Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Estimation of Urbanization Using Visual Features of Satellite Images

László Czúni*, Ágnes Lipovits**, Gábor Seress***
czuni@almos.uni-pannon.hu, lipovitsa@szt.uni-pannon.hu, seressg@almos.uni-pannon.hu
*Department of Electrical Engineering and Information Systems
**Department of Mathematics, ***Department of Limnology
University of Pannonia, Egyetem str. 10., Veszprém, Hungary

Abstract

The effects of urbanization have great impacts on the built and natural environment. There are different approaches to estimate urbanization; our paper discusses an evaluation method based on visual inspection of satellite images. Contrary to previous multispectral methods our approach uses only visible photographic information originating from various sensors and made in different time of the year. The motivation behind this approach is the wide availability of these types of images for researches over the World. Trained classification methods, based on automatic visual feature extraction, are used to describe the content of image blocks considering content such as roads, buildings, and vegetation. The analysis of several classification methods is given based on 41 manually scored reference images.
Keywords: Satellite image analysis, image features, data mining, classification, urbanization score, image processing

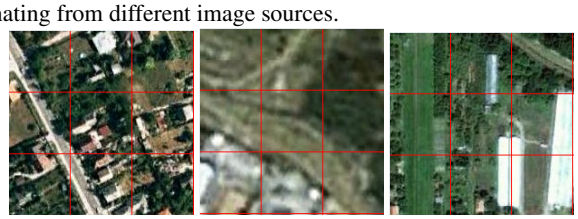
1 Introduction

The analysis of satellite images has great impacts on several areas of life such as urban planning, nature monitoring, agriculture, change detection, military applications. The traditional roaming field-work of biological surveys and ecological monitoring can be efficiently changed by visual evaluation of airborn or satellite images. In this case the visual observer has to evaluate the image pixels (or blocks of pixels) and recognize the different objects such as buildings, trees, fields, water, etc. This human based classification requires some professional expertise but can also be solved by machine vision techniques if multispectral information is utilized [17]. However, for many users only visual aerial/satellite information is available from GoogleEarth, from Bing Maps or from other public databases. The access to multispectral information is limited for many users and for many locations. Airborn images are also available only for some given locations while almost the whole Earth is covered by satellite images freely available. In our research we try to substitute an existing manual method, used for the estimation of urbanization, using only visual photographic satellite images and marginal manual work.

The effect or urbanization can be observed by several natural phenomena. Liker et al [13] uses data records of more than 10 years to investigate the change of weight, size, and condition of birds along the urbanization gradients. They observed house sparrows but other living beings or other natural or artificial phenomena can be the target of research of environmentalist or other researchers. In [4] the birds from areas of different urbanization are compared considering the relation between their condition and competitive behavior. In [14] the urbanization of a region is estimated with the help of the ratio of buildings, vegetation and roads.

Our purpose is to support such applications with an algorithm used for the automatic classification of satellite images, with different physical resolution, using no multispectral data. Some of the typical images used in our experiments are shown on Figure 1.

Figure 1: Example of 9 neighboring blocks of three images originating from different image sources.



While our approach has ecological motivations it can be used for other surveys such as change detection, or it can be used for the support of pixel-based object recognition by giving probability information about land cover for image blocks.

2 Related techniques

A typical approach of land cover and land usage estimation can be found in [8] where sensors of 3 satellites were used, namely Landsat-5 TM, SPOT-XS and Pan, IRS-1C and IRS-LISS, with several spectral bands. The classification method was trained with the help of ortophoto data supporting that visual information can be very useful for ground truth data generation about land cover and land usage.

In [11] vector geographic data, Spot-5 images, and NDVI are used for urbanization change detection. The algorithm proposed, besides NDVI classification, applies structure and texture extraction with Gabor filters. Fuzzy clustering and simple logical operations are used to fuse structural and NDVI information before the final step of change detection.

While [15] uses different images of six satellites (including NDVI and water index information), road maps, historical maps and applies only qualitative analysis, the method of [17] applies segmentation and segment description of multispectral images originating from Landsat TM. The value of mean, standard deviation of each band of poligons is manualy trained then Random forest [6] is used for classification. 10 segment types are specified but no information about the success of training or testing is given, only the changes of segment areas in time is described.

2.1 Manual estimation of urbanization

The manual characterization process is applied on images of physical size of 500m $\times$ 500m cut into 10 $\times$ 10 image blocks [13]. The content of each block is evaluated considering the type of land cover such as buildings, vegetation, and roads:

$$R_b = \begin{cases} 0 & \text{if ratio of buildings is 0\%} \\ 1 & \text{if ratio of buildings is between 0\% and 50\%} \\ 2 & \text{if ratio of buildings is above 50\%} \end{cases} \quad (1)$$

$$R_v = \begin{cases} 0 & \text{if ratio of vegetation is 0\%} \\ 1 & \text{if ratio of vegetation is between 0\% and 50\%} \\ 2 & \text{if ratio of vegetation is above 50\%} \end{cases} \quad (2)$$

$$P_r = \begin{cases} 0 & \text{if no solid road is present} \\ 1 & \text{if solid road is present.} \end{cases} \quad (3)$$

Based on the result of the three classifications all images get a 5 dimensional descriptor $D = [\hat{R}_b, \hat{R}_v, N(R_b = 2), N(R_v = 2), N(P_r = 1)]$ calculated as follows:

$$\hat{R}_b = \frac{\sum R_b}{100},$$
$$\hat{R}_v = \frac{\sum R_v}{100},$$
$N(R_b = 2)$ is the number of blocks where $R_b = 2$,
$N(R_v = 2)$ is the number of blocks where $R_v = 2$,
$N(P_r = 1)$ is the number of blocks where $P_r = 1$.

Finally, to get a scalar urbanization score used in ecological studies, the first principal component ($PC1$) is calculated from the 5 dimensional vector D.
Undoubtedly, the advantage of this manual method is the application of very simple classification rules with large tolerance ranges easy to carry out for large number of images. The main purpose of our research is to find a semi automatic estimation method that can substitute the time consuming and exhausting manual work.

3 Classification of rectangular regions

One possible solution would be to find ways to detect buildings, roads and vegetation for each pixel then to calculate the ratio of the three types of objects. However, these objects can have very different visual appearance depending on image resolution, time of the day, season, weather, type of vegetation, type of building structure, etc. On the other hand the manual annotation, used for training, would require a very meticulous and tiring manual work. To avoid this problem we propose a method for the classification of image blocks, as defined by the manual method, but based on more than 50 visual features almost automatically generated from the color images.

3.1 Calculation of visual features

The different appearance of the various objects makes it difficult to find the right characterization of them. For this reason we measure 52 different feature values based on color, texture, and local contrast (edges and corners) information and apply classification models on high dimensional data vectors. In the following we give a brief description of the feature generator image processing methods.

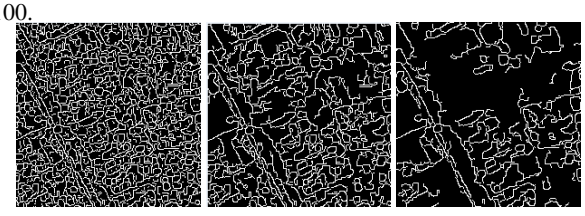
For each block the following features are calculated:

- number of edge points detected by the Canny detector [7] applied in 19 different settings (19 values);
- number of points belonging to each of the 5 specified segment classes: grass, tree/bush, building, road, others, with 4 settings of the Laws classifier (20 values);
- number of corner points detected by the Harris corner detector [10] applied in 6 different settings (6 values);
- average value of Red, Green and Blue channels within a block (3 values)
- average and modus of Hue (of the HVS color space), and the corresponding variance of Hue (4 values).

3.1.1 Number of edge points

The presence of edges can well characterize roads and buildings. For the detection of edge points the Canny detector [7] was used with different settings. The *low threshold* limit was set to 10, 15, 20, 25, ..., 100, while the upper value was set to 3$\times$ *low threshold* during the hysteresis thresholding. Finally the number of edge points was counted in the image block for the 19 cases. Figure 2 shows three examples with *low threshold* = 10, 50, and 100 respectively.

Figure 2: Three examples of the output of the Canny edge detector applied on the left of Figure 1 with threshold 10, 50, and 100.



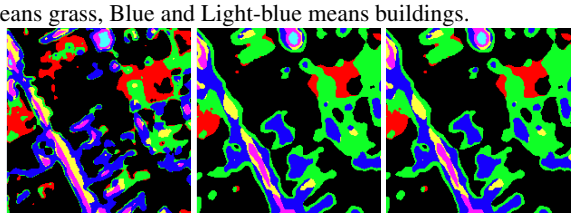
3.1.2 Laws texture classification

The Laws texture segmentation technique [12] uses 9 different convolution kernels to calculate the E_k energy at pixel (i, j):

$$E_k(i,j) = \frac{1}{(2\omega+1)^2} \sum_{x=i-\omega}^{x=i+\omega} \sum_{y=j-\omega}^{y=j+\omega} |K_k(x,y)|, \quad (4)$$

where K_k is the result of one of the convolution, $k = 1,2,...9$ and $\omega = 3,5,7,15$ is the radius of the averaging window.
We use manually selected training points to characterize specific object regions with their energy E_k then classify all pixels of a block to the closest training point using Euclidean metrics. 5 types of objects were trained (buildings, roads, trees/bushes, grass, others) with not more than 3 examples per image. Finally the number of pixels belonging to one of the three classes (road, vegetation, building) were counted and recorded for each rectangular block. Figure 3 illustrates 3 outputs for the left of Figure 1. We should note here that this is the only step where manual operation is required to select a very few, typically 3-6, training points per image.

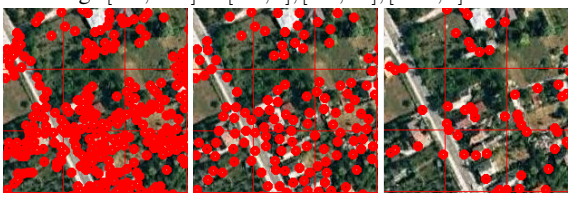
Figure 3: Three examples of the output of the Laws classifier on the left of Figure 1 with $\omega = 3,5$, and 7 respectively. Black means trees, Yellow and Pink means roads, Red and Green means grass, Blue and Light-blue means buildings.



3.1.3 Harris edge detection

One might think that corners can easily characterize man-made objects. The Harris edge detector [10] is one of the most efficient and most often used corner detector in image processing. Two types of parameters were varied to generate different outputs of the detector: the *eigenvalue threshold (ET)* was set to 0.02 and 0.16 and the *minimal distance of points (MD)* was set to 5, 10, and 20. The number of corner points per image block was used as descriptors for the classification. See Figure 4 for illustration.

Figure 4: Three examples of the output of the Harris detector with settings $[ET, MD] = [0.2,5],[0.2,10],[0.16,5]$.



3.2 Classification methods

The three classification tasks ($\hat{R}_b$,$\hat{R}_v$, $\hat{P}_r$) can be solved with several methods. In our experiments we tested 9 methods (C5, KNN, SVM, logistic regression, neural networks, Bayes network, discriminant analysis, C&R tree, and CHAID) but in our paper we describe the performance only of the 3 best basic methods and their combination got with information fusion.

- C5 decision trees: The commercial and improved version of the ID3 algorithm selects testing attributes by information gain; it allows missing values and can handle weighted variables [5]. It generates non-binary decision trees and applies pruning to avoid over-learning. Its performance can be usually improved by boosting [16].
- KNN algorithm: The advantage of the well-known K Nearest Neighbor algorithm is its simplicity and its fast operation. Its disadvantage is that it gives no importance value to the attributes. It can handle the case when some arguments are missing.
- Support Vector Machines (SVM) [9]: Contrary to KNN the Support Vector Machines technique can accept records with missing variables while trying to find the best separating hyperplanes to carry out classification.

4 Experiments and evaluation

4.1 Information about input images

We used 41 images freely available by GoogleEarth. All images were from 5 different locations in Central Europe, each covering 500m $\times$ 500m rectangular areas. We could not identify the satellites making the images only the image sources, as Table 1 shows. The different areas were captured at different seasons of the year as clearly visible on the vegetation. Table 2 gives the seasonal distribution of the 41 pictures. We had no information about the native surface resolution of the images; we normalized the image size for 134 cm / pixel. Three example of satellite images of different nature are illustrated on Figure 1.

Table 1: Sources of satellite images obtained by GoogleEarth.

Source	Number of pictures
Digital Globe [1]	21
GeoEye [2]	14
NASA [3]	6

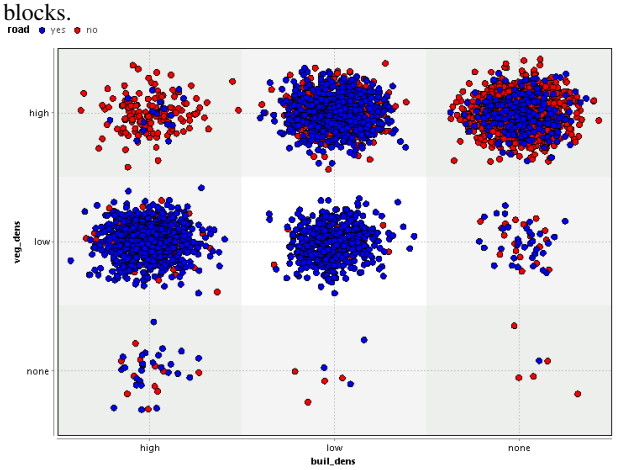
Table 2: Seasons of satellite images used in our tests.

Season	Number of pictures
Spring	16
Autumn	9
Summer	13
Winter	3

4.2 Training data

The manual annotation of training data was made by an experienced PhD student of ecology. The result of the manual classification of the 41 site locations (each having 10 $\times$ 10 blocks) is shown on Figure 5 where each point corresponds to one image block and the axis and color represent the type of classes. Please note that the classification method described above can be considered as a relatively robust decision making. In Figure 5 x-axis

Figure 5: The results of manual classification of the 4100 image blocks.



is for building density (high–low-none), in the y-axis the vegetation is specified (none–low–high). Colors code the presence of roads.
50% of the 4100 image blocks were randomly selected for training and the other 50% for testing for 5 times. The results given below are the average of the 5 tests.

4.3 Classification results

In the following sections we show the classification results including the confusion matrices and the list of the 7 most important explanatory attributes.

4.3.1 Estimation of vegetation

The best performance achieved in the case of the estimation of vegetation was 83.78% with the SVM technique. Table 3 gives the classification of the tested methods while Table 4 shows the confusion matrix. It can be seen that the combined method (C5+KNN+SVM), based on confidence weighted majority votes, could not exceed the SVM. The most important explanatory variables were:

- The Canny detector with the highest threshold value;
- The average of the Red and Blue color channels;
- Number of pixels belonging to buildings, trees, and grass estimated with the smallest window setting of the Laws technique;
- The number of corner points with the minimal distance.

Table 3: Classification of vegetation.

Method	Training	Testing
C5	94.33%	79.17%
C5 boosting	99.75%	81.93%
KNN	86.95%	81.69%
SVM	86.14%	**83.78%**
C5+KNN+SVM	89.81%	83.73%

Table 4: Confusion matrix for the SVM method for vegetation.

Training	0%	<50%	$\geq$ 50%
0%	9	6	7
<50%	1	534	126
$\geq$ 50%	0	136	1173
Testing	**0%**	**<50%**	**$\geq$ 50%**
0%	0	23	6
<50%	0	562	154
$\geq$ 50%	1	158	1204

4.3.2 Estimation of the presence of roads

The classification of roads could be greatly improved with the help of vegetation class information, obtained by a preceding classification. This is clearly visible if comparing Table 5 and Table 6. The confusion matrix of Table 7 tells us that the error rate is low and the main source of problem is that some roads are not recognized (20% more such errors occurred than roads detected at areas with no roads). This type of error will also be illustrated later in Table 10 and on Figure 7. The most important explanatory variables were:

- Vegetation label;
- The Canny detector with 3 higher threshold;
- The average of Blue channel;
- Number of pixels classified to building with the greatest window setting of the Laws technique;
- The modus of the Hue value.

Table 5: Classification of roads.

Method	Training	Testing
C5	82.28%	75.28%
C5 boosting	86.24%	77.7%
KNN	82.28%	66.22%
SVM	77.86%	77.42%
C5+KNN+SVM	81.38%	**77.85%**

Table 6: Classification of roads using vegetation information obtained by preceding classification.

Method	Training	Testing
C5	94.63%	87.1%
C5 boosting	99.2%	**90.61%**
KNN	87.95%	82.59%
SVM	89.31%	89.23%
C5+KNN+SVM	92.62%	88.9%

4.3.3 Estimation of buildings

The worst classification rate was achieved in case of buildings: the SVM ended up with 72.82%. Unfortunately, the usage of

Table 7: Confusion matrix for roads with C5 boosting.

Training	not present	present
not present	952	4
present	12	1024
Testing	not present	present
not present	904	88
present	110	1006

labels obtained by road and vegetation classification could not improve this recognition rate. The most important explanatory variables were:

- The number of corner points with both quality setting with the smallest distance;
- The Canny detector with the highest threshold;
- Variance of the modus of Hue;
- The average of the Blue channel;
- Number of pixels belonging to buildings and roads estimated with the smallest window setting of the Laws technique.

Table 8: Classification of buildings.

Method	Training	Testing
C5	77.61%	71.82%
C5 boosting	91.47%	70.89%
KNN	87.25%	68.26%
SVM	75.85%	**72.82%**
C5+KNN+SVM	83.08%	72.77%

Table 9: Confusion matrix for the SVM method for buildings

Training	0%	<50%	$\geq$ 50%
0%	611	60	4
<50%	75	669	74
$\geq$ 50%	11	86	402
Testing	0%	<50%	$\geq$ 50%
0%	600	94	5
<50%	126	573	170
$\geq$ 50%	13	172	355

4.4 Further analysis

Considering all three classification tasks on the 4100 image blocks the overall error rate was 16% during training and 18% during testing. The relative error of the urbanization score, compared to the manual results, is 12.5% and the relative score based position of the 41 images changed 6% (around 2.5 places) in average with the proposed semi automatic method.
The only manual interaction was the selection of 3-6 training points on all images for the Laws classifier. The usage of static training patterns, allowing the fully automatic operation, was outside the scope of our recent research but should be investigated in future.
By analyzing the images with the highest and lowest error rate we found that most errors occurred at highly urbanized regions. The left of Figure 6 shows the image with the highest error rate of the 41, and the right of this figure shows the one with the lowest error.

Figure 6: Left: image with the most classification errors. Right: image with the lowest error rate.

One reason for the large number of errors can be found in the presence of faults of subjective manual classification (in training and in the generation of ground truth data). As illustrated in Figure 7 and in Table 10 the manual operator did not recognize the road in shadow in blocks 90 and 100 while its presence is supported by the road map information obtained from Google Maps shown in the right of Figure 7. On the other hand the very similar image of blocks 89 and 99 are scored 2 and 1 respectively considering the ratio of buildings.

Figure 7: First row: Cell 89, Cell 90, Cell 89 with road highlighted, Cell 90 with road highlighted. Second row: Cell 99, Cell 100, Cell 99 with road highlighted, Cell 100 with road highlighted.

Table 10: Manual (reference) scoring of some image blocks of left of Figure 6.

Cell ID	Buildings	Vegetation	Roads
89	2	1	1
90	1	2	0
99	1	1	1
100	1	2	0

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

5 Conclusion and future work

Our purpose was to create a semi automatic way of urbanization estimation based on previously published ecological motivations. The greatest challenge of the research is whether it is possible to use only color images originating from different sensors with different resolutions taken at different seasonal and weather conditions.
At the current state of our work we found that the classification error rate is around 18% which makes a 12% error in the final urbanization score. However, there are lots of areas where we can improve our method in future:

- we plan to introduce new descriptors to detect the crossings of edges possibly belonging to buildings;
- we plan to incorporate regional neighborhood information into our model;
- we plan to incorporate road map information, easily available for many areas of the world, by Google Maps.

We also believe that the proposed method can be used to generate regional support information for other pixel-based object detection methods.

6 Acknowledgement

The work and publication of results have been supported by the Hungarian Research Fund, grant OTKA CNK 80368 and by TÁMOP-4.2.2/B-10/1-2010-0025. We also say thanks to János László Tóth for the help in data generation.

References

[1] Digitalglobe. http://www.digitalglobe.com/.

[2] Geoeye. http://www.geoeye.com/.

[3] National aeronautics and space administration. http://www.nasa.gov/.

[4] V. Bókony, G. Seress, S. Nagy, Á.Z. Lendvai, and A. Liker. Multiple indices of body condition reveal no negative effect of urbanization in adult house sparrows. *Landscape and Urban Planning*, 104:75–84, 2012.

[5] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Chapman and Hall, New York, 1984.

[6] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[7] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[8] H. Gonca Coskun, Ugur Alganci, and Gokce Usta. Analysis of land use change and urbanization in the Kucukcekmece water basin (Istanbul, Turkey) with temporal satellite data using remote sensing and GIS. *Sensors*, 8(11):7213–7223, 2008.

[9] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

[10] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[11] V. Lacroix, M. Idrissa, A. Hincq, H. Bruynseels, and O. Swartenbroekx. Detecting urbanization changes using Spot5. *Pattern Recognition Letters*, 27(4):226 – 233, 2006. Pattern Recognition in Remote Sensing (PRRS 2004).

[12] K. Laws. *Textured image segmentation*. Dept. of Electrical Engineering, University of Southern California, 1980.

[13] A. Liker, Z. Papp, V. Bókony, and Á.Z. Lendvai. Lean birds in the city: body size and condition of house sparrows along the urbanization gradient. *Journal of Animal Ecology*, 77:789–795, 2008.

[14] M. J. McDonnell and A. K. Hahs. The use of gradient analysis studies in advancing our understanding of the ecology of urbanizing landscapes: Current status and future directions. *Landscape Ecology*, 23:1143–1155, 2008.

[15] Bach Viet Pham, Dinh Duan Ho, Raghavan Venkatesh, and Shibayama Mamoru. Using satellite imagery to study spatial urban expansion of Hanoi city. In *Proceedings of the International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, 2006.

[16] Robert E. Schapire. *The boosting approach to machine learning: An overview. In D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, B. Yu, editors, Nonlinear Estimation and Classification*. Springer, 2003.

[17] Thi An Tran and Anh Tuan Vu. Application of remote sensing in land use change pattern in Da Nang city, Vietnam. In *Proceedings of the International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, 2008.