Supervised Classification of Multi-Spectral Satellite Images Using Neural Networks*

Tomasz Kubik†
tel: 320-2549, e-mail: tkubik@ict.pwr.wroc.pl
Witold Paluszyński†
tel: 320-2741, e-mail: Witold.Paluszynski@ict.pwr.wroc.pl
Adam Iwaniak‡
tel: 320-5686, e-mail: iwaniak@ar.wroc.pl
Przemysław Tymków‡
tel: 320-5686, e-mail: przemyslaw.tymkow@wp.pl

†Institute of Engineering Cybernetics
Wrocław University of Technology
Janiszewskiego 11/17, 50-372 Wrocław, Poland
fax: (71)321-2677
‡Agricultural University of Wrocław
Norwida 25/27, 50-375 Wrocław, Poland
fax: (71)320-5617

Abstract

Satellite images provide a wealth of information, but this information is not easy to use, so image processing methods and models targeted at satellite images processing are being developed. A rudimentary technique of satellite data interpretation is to compare the reflectance of the surface and objects there located to various spectral fractions of light. This paper reports on the results from a project where neural networks and other data processing techniques are being applied to satellite images to obtain data which could be used as source to the *Geographic Information Systems (GIS)*. Basic satellite image properties are presented and discussed. Then the process of data preparation and building artificial neural networks to process them is described. Finally, the results obtained are presented and evaluated.

Keywords: pattern classification, neural networks, satellite image interpretation

1 Introduction

Satellite images are an invaluable source of information which can be used for: navigation, weather forecasting, tracking environmental changes, and in various forms to support administrative decision making, urban, industrial, and even military planning, etc. Initially providing a resolution of 1×1km, now commercially available satellite images go down do 60×60cm. While this is still not as good as the finest resolutions provided by aerial photographs, satellite images can be acquired much faster.

*This research was supported in part by the Committee for Scientific Research through grant 5T12E02924

They provide a wealth of information, but this information is harder to use, so there is a growing demand for automatically processing such images in order to extract larger features from them. Therefore image processing methods and models targeted at satellite images are being developed.

This paper reports on the results from a project where neural networks and other data processing techniques are being applied to satellite images to obtain data which could be used as source to the *Geographic Information Systems (GIS)*, whose main tasks are to: collect, store, process, analyze, and visualize spatial data for the purpose of: spatial and ecological planning, providing mapping data for sale to institutions and the public, local administration support, emergency management, etc. In order to accomplish all these goals a high level of automation of data processing is required.

A rudimentary technique of satellite data interpretation is to compare the reflectance of the surface and objects there located to various spectral fractions of light. For uniform areas this method works very well taking advantage of unique spectral signatures of various types of land. It fails however for areas with a high heterogeneity of spectral response, such as urban areas.

band 2 band 6 band 7

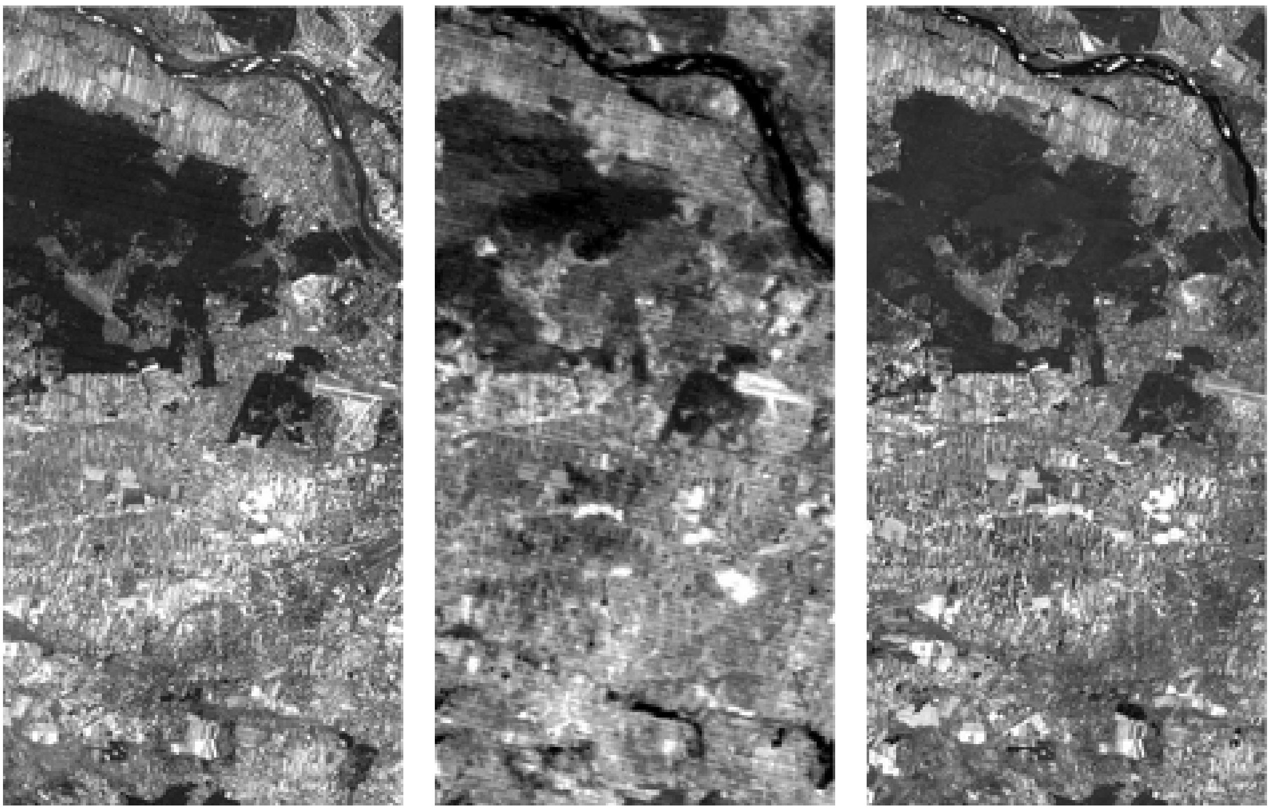

Figure 1: Landsat TM images of section of the city of Warsaw: band 2 is at $520 - 600$nm (visible green), band 6 is at $10.40 - 12.50\mu$m (thermal infrared), and band 7 is at $2.08 - 2.35\mu$m (middle infrared), the spatial resolution is 30×30m for all bands except 7, with 120×120m resolution, here resampled also at 30×30m.

Figure 1 shows representative images from the Landsat satellite TM scanner across the spectral range from the visible to thermal infrared regions of the spectrum. Landsat TM images provide 7 bands at of TM images is 30×30m resolution (except the thermal band 6). The radiometric resolution for all Landsat TM bands is 8-bit.

The multispectral analysis in GIS can be performed in several ways. In many cases the central role is played by the human operator, interpreting the data based on possessed knowledge and other data, such as topographics maps, digital terrain models, etc.

In most common approaches automatic land use classification is performed in two consecutive steps: *per-pixel* classification (which is a surface type classification) and *per-parcel* classification (which is a land use classification). Sometimes automatic land use classification is an aggregation of these two steps in one computational procedure.

One can observe that implementation of the first step must involve some data clustering methods, where values of pixels intensities constitute a features space. The feature space dimension is equal to the number of images taken, i.e. number of spectral channels.

But the second step needs a different approach. It must include computation of local areas statistical features at first. Then these features can be used for classification. Such methodology is closely related to the resolution of the problem which is known in the image processing community as *image segmentation*. This problem relies on partitioning of an image into related sections or regions regarding color, texture, intensity, etc.

2 Methodology

***Per-pixel* classification** relies on: (i) identification of distinctive classes corresponding to the different land surface types (classifier design) (ii) allocation of a particular class membership to each pixel (pixel classification). Implementation of these two points with respect to satellite (aerial) imagery generate number of questions. These are: "What are the classes we are interested in on this particular imagery? What can help in these class identification? Which spectral bands offer the best set of distinctive data? How are classes distributed in the feature space?"

The first question can be answered partially by the definitions included in the European Union CORINE classification system. This system contains 44 land-use classes defined in a hierarchical structure (CORINE land use-classes), which are broadly grouped into 5 major (first level) categories: artificial surfaces, agricultural areas, forests and semi-natural areas, wetlands and water bodies. But this structure contains artificially introduced classes which often incorporate regions of the same spectral characteristic. Such classes are difficult to separate when doing it in *per-pixel* mode and need to be redefined. The environmentally meaningful classes [1] created in that way may be better suited for *per-pixel* classification task.

Identification of distinctive classes on spectral image (i.e. classifier design) can be done with or without supervision. In supervised mode it is assumed that there is one image with all pixels already classified. This image with a set of corresponding images in all spectral bands can be used to builds classification rules. In unsupervised mode a cluster analysis is performed. Classifiers are defined then by the methods used for data clustering and centers of clusters found. Several approaches for supervised and/or unsupervised classifier design for *per-pixel* multispectral image analysis were discussed in the literature. This problem was tackled with the aid of Artificial Neural Networks, [2, 3], clustering algorithms (K-means, Gaussian maximum-likelihood, mixture discriminant analysis, fuzzy-c) [3, 4, 5, 6]. In some of these approaches *per-pixel* classification was combined with *per-parcel* classification. A hybrid segmentation procedure to integrate contextual information with per-pixel classification in a metropolitan area land cover classification project was described and evaluated in [7].

The spectral bands of remote sensors were designed mainly for vegetation and water detection. To enhance vegetation detection spectral bands are often combined [8, 9]. This combination might be:

- normalized difference vegetation index $NDVI = \frac{NIR+Red}{NIR-Red}$ (where Red is the red band and NIR is near-infrared band);
- three Tasseled Cap parameter channels (brightness: weighted sum of six channels; greeness: difference between the total reflectance in the near infrared bands and in the visible bands; wetness: difference between the total reflection capacity between the visible-near infrared channels and the shortwave infrared channels);
- image fusion $B_i = \left(\frac{MS_i}{\sum_i MS_i}\right) P$ where B_i is a fused channel, MS_i is the i multispectral channed to be fused, P is the panchromatic band and $\sum_i MS_i$ is the summation of all multispectral channels.

Per-parcel **classification** in general operates on larger areas rather than on points. It should provide categories of areas with respect what they are being used for by humans. This type of classification makes use of the results of *per-pixel* classification.

There are different approaches to implement *per-parcel* classification. Regarding the way of features extraction we have methods: (i) operating on the raster imagery with spatial kernels (like in textural analysis GLCM method [10, 11, 12], Gibbs models and Markov random field [6], multifractality [13], variogram [14]) (ii) transforming raster into an object-based representation with subsequent contextual analysis (like in expert systems or contextual based classification [15]) (iii) combining former both (like in vector and raster data coupling [16]). Regarding classifier design: supervised and unsupervised (ANN [17, 3, 18]; k-means and Gaussian maximum-likelihood classifier [3], region-growth algorithm[19].)

3 Data preparation

The images used in the experiments described here are obtained form Landsat satellite (figure 1) and produced by the TM (Thematic Mapper) scanner, which is a multispectral scanning system. The TM sensor records reflected electromagnetic energy from the visible to thermal infrared regions of the spectrum. The spatial resolution of TM images is 30x30 m for all seven bands except the thermal band 6, which has 120x120 m resolution.

The experiments started using the CORINE project land cover map as a reference material for supervised training. The CORINE Program (Coordination of Information on the Environment) database includes, among other things, information about land covers. The CORINE classification consists of twenty-two classes of land use. After conducting some initial experiments, it turned out that this material is insufficient. On the one hand, the twenty-two classes it contained classes which are distinguishable only by abstract attributes, and not by physical features, such as: leafy forests, and parks within a city. On the other hand, it turned out to contain numerous errors and omissions.

Therefore, an improved reference classification map was hand-created by an interpreter. The resulting map contained only 11 different classes of land use (figure 2).

A grey-level frequency analysis was performed on this map, producing histograms for all the reference map classes in all spectral bands, figure 3. Comparing the histograms for different classes one can see some separation in various spectral bands. However, this separation in some cases is minimal. On the other

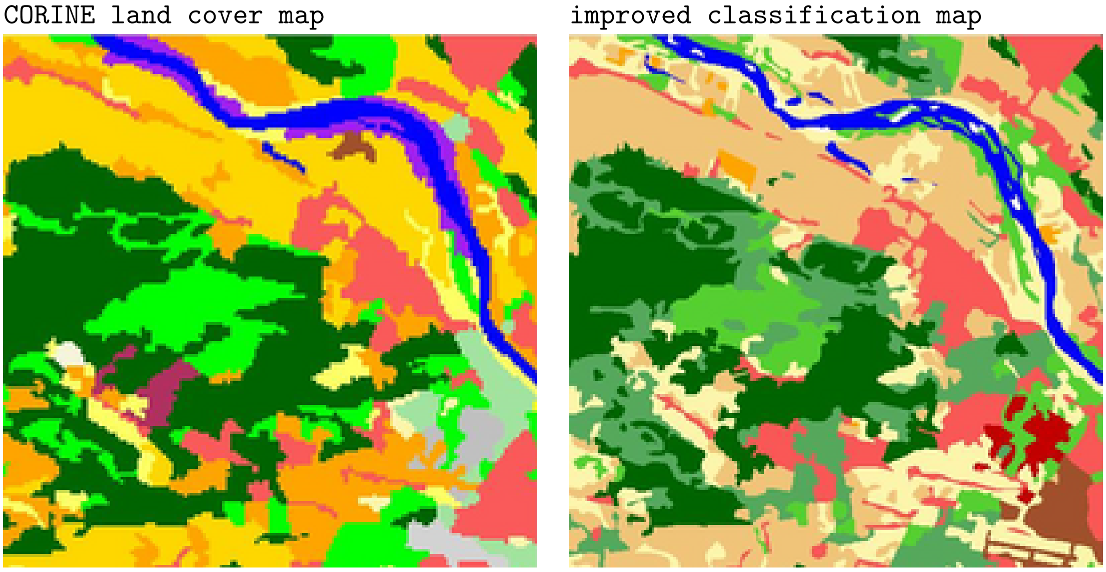

Figure 2: Reference images used in the experiments: the original CORINE project map and the simplified and corrected map created for the experiments described in this paper.

hand, the overlap between the combination of spectral response curves of different classes indicates, that classification is going to be difficult, and any simplistic approaches have a limited chance of success.

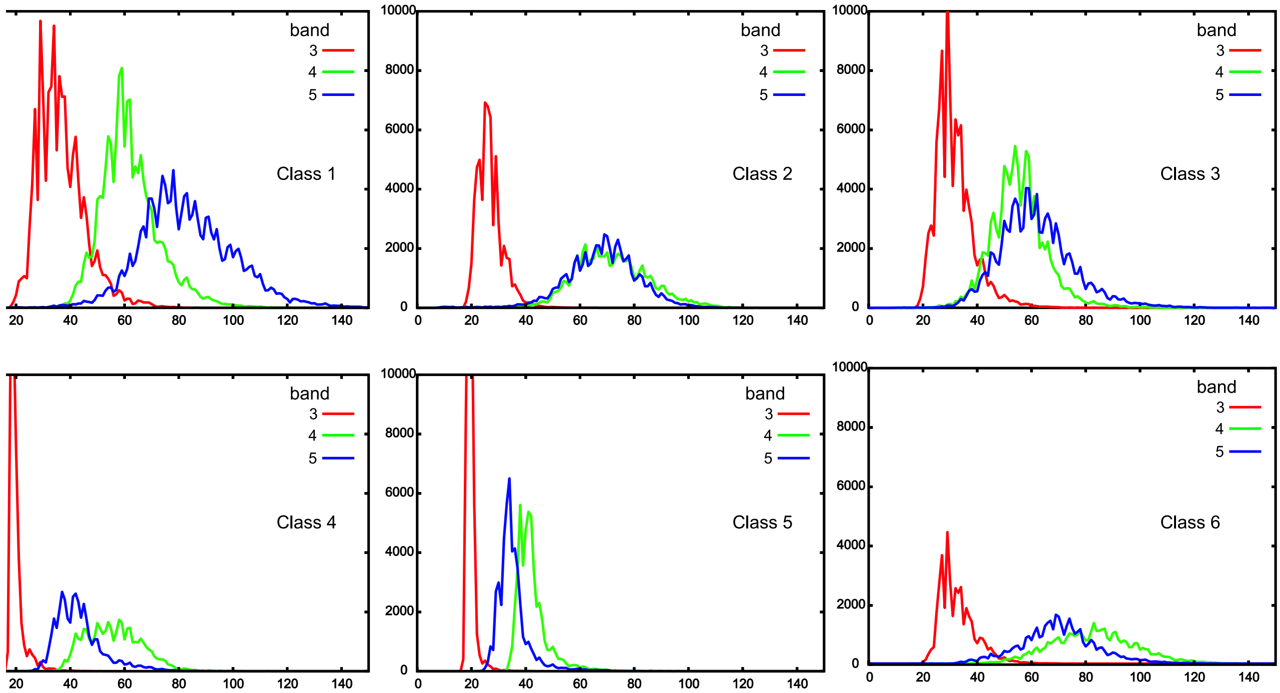

Figure 3: Histograms of 6 selected classes in 3 spectral bands: the horizontal axis designates grey levels, and the vertical axis shows the number of pixels for each class. The difference between the curves for different classes offers the possibility to distinguish between them, but in some cases this separation is minimal, eg. class 3 and 6.

4 Neural networks

The main aim of the preliminary experiments was defining possibilities of using neural networks in area classification. There are a few phases of the classification process:

- preparing data sets and patterns from images,
- building and training a network,
- processing and visualizing results,
- estimating the classification error.

The patterns files were usually made up of all Landsat spectral images and one raster image which is the map of area classification. The reference images have been split into two parts: the training set and the testing set, for result evaluation.

The next step was to build an optimal network architecture which would optimally resolve classification problem. A number of different architectures and sizes of *feed-forward* structures were tried by training them and comparing results on the training set, figure 4.

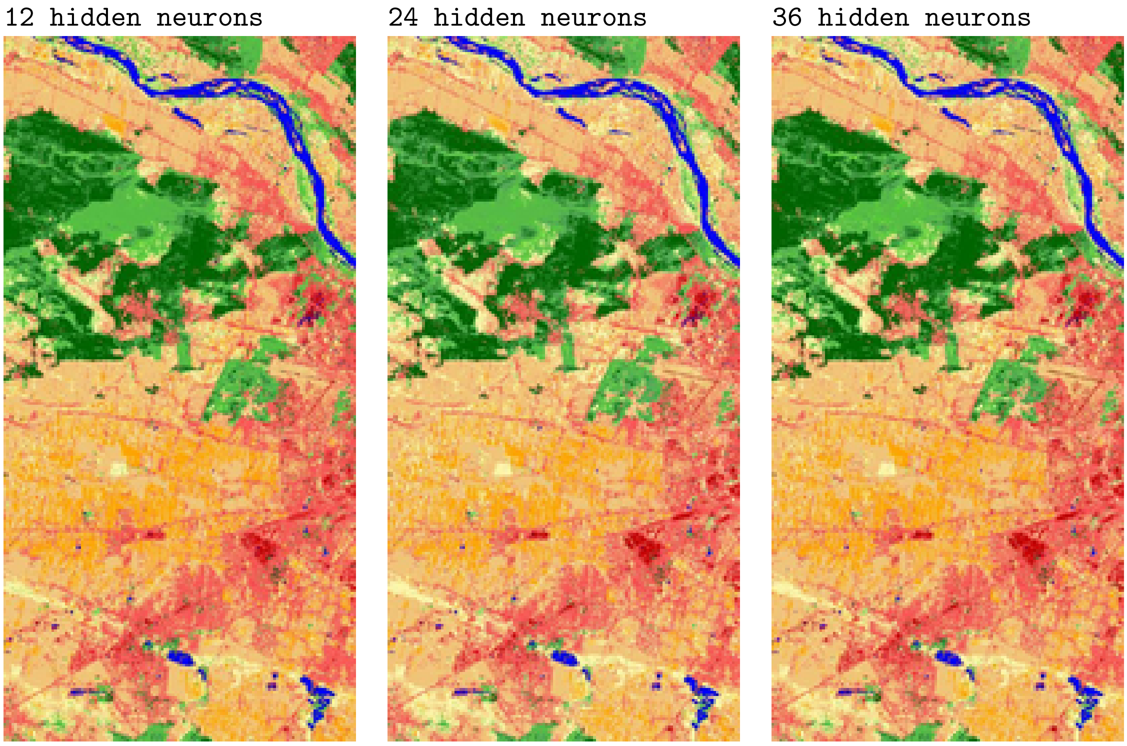

Figure 4: A comparison of different neural network architectures: the classification results for the training set. This helped determine an optimal neural network architecture for the computing task.

5 Quality measures

A number of experiments have been conducted resulting in many classification maps generated. They have been evaluated by human experts and found to be very valuable, readily comparable to the human-processed CORINE map and in some areas more accurate. It would be useful, however, to quantify the quality of the results, and to provide an evaluation framework, which could be used in working out optimal parameter settings. For this several statistical measures have been considered.

- Root mean squared error (RMSE)[4]

$$RMSE_j = \sqrt{\sum_{i=1}^{n} (\hat{x}_{ji} - x_{ji})^2 \times \frac{m_i}{M}} \qquad (1)$$

  where $\hat{x}_{ji}$ and x_{ji} are the predicted fraction and measured fraction for land cover class j in site i; m_i is the size of site i in number of pixels, M is the size of the whole testing data in number of pixels.

- For the confusion matrix $A = [a_{ij}]$ where a_{ij} is the number of sample pixels from the jth class that have been classified as belonging to the ith class, several coefficients have been proposed[20, 21]:
  - the users accuracy of class i: a_{ii}/a_{ri}, where $a_{ri} = \sum_i a_{i.}$ (sum of ith row entries);
  - the producers accuracy of class i: a_{ii}/a_{ci}, where $a_{ci} = \sum_i a_{i.}$ (sum of ith column entries);
  - the overall accuracy of the method: $\sum_i a_{ii}/a_t$, where a_t is the total number of pixels;
  - simple Kappa coefficient:

$$\hat{\kappa} = \frac{P_o - P_e}{1 - P_e} \quad , \text{ where } \quad P_o = \sum_i aii/a_t, \quad \text{and} \quad P_e = \sum_i a_{ri}a_{ci}/a_t^2 \qquad (2)$$

  - weighted Kappa coefficient:

$$\hat{\kappa}_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}} \quad , \text{ where } \quad P_{o(w)} = \sum_i \sum_j w_{ij}a_{ij}/a_t, \quad \text{and} \quad P_{e(w)} = \sum_i \sum_j w_{ij}a_{i.}a_{.j}/a_t^2 \qquad (3)$$

  The weights w_{ij} are constructed so that $0 <= w_{ij} < 1$ for all $i \neq j$, $w_{ii} = 1$ for all i, and $w_{ij} = w_{ji}$.

  Although there are no absolute cutoffs for kappa coefficients, two sources provide some rough guidelines for the interpretation of kappa coefficients. According to the one of them [22] values exceeding .75 suggest strong agreement above chance, values in the range of .40 to .75 indicate fair levels of agreement above chance, and values .40 are indicative of poor agreement above chance levels. In the other proposition [23] the following kappa interpretation was proposed: below 0.00 – poor, 0.00 – 0.20 – slight, 0.21 – 0.40 – fair, 0.41 – 0.60 – moderate, 0.61 – 0.80 – substantial, 0.81 – 1.00 – almost perfect.

6 Results

In the first series of experiments (figures 5, 6) neural networks have been separately built and trained to recognize and classify individual classes, such as forests of some type, buildings, etc., and then verified by processing a separate map, and comparing the results to a reference-classified map. Various quantitative measures, such as described above, have been applied to evaluate the quality of the results. Every care has been taken to make sure, that the testing data were built in the same way as the data used to train the neural network.

While the above experiments and the results obtained can be objectively evaluated, they do not make it possible to evaluate the interactions between classes, which come into play when a complete classification map is constructed from them.

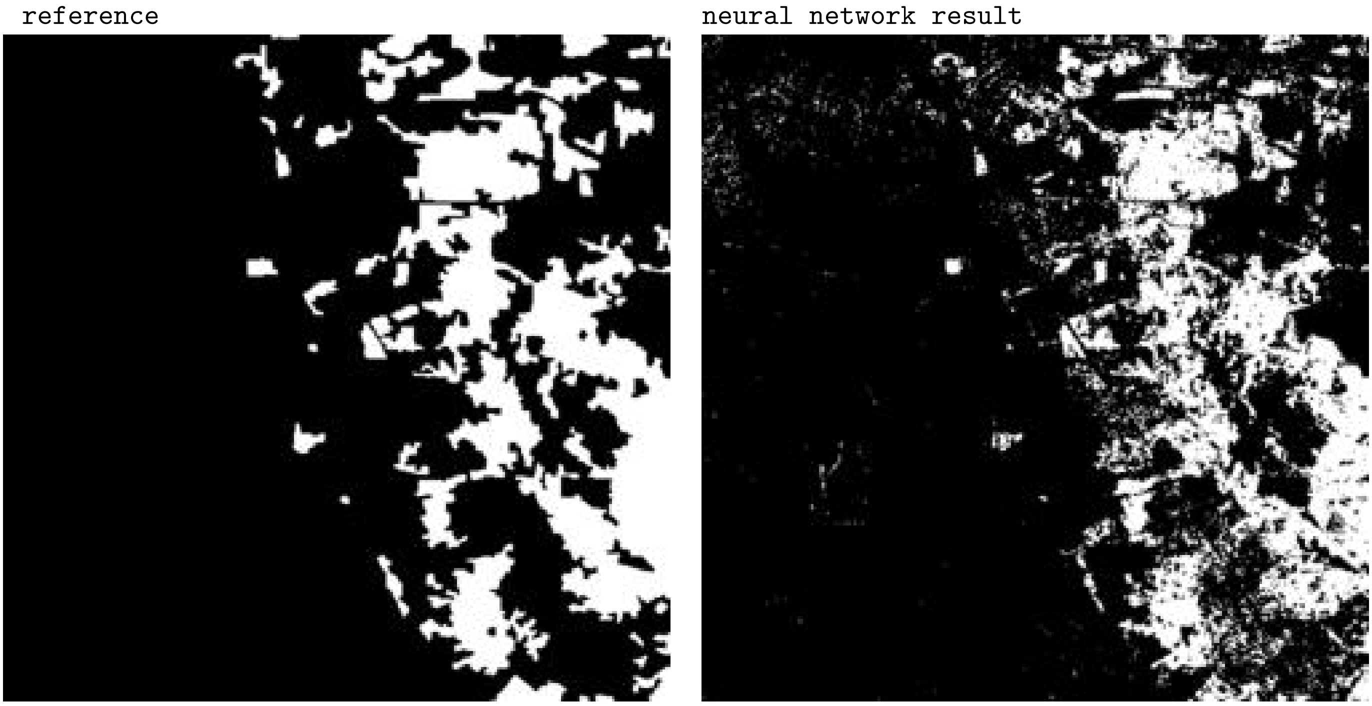

Figure 5: Coniferous forests: the reference map and the image obtained with a neural network. The interrater agreement coefficient Kappa for this example is 0.945 (see text).

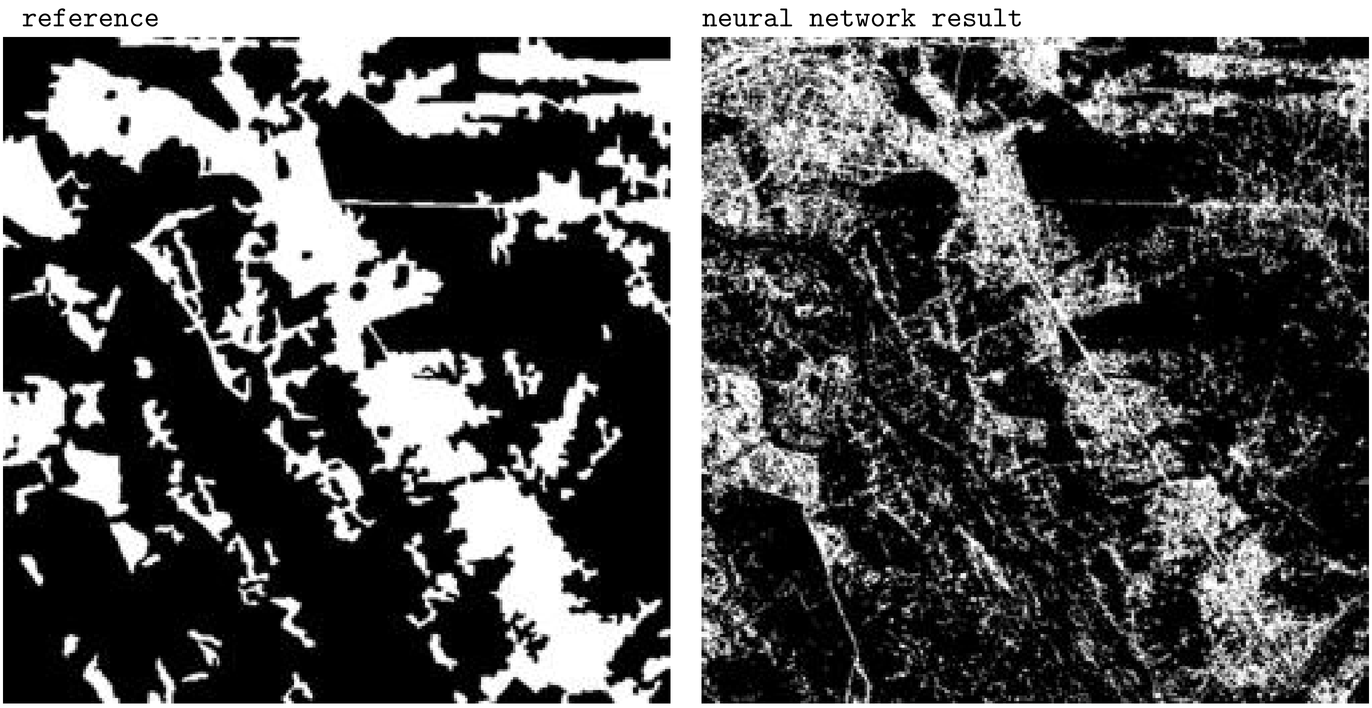

Figure 6: Urban areas: the reference map and the image obtained with a neural network. The interrater agreement coefficient Kappa for this example is 0.325 (see text).

A combined neural network can be built, and trained, to recognized all classes simultaneously, and produce colored maps similar to the original classification map, figure 7. Black pixels in this image represent pixels, which have not been determined to belong to any class, with respect to some threshold level. At the same time, a pixel can appear to belong to two or more classes, if it satisfies threshold levels for all these classes. Color mixing was used in this combined image to represent such pixels. It is much harder to achieve optimal classification this way, since many network parameters have to be optimized at the same time.

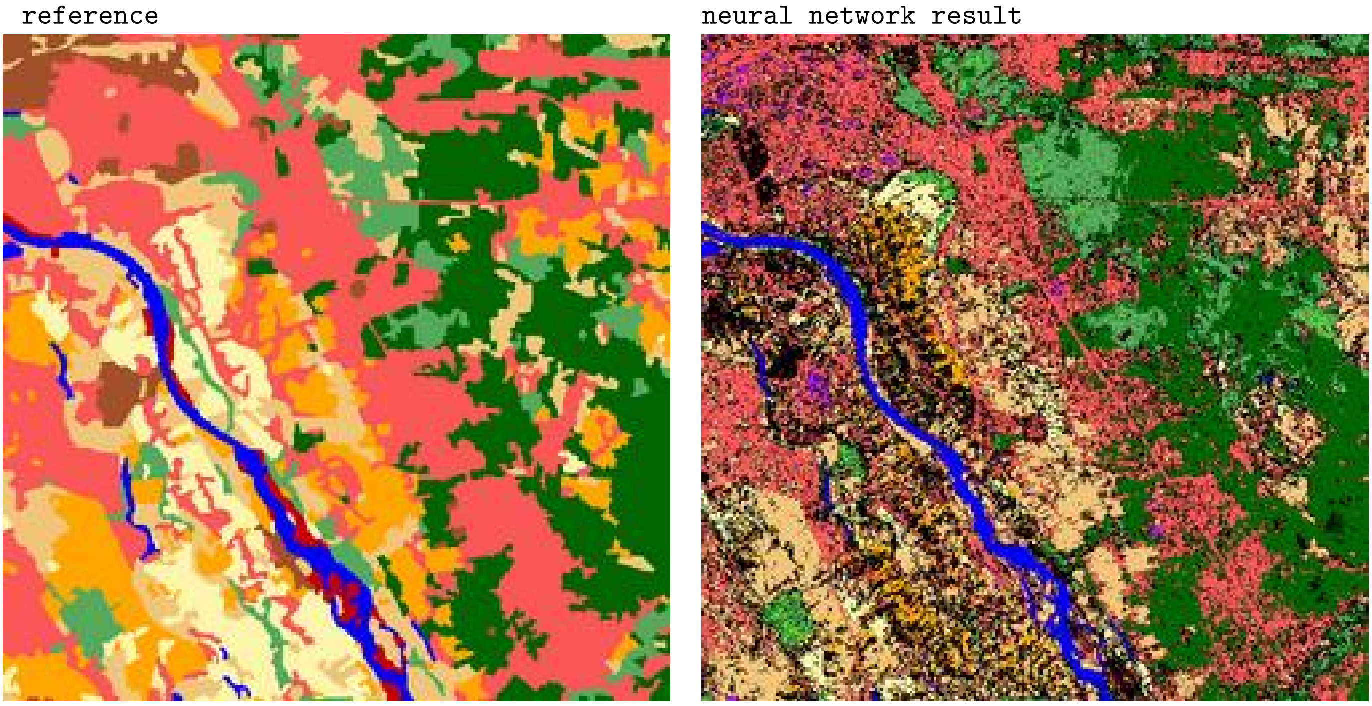

Figure 7: Complete image classification: the reference map and the image obtained with a neural network. Black areas are pixels not classified at the threshold level applied. Lowering threshold results in a more completely classified image, but some areas are doubly classified.

7 Conclusions

These experiments, and the results obtained, prove that satellite images can be processed automatically by supervised learning with a neural network. In some cases the results are very good, such as those in figure 5, and in some cases only barely acceptable (*fair*), figure 6. The explanation is that the neural networks in these experiments were trained for *per-pixel* classification, but on the basis of *per-parcel* classification results, against which the they were also evaluated. The training patterns included classes with a high differentiation of spectral response at pixels level. At the same time, classes' spectral responses overlapped in some parts.

Nevertheless, the neural networks proposed can serve as *per-pixel* classifiers. The results of classification can be further improved in the process of *per-parcel* classification, where the information about the pixel's neighborhood can aid in proper identification of its true nature. The authors are also working on some such techniques, like textures, to improve the results in these most difficult areas.

References

[1] Maselli Fabio (2001), *Remote Sensing of Environment*, 75(1):29–38

[2] German G. W. H., Gahegan M. N. (1996), *Computers and Geosciences*, 22(9):969–979

[3] Liu Xue-Hua, Skidmore A.K., Van Oosten H. (2002), *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(4):257–268

[4] Ju Junchang, Kolaczyk Eric D., Gopal Sucharita (2003), *Remote Sensing of Environment*, 84(4):550–560

[5] Foody Giles M. (2000), *Computers and Geosciences*, 26(4):469–478

[6] Keuchel Jens, Naumann Simone, Heiler Matthias, et al. (2003), *Remote Sensing of Environment*, 86(4):530–541

[7] Stuckens J., Coppin P.R., Bauer M.E. (2000), *Remote Sensing of Environment*, 71(3):282–296

[8] Atkinson Peter M., Sargent Isabel M.J. (2002), Annexe 5: High Resolution Sensor data for Automatic Land Use Classification. In Holland David, Guilford Bob, Murray Keith, editors, *OEEPE-Project on Topographic Mapping from High Resolution Space Sensors*, volume Official Publication No 44, pp. 91–104 Bundesamt für Kartographie und Geodäsie, Frankfurt am Main

[9] Zumsprekel Heiko, Prinz Torsten (2000), *Computers and Geosciences*, 26(1):87–100

[10] Carr James R. (1996), *Computers and Geosciences*, 22(8):849–865

[11] Tuceryan M., Anil K.J. (1998), *The handbook of Pattern recognition and computer vision*, chapter 2.1 Texture Analysis, pp. 207–248 World Scientific Publishing Co., 2nd edition

[12] Berberoglu S., Lloyd C.D., Atkinson P.M., et al. (2000), *Computers and Geosciences*, 26(4):385–396

[13] Cheng Qiuming (1999), *Computers and Geosciences*, 25(9):949–961

[14] Chica-Olmo M., Abarca-Hernndez F. (2000), *Computers and Geosciences*, 26(4):373–383

[15] Barr S., Barnsley M.J. (1999), A syntactic pattern recognition paradigm for the derrivation of second-order thematic information from remotely sensed images. In Atkinson P.M., Tate N.J., editors, *Advances in Remote Sensingand GIS Analysis*, pp. 167–184 Chichester: WIley

[16] Aplin P., Atkinson P. M. (2001), *International Journal of Remote Sensing*, 22(14):2853–2858

[17] Miller Diane M., Kaminsky Edit J., Rana Soraya (1995), *Computers and Geosciences*, 21(3):377–386

[18] Ruiz-del Solar Javier (1998), *Neurocomputing*, 21:7–18

[19] Lira Jorge, Maletti Gabriela (2002), *Computers and Geosciences*, 28(8):951–959

[20] Hubert-Moy L., Cotonnec A., Le Du L., et al. (2001), *Remote Sensing of Environment*, 75(2):174–187

[21] Foody Giles M. (1960), *Educational and Psychological Measurement*, 20:37–46

[22] Fleiss J. L. (1981), *Statistical Methods for Rates and Proportions*, volume 20 John Wiley & Sons, New York, second edition edition

[23] Landis J., Koch G. G. (1977), *Biometrics*, 33:159–174