Combining Extended Dependency Tree –HMM based Recognition and Unsupervised Segmentation for Land Cover Mapping in Aerial Images

Mohamed El Yazid Boudaren, Abdel Belaïd

Combining Extended Dependency Tree –HMM based Recognition and Unsupervised Segmentation for Land Cover Mapping in Aerial Images

Mohamed El Yazid Boudaren, and Abdel Belaïd

***Abstract*— An important challenge to any image pixels classification system is to correctly assign each pixel to its proper class without blurring edges delimiting neighboring regions. In this paper, we present an aerial image mapping approach that advantageously combines unsupervised segmentation with a supervised Markov model based recognition. The originality of the proposed system carries on three concepts: the introduction of an auto-adaptive circular-like window size while applying our stochastic classification to preserve region edges, the extension of the Dependency Tree –HMM to permit the computation of likelihood probability on windows of different shapes and sizes and a mechanism that checks the coherence of the indexing by integrating both segmentations results: from unsupervised over segmentation, regions are assigned to the predominating class with a focus on inner region pixels. To validate our approach, we achieved experiments on real world high resolution aerial images. The obtained results outperform those obtained by supervised classification alone.**

***Index Terms*— Aerial image indexing, Auto-adaptive window size, Combining supervised and unsupervised segmentations, Dependency Tree –HMM, Land cover mapping.**

I. Introduction

Land Cover Mapping (**LCM**) in high resolution aerial images is an important application of remote sensed data. It consists of identifying the natural objects present in a high resolution aerial image given a set of known patterns. In the most general case of aerial images, when the image contains several regions of different patterns, the aim is to label each pixel with the corresponding texture. Evidently, the labeling process subsumes image segmentation but besides segmenting the image to different regions, it assigns each region to one of the natural objects patterns.

Achieving the classification at pixel level is a big issue in LCM problem. In fact, it is easier to identify an image of a relatively big size than identifying a lonely pixel.

Manuscript received March 23, 2009.

This work was supported in part by the Military Polytechnic School, Algiers, Algeria.

Mohamed El Yazid Boudaren is with the Applied Mathematics Laboratory of the Military Polytechnic School, Algiers 16111, Algeria. (e-mail: boudaren@gmail.com).

Abdel Belaïd is a full Professor at Nancy II University, France and the head of READ team of LORIA research Center, Nancy, France (e-mail: abelaid@loria.fr).

In fact, pixel-wise approaches for image classification are not usually suitable to solve problems often found in remote sensing application [1] [2]. They result in a disgusting salt and pepper effect.

Recent researches clearly show the advantages of integrating spatial dimension to spectral features by using segmentation based classification methods and, hence, focusing into image regions instead of pixels [3][4].

More elaborated approaches use a family of Markov models to model the contextual interactions between labels. However, genuine 2D-Markov modeling of the contextual information is a time consuming iterative process [5].

On the other hand, reasonable complexity approaches identify each pixel by taking into account its neighboring pixels, usually by computing a similarity measure (likelihood probability for instance) on square windows centered at concerned pixels [6]. The drawbacks of such approaches are the followings:

- They adopt square windows which may introduce a bias toward rectangular regions. Moreover, corner pixels are more distant than other pixels. Adopting non-square windows is usually unaffordable due to the used model or measure nature.
- The bigger is the window, the more likely the identification is correct. However, adopting a too big window may penalize small regions. A tradeoff is generally made.
- Since a static window size is adopted, the window size is then too small to perform efficient classification for all image pixels and too high to preserve edges since the classification of frontier pixels are biased through introduction of neighborhood pixels.

In this paper, we propose a system that overcomes the previous difficulties by introducing the following:

- Segmentation is achieved through unsupervised segmentation which preserve region edges even if it provides an over-segmented image.
- Each region is identified through stochastic supervised classification.
- Likelihood probability may be computed on windows of different sizes and shapes centered at considered pixels.
- To determine window size and shape, an auto-adaptive distance is computed based on the considered pixel position towards region edges.
- To permit likelihood probability computation on

non-rectangular windows, we extended the Dependency Tree- hidden Markov model (DT-HMM) by allowing four directional dependencies instead of two, and adopting the central pixel as root instead of upper-left pixel when dealing with rectangular windows.

The reminder of the paper is organized as follows: in Section II, we introduce our Extended Dependency Tree -HMM (EDT-HMM) that extends DT-HMM. Section III describes our indexing scheme. Section IV shows experimental results. Conclusion and future works are given in Section V.

II. Extended Dependency Tree –Hidden Markov Models

Markov models (Markov Random Fields, Hidden Markov Fields, Hidden Markov Models, Hidden Markov Trees…) were extensively and successfully used for texture modeling and segmentation [7]. This is majorly due to their ability to model contextual dependencies and noise absorption [8]. However, their performance depends widely on the model architecture: genuine 2D-models yield better results but exhibits much higher computational complexity [8]. In general, the more complex is the model, the better are the performances.

Nevertheless, for computational complexity reasons, several approaches consider linear models like HMM even if such a model is not suited for two-dimensional data [9]. More elaborated approaches resort to 2D-models with simplifying assumption. One simplifying assumption that provides good results with a linear complexity is that assumed in DT-HMM [10][11]: one site (image pixel) may depend on either the horizontal or vertical predecessor, but not on both the same time.

The extension of DT-HMM in this work is motivated by two reasons:

- The need to compute likelihood probability on non-rectangular shaped windows of different sizes.
- The need to adopt central pixel (to be labeled) as the dependency tree root since the root shows more interactions with neighbors than other pixels do.

A. EDT-HMM Overview

Before describing our model principles, let us define the applicability conditions of the EDT-HMM model on a window w with respect to root r.

The window w must fit the following condition:

- For each site s of w, s must have at least one neighbor $v \in N_s$ that belongs to w and fulfills: $\|v,r\| < \|s,r\|$ where N_s is the 4-neighborhood of s and $\|\ \|$ is the Euclidean distance.

Let w be a window verifying the condition above, and let r be the center of the window and $Y_r = \{y_s / s(i,j) \in w\}$ be the set of features vectors (RGB for instance) of pixels inside w. Y_r is then the observable process. Let X be the hidden process. The likelihood probability is given by:

$$P(Y_r/\lambda) = \sum_X P(Y_r/X,\lambda)P(X/\lambda) \qquad (1)$$

Unlike DT-HMM, where each pixel may have a predecessor chosen between two directions, in the EDT-HMM modeling, a pixel s may have a predecessor v chosen randomly from the 4-Neighborhood (up, down, right or left) and verifying the Euclidean distance property. Note that, the neighborhood directions of all pixels of w define a tree structure T like depicted in figure 1. We note $T(s) = v$.

The likelihood probability to observe Y_r given the parameters of the DT-HMM $\lambda(\pi, A, B)$ can be approximated as follows:

$$\begin{aligned} P(Y_r/\lambda) &\approx \sum_T P(Y_r/T,\lambda) \\ &\approx \sum_T \sum_X P(Y_r/X,\lambda)P(X/T,\lambda) \\ &\approx \sum_T \sum_X \left\{ \prod_{s\in w} P(y_s/x_s,\lambda)P(x_s/T,\lambda) \right\} \end{aligned} \qquad (2)$$

In this paper, we propose to evaluate the likelihood on a set of random dependency trees τ. The previous equation becomes:

$$\begin{aligned} P(Y_r/\lambda) &\approx \sum_{T\in\tau} P(Y_r/T,\lambda) \\ &\approx \sum_{T\in\tau} \sum_X P(Y_r/X,\lambda)P(X/T,\lambda) \\ &\approx \sum_{T\in\tau} \sum_X \left\{ \prod_{s\in w} P(y_s/x_s,\lambda)P(x_s/T,\lambda) \right\} \end{aligned} \qquad (3)$$

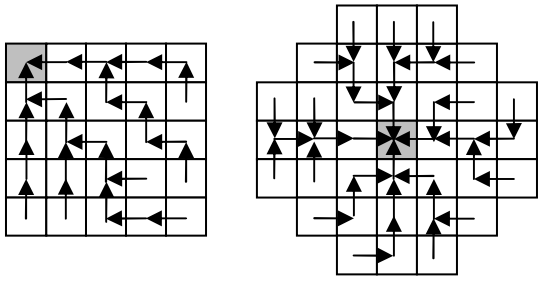


Figure 1: Examples of random dependency trees according to DT-HMM (left) and EDT-HMM (right) formalisms

Thereafter, we remind the definition of the Model parameters π, A and B.

$$b_i(O) = P(y_s = O / x_s = i) \qquad (4)$$

$$P(x_s = j / x_v = i, \lambda) = \begin{cases} \pi_j \text{ if } s = r \\ a_{ij} \text{ otherwise} \end{cases} \qquad (5)$$

where $i, j = 1, ..N$ represent hidden states.

Note that a_{ij} only depends on i and j and not on the direction (horizontal or vertical).

To compute the likelihood probability of equation 3, we define the backward function $\beta_i(s)$ representing the probability of observing the data contained in the sub-tree of T with s as a root starting from the hidden state i.

$$\beta_i(s) = \begin{cases} b_i(y_s) \text{ if } s \text{ is a leaf} \\ b_i(y_s) \prod_{T(v)=s} a_{ij}\beta_j(v) \text{ otherwise} \end{cases} \qquad (7)$$

Note that the likelihood probability of equation 3 can be

evaluated as follows for each dependency tree T :

$$p(Y_r/T,\lambda)=\sum_{i=1}^{N}\pi_i\beta_i(r) \qquad (8)$$

This computation exhibits a reasonable complexity (linear with window size).

The extension of the DT-HMM only concerns likelihood probability computation and Viterbi decoding whereas learning is performed the same way as in DT-HMM context.

The Viterbi decoding process can be achieved in a similar way to the likelihood probability computation.

In this work, we only resort to likelihood probability computation.

On the other hand, learning is performed via an iterative way the same as for the DT-HMM model, since the parameters are the same:

- Initialize model parameters.
- Choose a random dependency tree T as described above (respecting the Euclidean distance constraint).
- Perform learning as in a linear framework (like in 1D-HMM).

III. Classification Scheme

To produce a class map of a given high resolution aerial image, we follow the scheme depicted in figure 2. In the following paragraphs, we describe each step.

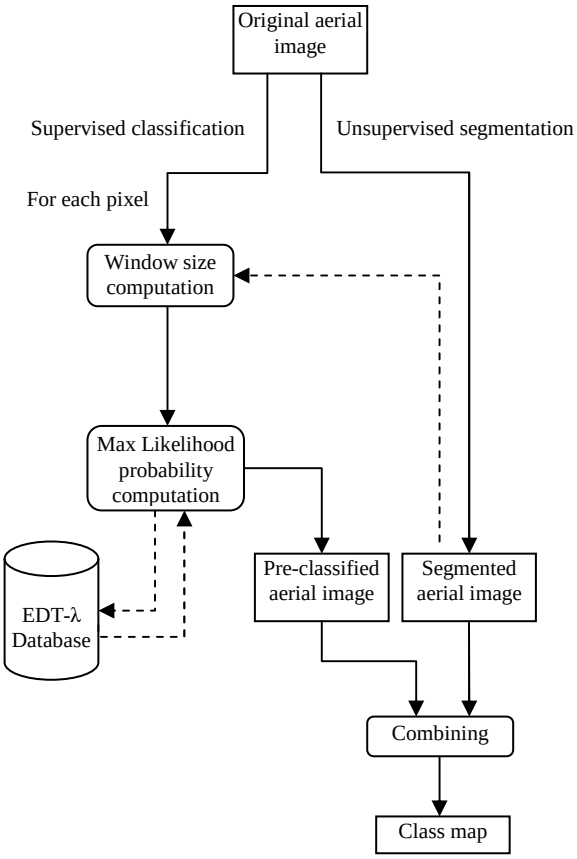


Figure 2: Classification scheme

A. Image unsupervised segmentation

Before classifying the image pixels, we need to perform an image unsupervised segmentation that fits the following conditions:

- Image edges are preserved.
- Pixels of the same region belong necessarily to the same natural object class, i.e. we may have an over-segmentation but not under-segmentation.

This step serves as a pre-processing one, and will guide the rest of steps of the classification process.

One unsupervised segmentation that has been shown to provide good results is the one produced by the EDISON system [12][13] which we use in this work.

A sample of a high resolution aerial image (50cm per pixel) and the corresponding unsupervised segmentation via the EDISON system are provided in figure 3.

Figure 3: A sample of high resolution aerial image [14] (left) and its corresponding unsupervised segmentation (right)

B. Window size computation

Since texture is not a local phenomenon, in order to classify a pixel s , we consider it with its neighborhood. More explicitly, we will compute the likelihood of the data inside a window centered at the pixel under consideration.

Let us denote w_s such a window and Y_s the data associated to that window. The class λ_s of the central pixel s is the class that maximizes the likelihood probability:

$$\lambda_s^{*}=\arg\max_{\lambda\in\Lambda} P(Y_s/\lambda) \qquad (9)$$

Most approaches adopt a square window of a fixed size for all pixels. A trade-off is usually made so that the window is enough big to correctly classify the central pixel and enough small to preserve the region edges.

In this work, we propose to dynamically compute the window size to allow our system to deal with a maximum amount of information without distorting region edges. The more the pixel to classify is far from the region boundary, the larger is the window whereas edge pixels are classified without considering their neighborhood.

Hence, window size is chosen so that pixels within the window belong to the same region according to an unsupervised over-segmentation of the image.

Window shape and size depend on a unique parameter Ray_s that represents the maximum Euclidean distance between neighbors and central pixel s . Figure 4 shows samples of windows of different shapes and sizes.

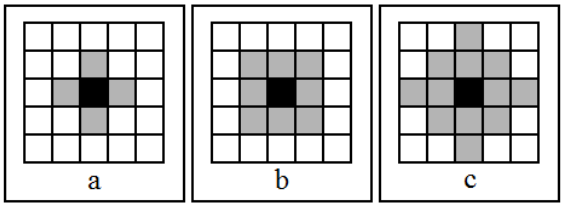
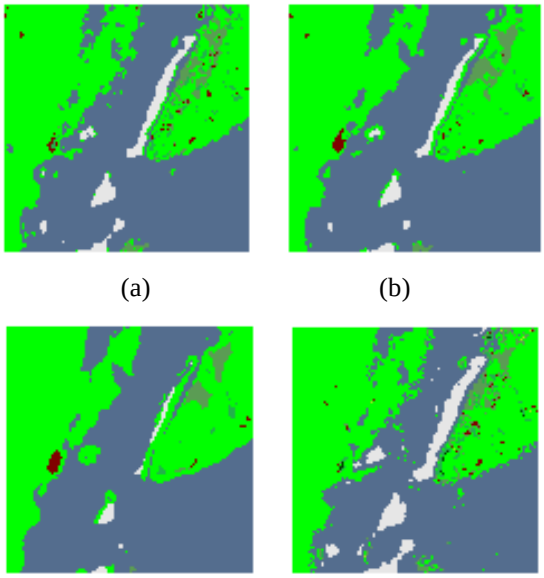


Figure 4: window shape and size for different values of Ray_s .

a- $Ray_s^2=1$, b- $Ray_s^2=2$, c- $Ray_s^2=4$.

This parameter is obtained from the pre-segmented image. Its value is the maximum value possible so that pixels within the window belong to the same region.

A comparative analysis of the window size impact on the accuracy of classification of the aerial image of figure 3 is shown in figure 5.



(a) (b)

(c) (d)

Figure 5: Impact of window size on the pre-classification accuracy

a- $Ray_s^2 = 1$, b- $Ray_s^2 = 4$, c- $Ray_s^2 = 10$, d- Auto-adaptive Ray_s.

C. *Image pre-classification*

To assign a pixel to a class, we compute the likelihood probability of observing the window centered on that pixel according to the EDT-HMM of each natural object class as described in the previous section. The pixel is then allocated to the class that maximizes this probability as shown in equation 9.

The parameters of the EDT-HMM corresponding to each natural object class are obtained after a learning process achieved on mono-class aerial images.

To represent each pixel, we used the classical RGB color space. To estimate the parameters of the DT-HMM of each class, we achieve K-Means clustering on pixels of mono-class image of the corresponding class to divide the image pixels on *N* sub-classes. Subsequently, we obtain the parameters of *N* Gaussian functions. These parameters serve as an initialization of our EDT-HMM. The final parameters of the model are then obtained after an iterative process as described in the previous section.

We remind that the observation probability density function for each state is given via a Gaussian function.

$$b_i(O) = \mathcal{N}_{\mu,\delta}(O) \qquad (10)$$

An example of image pre-classification via the EDT-HMM labeling is provided in figure 5-d.

D. *Classification Correction*

After the previous steps, the resulted class map suffers from the so called salt and pepper phenomenon. This is majorly to the difficulty to distinguish between several similar textures, especially for pixels near boundaries. In fact, such pixels are classified considering small-sized windows. To overcome this involvedness we propose to merge pixels of the same region (in the sense of the unsupervised segmentation) into the same natural object class with a focus on inner pixels of the region, since those pixels were classified considering larger windows.

Explicitly, each region *R* is assigned the natural class that fits the following rule:

$$X_R^* = \underset{X \in \Lambda}{\arg\max} \sum_{\lambda_s = X, s \in R} size(w_s) \qquad (11)$$

Figure 6 shows the result of labeling correction on pre-classified image from figure 5-d.

Figure 6: Image classification correction.
Original aerial image (left), class map (right)

IV. Experimentation

In this section, we show the results of our classification scheme on real world high resolution aerial images. We first give an overview of the data and the learning database. Then, we provide experimental results on test images.

A. *Data Overview*

For our experimentation, we consider real world aerial images with a resolution of 50cm per pixel. The images were provided by La Régie de Gestion des Données des pays de Savoie (RGD 73-74), France [14].

The pictures were taken in relatively good light conditions; however, some images suffer from presence of shadow in some parts.

Samples of the aerial images used in our experiments are shown in Figure 7.

Figure 7: Sample of aerial images from RGD 73-74[14].

B. *Learning Database*

Learning was performed on mono-class images. These images were carefully extracted from the aerial images of the same area of study.

Some samples of mono-class images used for learning process are provided in figure 8.

Figure 8: Samples of learning images of classes: Tree (left), Snow (middle) and Water (right).

Thereafter, we present the legends used to represent the classes.

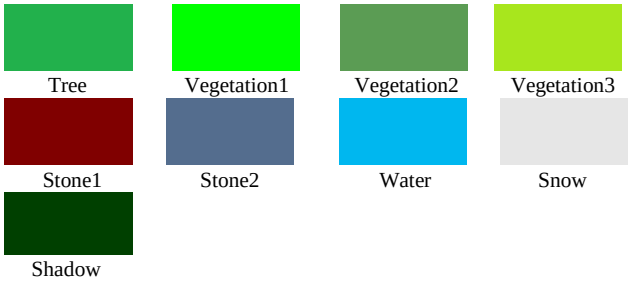


Figure 9: Natural object classes' legends

C. *Mono-Class Image Generation*

To demonstrate the capacity of the DT-HMM to represent natural object textures, we generate mono-class images using the corresponding DT-HMMs and compare them to generated images using classical 1D-HMM and GMM. Some generation results are given in figure 10.

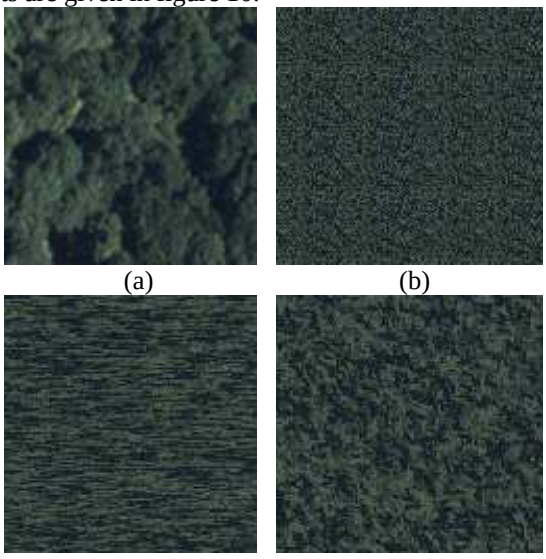


Figure 10: Mono-class image generation of class Tree using: b- GMM, c- 1D-HMM, d-DT-HMM.

D. *Experimental Results*

To evaluate the robustness of our aerial images pixels' classification system, we considered three types of test images:

- Mono-class images, for which the classifier is expected to assign all pixels to the corresponding class.
- Mosaic images, assembled by combination of different classes' images into regular boxes so that we can easily produce a corresponding ground truth map.
- Natural aerial images, for which we don't have a precise ground truth map. Thus, only a visual evaluation can be achieved in this case.

To produce the unsupervised segmentation of areal images, we acknowledge the use of EDISON system software [12][13]. Notice that experiments were performed on a large number of aerial images, and only some of these are shown in this paper.

Some experimental results on mono class images are given in Figure 11.

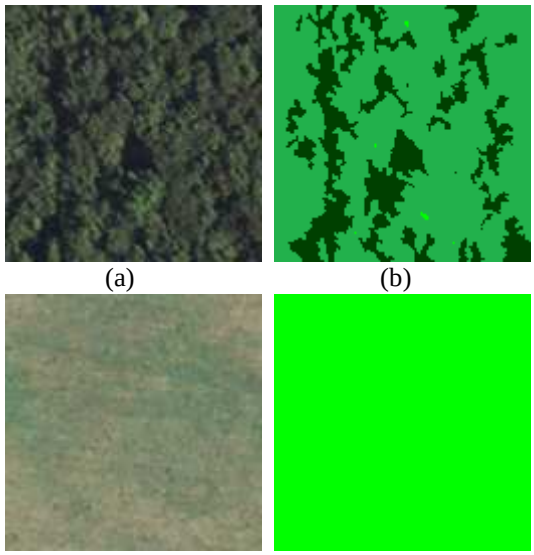


Figure 11: Mono-class image classification
Mono-class aerial images of classes Tree (a) and vegetation1 (c) and their corresponding class maps (b, d).

Experiments were then performed on mosaic images. These latter were assembled using different classes sub-images chosen randomly from the test database. Some results are presented in figure 12.

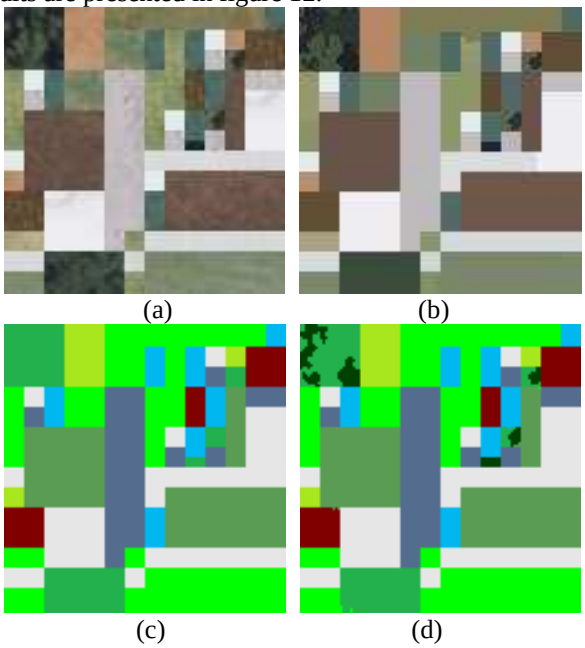


Image 12: Classification result on mosaic image
a- Original mosaic image, b- Pre-segmented image, c-Ground truth, d- Image class map.

The obtained classification of the mosaic image is very similar to the corresponding ground truth. Notice that the shadow class is not included in the ground truth map since we know to which class the shadowy pixels belong.

Finally, we give experiments results on natural aerial images.

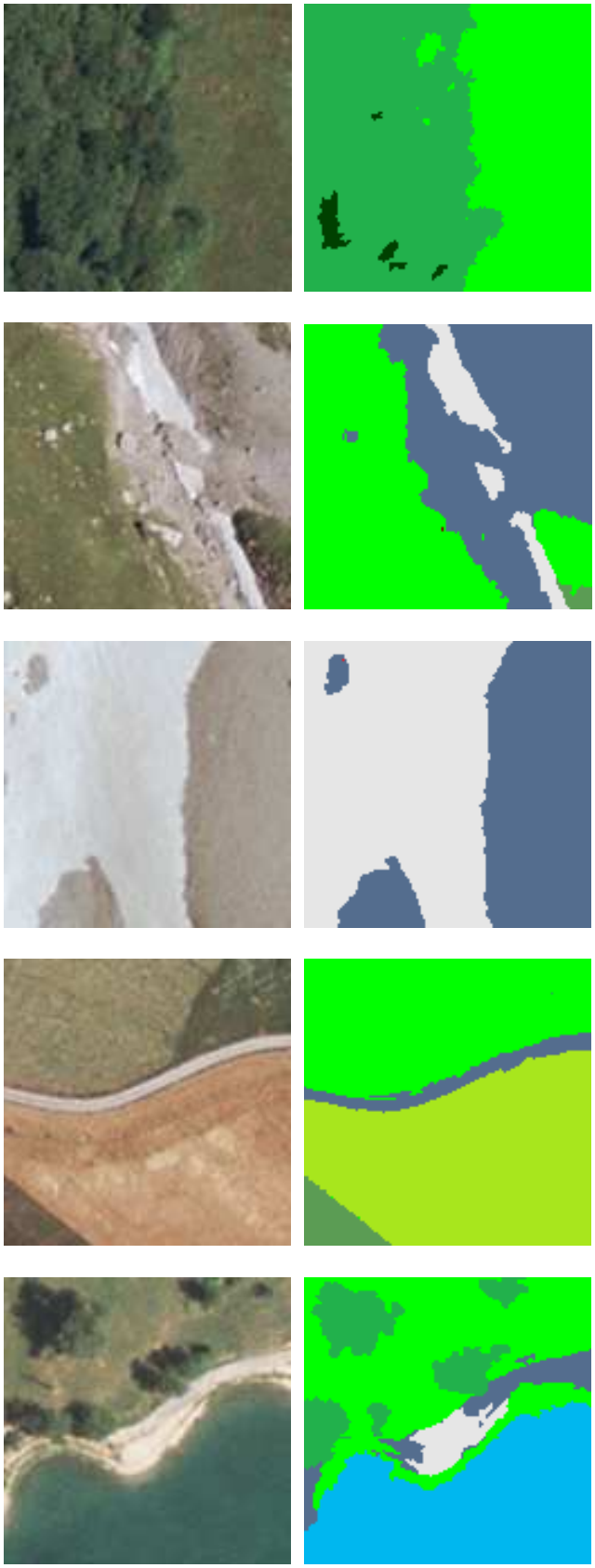


Figure 13: Classification results on real aerial images: Original images (left), the corresponding class maps (right)

V. Conclusion

In this paper, we proposed an approach that advantageously combines supervised EDT-HMM modeling and unsupervised segmentation to classify land cover pixels.

Instead of achieving our classification using a static window size, we resorted to an auto-adaptive window size depending on the position of pixel under consideration towards region boundaries.

Overall, the experimental results show that our system produces satisfactory class maps in a reasonable time given the linear complexity of the modeling. Note that several textures are so similar to each other that it is sometimes very difficult even for a human to distinguish between them.

As future work, we propose to apply EDT-HMM modeling to other kinds of problems.

References

[1] V. Mesev, *Remotely Sensed Cities*, Taylor and Francis, 2003.

[2] Thomas N., C., C. H. and Congalton, R.,"A comparison of urban mapping methods using high-resolution digital imagery". *Photogrammetric Engineering and Remote Sensing* 69(9), 2003, pp. 963–972.

[3] De Jong, S. M. and Freek, D. M., 2006. Remote Sensing ImageAnalysis: including the spatial domain. Springer.

[4] Jensen, J., 2006. Introductory Digital Image Processing. Prentice Hall.

[5] E. Levin, R. Pieraccini, "Dynamic planar warping for optical character recognition", *IEEE International Conference on Acoustics,Speech and Signal Processing*, Volume 3, pp.149 –152,March 1992.

[6] H. Permuter, J. Francos, I. H. Jermyn, "Gaussian mixture models for texture and colour for image database retrievel", *IEEE International Conference on Acoustics,Speech and Signal Processing*, Volume 3, pp.569 –572,2003.

[7] Hideki Noda, Mahdad N. Shirazi and Eiji Kawaguchi, "MRF based texture segmentation using wavelet decomposed images", *Pattern Recognition*, Volume 35, pp. 771-782, 2002.

[8] Wojciech Pieczynski, "Markov models in image processing", Traitement de Signal, Volume 20 N°3, pp.255-277, 2003.

[9] Mohamed El Yazid Boudaren, Abdenour Labed, Adel Aziz Boulfekhar and Yacine Amara, *Hidden Markov model based classification of natural objects in aerial pictures*, Proceeding of IAENG International Conference on Signal and Image Engineering, London, July 2-4, 2008.

[10] B. Merialdo, *Dependency Tree Hidden Markov Models*, Research Report RR-05-128, Institut Eurecom, 2005.

[11] B. Merialdo, J. Jiten, E. Galmar, B. Huet, *A new approach to probabilistic image modeling with multidimensional hidden Markov models*, Adaptive multimedia retrieval 2006, pp. 95-107.

[12] D. Comaniciu, P. Meer, "Mean shift: A robust approach towards feature space analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, NO. 5, May 2002.

[13] P. Meer, B. Georgescu, "Edge detection with embedded confidence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, NO. 12, December 2001.

[14] RGD73-74: Régie de Gestion des Données des Deux Savoies. http://www.rgd73-74.fr.