A new approach to natural object labelling in colour aerial images

J. Martí, J. Freixenet, X. Lladó, X. Muñoz, and X. Cufí
Computer Vision and Robotics Group - IIiA
University of Girona. 17071 Girona - Catalonia - Spain
{joanm,jordif,llado,xmunoz,xcuf}@eia.udg.es

A new approach to natural object labelling in colour aerial images

Abstract

This paper presents a new approach to aerial scene description, by providing a methodology for labelling pixels of natural image objects in a top-down way. The approach is based on an initial learning process which permits to generate flexible models that can fit the natural objects of aerial scenes. The proposal of the paper is to present the mechanism for interactively collecting object models that is designed as a supervised learning task, useful for objects that vary significantly under different environment conditions. A teacher provides meaningful examples of the objects of interest in training images, and the learning system selects the most interesting features for each object of interest. The proposed method constitutes a fast, simple and easy model acquisition process for pixel-labeling systems. The results of a series of tests carried out under different weather and seasonal conditions demonstrate the feasibility of the approach.

Keywords: Aerial scene description; natural object recognition; knowledge acquisition; pixel classification; supervised learning

1 Introduction

A scene description system can also be considered as a knowledge-based vision system, because such a system requires models that represent prototype objects and scenes. Any information concerning the specific characteristics of an object is contained in the object model database, while a list of objects that are expected to be in an image, and their relationships, are contained in the scene model databases.

Outdoor scene description systems entail many problems due to the scenes variability. Until now, most existing systems have not taken into account such variations, and are restricted to the descriptions of some predetermined images in specific outdoor conditions only (e.g., sunny images without shadows, only green trees, etc.).

On the other hand, many knowledge-based vision systems built their databases from scratch. Programmers are left to choose vision procedures for object recognition based only on their own intuition, and often incorporating model knowledge implies a difficult and tedious task because up until now, most of the existing knowledge-based systems had to incorporate this knowledge by hand; and all the more so for code-encapsulated data. Examples are the VISION/Schema system [4], EYE [3], and the region analyser of Ohta et al. [7] among others, which are successful demonstration systems, and compulsory works of reference. A major drawback of these systems is the significant effort needed relating to the development of the model knowledge base.

With the aim to provide some improvements in such knowledge engineering tasks, our current work focuses on detailing a knowledge acquisition method designed as a supervised learning task. Concretely, the presented method has been designed to be applied to images corresponding to outdoors, and consequently takes into account the variability of object appearance in such scenes. The method constitutes a fast, simple and easy model acquisition process for pixel-labeling systems.

2 Learning as a supervised knowledge acquisition task

Models constitute representations of the real world, and thus, modelling implies the choice of a suitable set of parameters in order to build those representations. Obviously, the selection of the parameters will affect how well the models fit reality, and this becomes a central issue in any modelling process. Due to the complexity of some scenes, in our approach we make use of generic modelling to characterize every object in the scene, but include the possibility that every single object can be described by specific features, in order to facilitate later recognition processes.

Object modelling has been designed as a supervised task, where a teacher presents representative examples of objects in training images. Each real object is modelled as a set of *object classes*. An *object class* is defined as a prototype of a real object described in terms of colour and textural features under specific outdoor conditions. A proper combination of such *classes* will produce a suitable description of the object under any outdoor conditions.

Basically, the interactive modelling process includes:

- A set of different images, corresponding to several outdoor conditions, is used to train the system. Three different attributes are used to define the outdoor conditions: (i) the season, (ii) the time of day, and (iii) the weather (similar to those considered by Bhanu et al. in [1]). With the aim of facilitating the learning task, these attributes can be initialized using linguistic values.

- Data acquisition. The teacher first selects meaningful regions representing *object classes* (i.e., real objects in specific outdoor conditions). Then, the learning application computes and registers a number of colour and texture features for each *object class*.

- Feature selection. The learning application selects a particular subset of features for characterizing each *object class* of interest.

- Learning verification. The learning application produces a visual feedback for the *classes* of interest.

2.1 Feature selection

The whole set of colour and texture features were not necessary for object characterization, so only a subset were used. Thus, the feature selection task is devoted to choosing the best set of colour and texture features that effectively characterize every *object class*. We have proposed two ways to carry out this task:

- an exhaustive search, which includes all the combinations of all the features, and then choosing the combination that achieves the lowest misclassification ratio. Here the processing time is not a real constraint, because the computational cost is very high and does not permit real interactivity. In the experiments we perform we use 30 features, so the exhaustive search does not suppose an impossible computationally cost, and it takes hours of processing time.

- a reduced combinatorial exploration, which means clipping the former combination process in an intelligent way, trying to achieve a real interactive *teacher-system* by means of reducing the search space. This reduction is performed by a genetic algorithm which combines random and theoretical behaviours (see [5] for details).

2.2 Feature evaluation

An optimal subset is always relative to a certain evaluation function. Typically, an evaluation function tries to measure the discriminating ability of a feature or a subset, to distinguish the different class labels [2]. Following the "divide and conquer" paradigm, we have solved the problem of feature evaluation by using a classifier based on a decision tree that separates the samples in a recursive way providing for each subset a misclassification ration (see [6] for details).

2.3 Verification

Once the process of feature comparison is completed, the set of features having the lowest misclassification ratio is selected. Then, the system allows the possibility to classify the pixels of the current image, providing immediate visual feedback to the teacher. This operation is performed by dropping each pixel of the image down the tree structure of the evaluation function, so that every pixel is classified as belonging or not to the classe of interest. The visual feedback guides the teacher, giving him or her the option to interfere in the process of feature selection, by introducing new training images, or by constraining the selection of some texture and colour parameters.

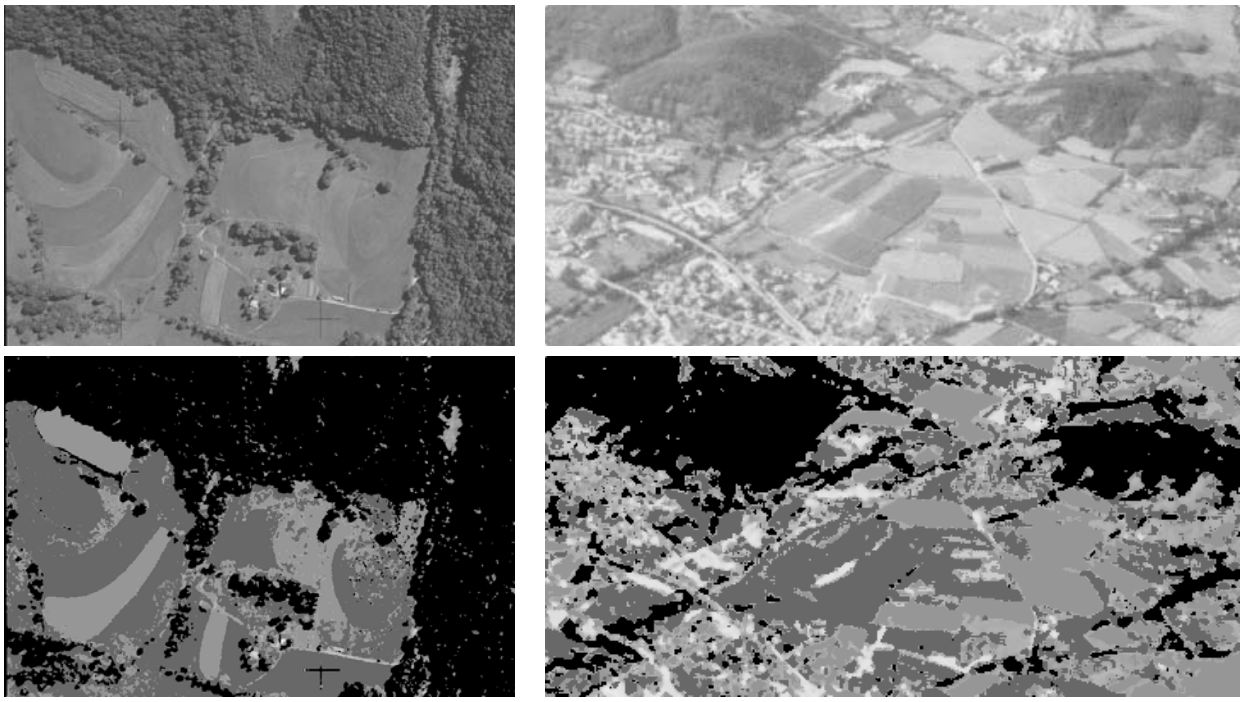
3 Performance

The reliability of our proposal is proven through a goal-directed segmentation that is performed by means of a classification function that decides whether each pixel of the image belongs to an object or not. The adopted classification function is the same as that used to show the visual feedback to the teacher during the object modelling process. As is described in section 2, each *class* representing an object under a specific outdoor condition has an associated decision tree that acts as a classification function. Conscious that not all outdoor conditions can be modelled for each one of the objects of interest, the system attempts to maximize outcomes from the available object knowledge. Furthermore, it is obvious that in the outdoor world, the appearance of some objects could differ widely from the modelled prototype. Therefore, in order to handle these problems, the system combines the available information of the object databases in a fuzzy way. Considering the three input parameters (the `season`, the `weather`, and the `time of day`) as fuzzy sets, each classification function has a degree of confidence, and a pixel is classified according to the following rule:

$$\sum_{j=1}^{m} \left(\frac{\alpha_j + \beta_j + \gamma_j}{3} \right) (CF_j) \geq 0$$

where α,β, and $\gamma \in [0, 1]$, represent the value of the input variables `season`, `weather`, and `time of day`, respectively. If the classification function of the *object class* j, classifies the pixel as belonging to the object, then CF_j is set to 1, otherwise it is set to -1. Finally, pixels classified as belonging to the *object class* j, are grouped into regions.

To verify system reliability, a set of experiments were performed. Recognition of objects such as the `tree-crown`, the `grass` and the `ground` was achieved with a high degree of success. Here, we report on some of the results in order to illustrate the system performance, describing test images which have been taken in different outdoor conditions to those in training. The training set includes only few images reproducing three different outdoor conditions (only three classes were considered). Figure 1 shows the segmentation results obtained in the description of two aerial images.



(a) (b)

Figure 1: Two original images and the corresponding results achieved in recognizing the objects `tree-crown`, `grass`, and `ground` (in black, iron-gray and pearl-gray respectively)

4 Conclusions

Most vision researchers agree that the success of knowledge-based vision systems falls on their ability in, and ease of use in *learning* and *growing*. This paper has presented a fast, simple and easy model acquisition process for vision systems. Furthermore,

the proposed features selection approach provides a powerful tool in such situations where it is not clear which features are to be used in classification tasks, as happens in most outdoor scenes under variable lighting conditions. The ability to select the most appropriate features attenuates the lack of a consolidated colour space for characterizing complex scenes, as it has been proven testing the algorithm on an outdoor scene description system.

References

[1] B. Bhanu, S. Lee, and J. Ming. Adaptative image segmentation using a genetic algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 25(12):1543–1567, December 1995.

[2] M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1(3), 1997.

[3] R.J. Douglass. Interpreting three-dimensional scenes: A model building approach. *Computer Graphics and Image Processing*, 17:91–113, 1981.

[4] B.A. Draper, R.T. Collins, J. Brolio, A.R. Hanson, and E. Riseman. The schema system. *International Journal of Computer Vision*, 2:209–250, 1989.

[5] X. Lladó, J. Martí, J. Freixenet, and X. Muñoz. Feature selection through genetic algorithms. In *Proc. of the 3r Congrés Català d'Intel.ligència Artificial*, pp.152–156, Vilanova i la Geltrú, oct. 2000.

[6] J. Freixenet, X. Lladó, J. Martí, and X. Cufí. Use of decision trees in colour feature selection. Application to object recognition in outdoor scenes. In *Proc. of the IEEE Int. Conf. on Image Processing (ICIP'2000)*, vol.3, pp.496–499, Vancouver, sept. 2000.

[7] Y. Ohta. *Knowledge-based Interpretation of Outdoor Natural Color Scenes.* Pitman Publishing, 1985.