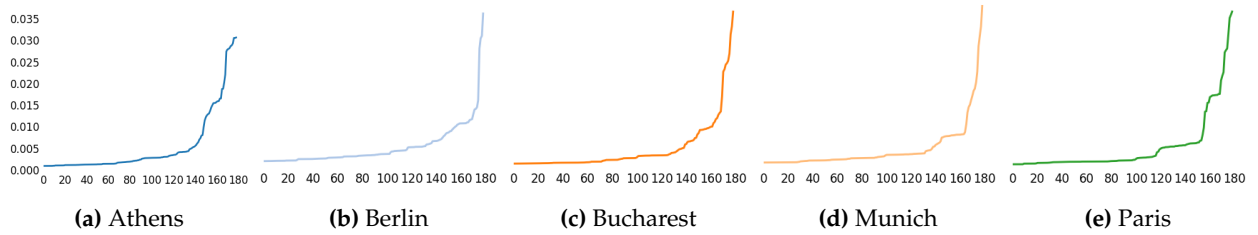
The features computed are all 15 spectral indices, eight GLCM features for each band, and the average and mean of Gabor filter using three scales and three angles. The total number of features is 178 for each scene.

### 4.2. Best features

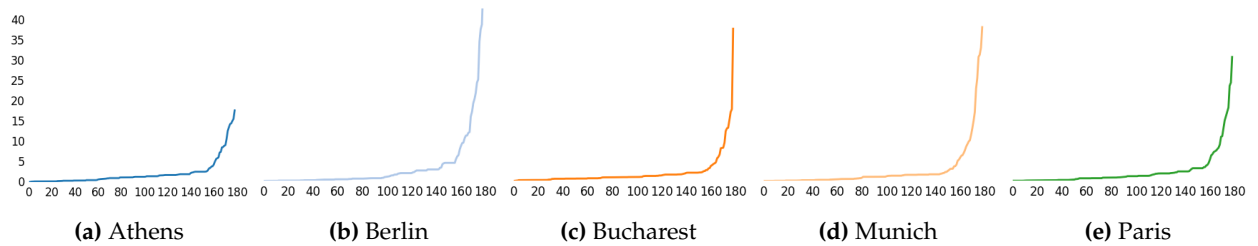
To select the best sub-set of features, we used the different methods discussed in section 3. The ranking of mutual information methods is shown in Figure 1. Each method produces different ranking for the features, the graph depicts this ranking sorted from less important to highest importance. The ranking according to wrapper methods are shown in Figure 2 for the Random Forest, and Figure 3 for SVM based.



**Figure 1.** Ranking of features using information theoretic based measures. Legend: ADC and  $U_s$  are almost identical shown in sky blue,  $S_u$  is the orange curve



**Figure 2.** Ranking of features using Random Forest



**Figure 3.** Ranking of features using Linear SVM

The measures have been sorted separately for better visualization, but the index of each feature is not the same along the different plots.

The best features after applying the two-step voting algorithm are:

Band 0: GLCM Dissimilarity, Energy, Homogeneity.

Band 1: GLCM Homogeneity.

Band 2: GLCM Energy.

Spectral indices: NDWI, SAVI, Index 7, Index 10, Index 14.

## 5. Conclusion

As a first step towards a global urban classifier based on Landsat data, we investigated 178 different features to select a feature set which guarantees good performance locally, and provide good generalization capacity, so it can be used in other areas. Using different feature selection methods, the best features for each city were chosen in a two-step voting process. The chosen features are not the same for all cities, which emphasizes the difficulty of selecting features for a global conventional classifier.

## Bibliography

1. A. J. Torija and D. P. Ruiz, "A general procedure to generate models for urban environmental-noise pollution using feature selection and machine learning methods," *Science of The Total Environment*, vol. 505, pp. 680–693, 2015.
2. P. Tokarczyk, J. Montoya, and K. Schindler, "An evaluation of feature learning methods for high resolution image classification," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 22nd ISPRS Congress, Melbourne, Australia, 2012*.
3. J.-W. Chan, R. S. DeFries, X. Zhan, C. Huang, and J. R. G. Townshend, "Texture features for land cover change detection at 250 m resolution-an application of machine learning feature subset selection," in *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, vol. 7. IEEE, 2000, pp. 3060–3062.
4. C. Li, J. Wang, L. Wang, L. Hu, and P. Gong, "Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery," *Remote Sensing*, vol. 6, no. 2, pp. 964–983, Jan. 2014.
5. X. Li, P. Gong, and L. Liang, "A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data," *Remote Sensing of Environment*, vol. 166, pp. 78–90, Sep. 2015.
6. T. Hu, J. Yang, X. Li, and P. Gong, "Mapping Urban Land Use by Using Landsat Images and Open Social Data," *Remote Sensing*, vol. 8, no. 2, p. 151, Feb. 2016.
7. M. Shimoni, J. Lopez, Y. Forget, E. Wolff, C. Michellier, T. Grippa, C. Linard, and M. Gilbert, "An urban expansion model for African cities using fused multi temporal optical and SAR data," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. IEEE, 2015, pp. 1159–1162.
8. K. C. Tan, H. S. Lim, M. Z. MatJafri, and K. Abdullah, "Landsat data to evaluate urban expansion and determine land use/land cover changes in Penang Island, Malaysia," *Environmental Earth Sciences*, vol. 60, no. 7, pp. 1509–1521, Jun. 2010.
9. J. G. Masek, F. E. Lindsay, and S. N. Goward, "Dynamics of urban growth in the Washington DC metropolitan area, 1973-1996, from Landsat observations," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3473–3486, Jan. 2000.
10. J. Zhang, P. Li, and J. Wang, "Urban Built-Up Area Extraction from Landsat TM/ETM+ Images Using Spectral Information and Multivariate Texture," *Remote Sensing*, vol. 6, no. 8, pp. 7339–7359, Aug. 2014.
11. M. Hall-Beyer, "Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales," *International Journal of Remote Sensing*, vol. 38, no. 5, pp. 1312–1338, Mar. 2017.
12. W. Duch, T. Wiecek, J. Biesiada, and M. Blachnik, "Comparison of feature ranking methods based on information entropy," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. IEEE, pp. 1415–1419.
13. P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir, "Feature selection of time series MODIS data for early crop classification using random forest: A case study in kansas, USA," *Remote Sensing*, vol. 7, no. 5, pp. 5347–5369, Apr. 2015.
14. A. Rakotomamonjy, "Variable selection using SVM-based criteria," *Journal of machine learning research*, vol. 3, pp. 1357–1370, 2003.
15. T. Esch, A. Schenk, T. Ullmann, M. Thiel, A. Roth, and S. Dech, "Characterization of land cover types in terrasar-x images by combined analysis of speckle statistics and intensity information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, p. 1911 – 1925, 2011.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).