REGION-BASED SATELLITE IMAGE CLASSIFICATION: METHOD AND VALIDATION

X. Gigandet, M. Bach Cuadra, A. Pointet, L. Cammoun, R. Caloz* and J.-Ph. Thiran*

Signal Processing Institute (ITS), http://lts5www.epfl.ch/
*Geographical Information Systems Laboratory (LaSIG), http://lasig.epfl.ch/

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

{Meritxell.Bach,Abram.Pointet,Leila.Cammoun,Regis.Caloz,JP.Thiran}@epfl.ch

ABSTRACT

We propose an algorithm for very high-resolution satellite image classification that combines non-supervised segmentation with a supervised classification. Both multi-spectral data and local spatial priors are used in the Gaussian Hidden Markov Random Field (GHMRF) model for the segmentation. Then, two classifiers, Mahalanobis distance classifier and SVM, are studied using intensity, texture and shape features. Validation is done qualitatively and quantitatively by comparison with a manual classification used as a ground truth. Results show very good performance of our approach in comparison to existing techniques. Also, we demonstrate that spectral and spatial features calculated on segmented regions are much more discriminant than the spectral features of the pixels taken individually for the classification task.

1. INTRODUCTION

The extraction of information on the land cover from remote sensing data has for long been driven by the use of the spectral dimension of the image alone. This approach, based on the consideration of the spectral distance and the decision criterion, proved to be satisfactory for the classification of medium and high resolution images (> 10 m resolution). Given the strong heterogeneity of the spectral information induced by the current very high resolution images (up to 0.5 m resolution), the pixel by pixel approaches of image classification are no more satisfactory.

Solutions are brought by the contextual approaches, considering the pixel neighborhood. The segmentation acts as an homogenization factor, strengthen the discriminating ability of the classification. Moreover, spatial features are calculated for each homogeneous region, supplementing the spectral features to be considered by the classifier.

Apart from rare cases, remote sensing dedicated studies start to introduce contextual methods for the thematic classification of images. Despite the advanced research status and the wide use of the contextual approaches in other image treatment domains (such as medical imaging [1]), the remote sensing domain remains less explored. Some early studies by [2] where followed by several contributions to segmentation of remote sensing data using clustering, region growing and edge detection methods [3]. Recently, Markov Random Fields (MRF) have an increasing interest in remote sensing [4, 5] since they are very well suited to model stochastic interactions among pixels. Computational solutions for the treatment of remote sensing data remains rare and far to provide the user with accurate and relevant classification results [6, 7].

A method for satellite image analysis combining several well-known algorithms is presented in this paper. First, both multi spectral data and contextual information are used in the Gaussian Hidden Markov Random Field (GHMRF) model for segmenting the image into homogeneous regions. Then, features such as mean intensity, texture or shape are extracted from the segmented areas in order to attribute a label to each region, using either a Mahalanobis distance or a Support Vector Machine classification. The first goal of this paper is to study the sensitivity of classification step to the prior segmentation. The second aim is to show the advantage of using region-based features from segmented images versus pixel-based features for high-resolution image classification.

The text is organized as follows. First, a brief description of the methods is done. Then, results and validation are presented followed by the discussion and conclusion.

2. METHODS

2.1. Segmentation

A Gaussian Hidden Markov Random Field (GHMRF) model is used for segmentation. Because of limited space, a summary of the main concepts is done, the reader is refered to [8, 9] for further details.

Intensity model. Let us suppose that the image intensity (each spectral band) has a density function $p(\boldsymbol{x}|\boldsymbol{\Theta})$ that is governed by the set of parameters $\boldsymbol{\Theta}$ in a way that:

$$p(\boldsymbol{x}|\boldsymbol{\Theta}) = \sum_{k=1}^{M}(\alpha_k p_k(\boldsymbol{x}|\theta_k)), \tag{1}$$

where M is the number of components, $\boldsymbol{\Theta} = (\alpha_1, ..., \alpha_M, \theta_1, ..., \theta_M)$ are such that $\sum_{k=1}^{M} \alpha_k = 1$ and each p_k is a probability density function characterized by θ_k[1]. Let $\mathbf{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ be a set of data drawn from this distribution. Let us suppose that there is a set of data $\mathbf{Y} = \{y_i\}_{i=1}^{N}$ whose values indicates which probability density function p_k generated each $\boldsymbol{x}_i$, where $y_i \in 1, ..., M$ for each i, and $y_i = k$ if the sample i is generated by the distribution k. If Gaussian distribution is considered, we note $\boldsymbol{\Theta}^g =$

[1] In the particular case of using Gaussian distributions, this is the well-known Finite Gaussian Mixture Model (FGMM).

0-7803-9134-9/05/$20.00 ©2005 IEEE

$(\alpha_1^g, ..., \alpha_M^g, \theta_1^g, ..., \theta_M^g)$, and $p_k(\boldsymbol{x}_i|\theta_k^g)$. The mixing parameters α_k are the prior probabilities of each mixture component. By using Bayes rule, we can compute *a posteriori* probability:

$$p(y_i|\boldsymbol{x}_i, \boldsymbol{\Theta}^g) = \frac{\alpha_{y_i}^g p_{y_i}(\boldsymbol{x}_i|\theta_{y_i}^g)}{p(\boldsymbol{x}_i|\boldsymbol{\Theta}^g)} = \frac{\alpha_{y_i}^g p_{y_i}(\boldsymbol{x}_i|\theta_{y_i}^g)}{\sum_{k=1}^M \alpha_k^g p_k(\boldsymbol{x}_i|\theta_k^g)}. \quad (2)$$

Spatial distribution model. The Hidden Markov Random Field (HMRF) theory asserts that the total image information can be reduced to a local information according to a neighborhood system. The local spatial information is taken into account by the GHMRF approach as prior probability, replacing α_k in Eq. 2 by

$$p(y_i|y_{N_i}) = \frac{e^{\beta V_{i,y_i}}}{\sum_{k=1}^M e^{\beta V_{i,k}}}. \quad (3)$$

y_{N_i} is the neighborhood around i, (8 direct neighbors are considered [5]), $V_{i,k}$ is called energy function, and β is a constant which represents the importance of the spatial prior. β is chosen experimentally, knowing that a small value will provide noisy segmentations, while a large one could damage the geometry of the objects. Usually, optimum value is in between 0.5 and 1.5.

Initial settings and implementation. The number of classes is usually fixed by the user. However, a *normalized entropy criterion (NEC)* [8] is useful in this choice. Each class is modelled by a Gaussian distribution whose initial parameters, θ^g, are obtained using a k-means algorithm. Initial segmentation is obtained through a thresholding of the image histogram[2] (step 1.1 in Table 1). GHMRF implementation follows an iterative scheme that solves the estimation parameter problem in an adapted version of the EM algorithm as suggested in [10]. The algorithm stops when the relative difference between estimated means is less than 0.1%. Segmentation is done by maximizing the *a posteriori* probability (MAP):

$$p(y_i|\boldsymbol{x}_i, \boldsymbol{\Theta}^g) = \frac{p(y_i|y_{N_i}) p_{y_i}(\boldsymbol{x}_i|\theta_{y_i}^g)}{\sum_{k=1}^M p(y_i|y_{N_i}) p_k(\boldsymbol{x}_i|\theta_k^g)}. \quad (4)$$

2.2. Classification

Classification is done in three steps: feature extraction, feature selection and labelling (step 2 in Table 1). They are briefly described in what follows.

Feature selection. Several features are extracted from the segmented areas: the mean and standard deviation for each channel, the red-infrared and red-green ratios, the area, perimeter and compactness. Feature selection is a critical step to ensure the best results. One way is to let the expert selecting the features. However, we suggest two approaches that could help in this choice, the *cross-validation* [11] and the *sequential generation* [12].

Mahalanobis distance classifier. This is a minimal distance classifier that tries to minimize, in the principal component space, the distance between each testing instance and the center of its class. Each class, defined by a set of training samples, is assumed to follow a Gaussian distribution.

[2]A FGMM is used to fit the image histogram. Then, initial segmentation is obtained by thresholding.

Classification process
1 Segmentation step
1.1 Initialization: k-means, FGMM
1.2 GHMRF
2 Classification step
2.1 Region-based Feature Extraction
2.2 Feature selection
2.3 Mahalanobis or SVM

Table 1. Segmentation and classification steps.

Training set				
Classifier	1	2	3	4
Mahalanobis	79.4	84.3	80.2	74.3
SVM	81.5	86.2	86.3	81.5
	5	6	7	8
Mahalanobis	85.7	79.3	81.7	85.3
SVM	86.6	83.3	83.5	80.8

Table 2. Classification process: overall accuracy [%]

Support Vector Machine (SVM). The SVM classifier is widely used in supervised classification. Its goal is to transform input data into a higher-dimensional space where classes can be better separated. Here, radial basis functions are used as kernel functions and the kernel parameters are experimentally chosen. We refer the reader to [8, 13] for further details.

3. RESULTS

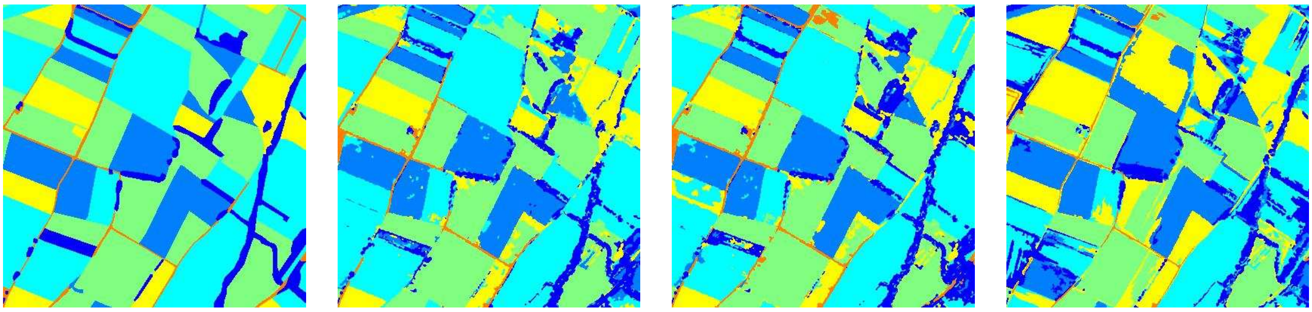
The segmentation and classification methods presented above have been extensively tested an validated on different data sets [8]. Here only the most significant results will be presented.

3.1. Results

Original image is from the Quickbird satellite (blue, green, red and infrared channels are used, image size is 401x401 pixels with 2.4m resolution) showing mostly rural areas. Fig. 1(a) shows a manual segmentation that will be considered as a ground truth classification image. Six classes have been selected by the expert: wood, farm 1, farm 2, open land, scarce vegetation and roads.

First, qualitative assessment of the results of our classification approach is done by visual inspection (see Fig. 1(b) and (c)). Results are very satisfactory since few regions are lost comparing to the reference. The three main factors of error are due to the initial segmentation errors, the noise and the shadows. Then, in Table 2, quantitative validation is shown by the overall accuracy (percentage of pixels correctly classified) using 8 different training sets manually selecterd (that cover from 22% to 50% of the whole image size). Globally, SVM provides slightly better results than Mahalanobis and both are robust among different training sets. Performance for each class is shown by the Dice Similarity Measure (DSM) in Table 3. DSM is an index of relative similarity, defined as $DSM_{a,b}^c = \frac{2N_{a\cap b}}{N_a+N_b}$, sensitive to both differences in size and location[3]. Results are shown for the Mahalanobis dis-

[3]N_x is the number of pixels that method x classifies as class c. $N_{a\cap b}$ is the number of pixels that both methods a and b classify as c.



(a) Manual classification (b) Mahalanobis (84.3%) (c) SVM classifier (86.2%) (d) E-cognition (66.8%)

Fig. 1. Reference and classified images (overall accuracy)

DSM [%]					
1	2	3	4	5	6
68.1	85.6	86.9	92.1	70.2	53.6

Table 3. Dice Similarity Measure (DSM), Mahalanobis distance classification. Wood, farm 1, farm 2, open land, scarce vegetation and roads

tance classifier using the training set 2 (overall accuracy of 84.3%). DSM shows that wood, roads and scarce vegetation are the classes with lowest rate accuracy (53, 68, and 70% respectively) mainly because they are not widely represented in the test image.

As far as we know, only the E-cognition software [14] performs a previous segmentation (using watershed algorithm) step for classification. Comparing with our approach, its resulting classification (see Fig.1(d)) is much more noisy and the overall accuracy is significantly lower (66.8%). If only the classification step is considered (E-cognition segmented image is used as input of our Mahalanobis classifier), an accuracy higher than 75% is obtained. The main limitations of their approach are the following. First, a single threshold parameter can be used in the watershed approach to tune the homogeneity of the segmentation. This threshold parameter acts simultaneously on both spectral and spatial dimensions. Therefore all the segments will have a similar size while this does not reflect the territory reality (having both small and large objects). Actually, GHMRF has better performance than watershed approach, as proven in [8]. Second, no automatic and class specific method is provided for the selection of features. Therefore, a common use is to integrate all the features in the classification and this may interfere with an optimal classification due to the use of non-discriminant features.

4. DISCUSSION

4.1. Sensitivity of the classifier to prior segmentation

Since feature extraction is done on the segmented regions, the selection of the number of classes, K, becomes a critical choice. If few classes are used, image geometry is not respected. On the other hand, if K is too large, regions are over-segmented and the extracted features are not significant. We then have to find the best compromise between segmentation and classification. Our aim is to determine which segmentation gives the best classification. Accuracy using different initial segmentations is shown

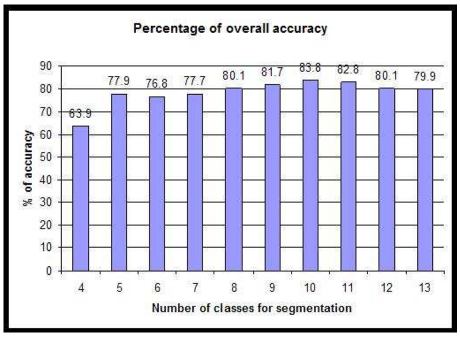


Fig. 2. Influence of the prior segmentation

(for Mahalanobis classifier) in Fig. 2. Note that the initial segmentations have slightly different training regions. Few number of classes, $K = \{2, 3\}$, do not allow to perform a classification since no relevant training regions can be selected. From $K = 4$ to $K = 10$ the accuracy progressively increases, mainly because of the improvement of the segmentations. Between $K = 9$ to $K = 11$, the best classification results are obtained. Then, performance slightly decreases from $K = 12$. This is not surprising because at this point images are over-segmented. Obviously, selecting an optimum number of components is not straightforward. However, maximum accuracy is clearly obtained between $K = 9$ and $K = 11$. Thus, a range of optimum values can be first determined visually by expert criteria and, finally, the optimum can be automatically chosen using the NEC criterion.

4.2. Pixel-based vs. region-based classification

The aim of this section is twofold: first, to compare pixel-based classification vs our region-based classification and second, to compare the effectiveness of different features. The Mahalanobis classifier is used in this section because of its best compromise between performance and computing time. In pixel-based classification texture is computed as the standard deviation of the intensity for each band in a 3 by 3 neighborhood. All the results are reported in Table 4. It is clearly shown that region-based segmentation performs around 8% better in terms of accuracy. Moreover, region-based classification is much less noisy thanks to the HMRF model that create homogenous areas. Finally, training is

	Overall accuracy [%]	
Features	Pixel-based	Region-based
Intensity	72.5	80.0
Int.+Tex.	77.9	84.2
Int.+Shape	-	82.5
Int.+Tex.+Shape	-	84.3

Table 4. Pixel-based vs. region-based classification

much easier when using region-based classification, since selecting regions is nicer than manually drawing areas to train the classifier. Texture features improve the results of more than 4%, in both pixel and region-based classifiers. Despite the improvement of the overall accuracy we do not suggest to use texture features in pixel-based classification because of the border effect. On the contrary, the use of texture is strongly advised for region-based classification. Shape features increase around a 2% the classification using only intensity information. However, differences between intensity-texture and intensity-texture-shape classification are not significant. Shape features do not seem to be useful in this case but we though suggest to extract shape features, even if the results are not improved in some cases, since they are naturally useful for some classes such as roads or high-ways.

4.3. Limitations and future work

As seen in the qualitative validation, poor prior segmentation, shadows and noise are the main factors classification errors. The influence of the prior segmentation quality has already been discussed in Sec. 4.1. Shadows and noise (such as speckle) are not still considered in our image model. Special attention should be done to these artifacts. Shadows are particularly difficult to deal with. In some cases, even if they can be detected is not easy to determine which is the underlying class. This is currently solved by arbitrarily splitting the shadowed area into the neighboring classes.

In a near future, a multi-level classification approach will be considered. As it is known, classes do not have the same spectral and geometrical characteristics. For instance, compactness is relevant for urban areas, since these regions are very elongated. On the contrary, farming areas do not present singularities in shape features. So, this new approach could discriminate at each level one class from the remaining ones, and the regions belonging to this class would be excluded in the next level. Then, the same principle would be applied to the other regions, until all the classes would have been assigned. Only relevant features would be used at each level, so that each class would be constructed with its best features.

5. CONCLUSION

We presented an algorithm for high-resolution satellite image classification that combines non-supervised segmentation with two different approaches for supervised classification. Segmentation uses both multi spectral data and contextual information. Then, features such as mean intensity, texture or shape are extracted to attribute a label to each region, using either a Mahalanobis distance or a Support Vector Machine classification. Results are very satisfactory: high accuracy rates have been obtained and the advantage of using a prior segmentation for feature extraction and training have been shown. Also, our approach has demonstrated its higher performance compared to softwares currently used. Finally, the feature selection problem has been discussed for intensity, texture and shape features.

6. ACKNOWLEDGEMENT

We would like to thank Dr. Vlad Popovici for his helpful discussions regarding the classification algorithms.

7. REFERENCES

[1] M. Bach Cuadra, B. Platel, E. Solanas, T. Butz, and J. Thiran, “Validation of tissue modelization and classification techniques in t1-weighted mr brain images,” *MICCAI 2002*, pp. 290–297, November 2002.

[2] R.L. Kettig and D. A. Landgrebe, “Classsification of multispectral image data by extraction and classification of homogeneous objects,” *IEEE Transactions on Geoscience Electronics*, vol. GE-14, no. 1, pp. 19–26, 1976.

[3] A.Guarnieri and A. Vettore, “Automated techniques for satellite image segmentation,” *Symposium on Geospatial Theory, Processing and Applications*, vol. WG IV /7, 2002.

[4] D.E. Melas and S.P. Wilson, “Double markov random field and bayesian image segmentation,” *IEEE Transactions on Image Processing*, vol. 50, no. 2, pp. 357–365, 2002.

[5] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, “Unsupervised classification of radar images using hidden markov chains and hidden markov random fields,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 675–686, 2003.

[6] G. Meinel and M. Neubert, “A comparison of segmentation programs for high resolution remote sensing data,” *International Archives of the ISPRS*, vol. XXXV, pp. 1097–1105, 2004.

[7] R. Caloz and A. Pointet, “Contextual classification of very high resolution images: Application on aerial photograph,” *Teledetection*, vol. 3, no. 2-3-4, pp. 311–322, 2002.

[8] X. Gigandet, “Satellite image segmentation and classification,” Tech. Rep., Swiss Federal Institute of Technology (Lausanne), 2005, http://ltswww.epfl.ch/˜bach/Gigandet2005.pdf.

[9] M. Bach Cuadra, *Atlas-based segmentation and classification of magnetic resonance brain images*, Ph.D. thesis, Swiss Federal Institute of Technology (Lausanne), November 2003.

[10] J. Zhang and W. Modestino, “Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation,” *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 404–420, July 1994.

[11] C. J. Lin C. W. Hsu, C. C. Chang, “A practical guide to support vector classification,” 2003.

[12] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer, 1998.

[13] P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, “Evaluation of factors affectiong support vector machines for hyperspectral classification,” *The American Society for Photogrammetry & Remote Sensing ASPRS*, 2004.

[14] “http://www.definiens-imaging.com/,” .