Pattern Recognition Letters 31 (2010) 1109–1119

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

How to involve structural modeling for cartographic object recognition tasks in high-resolution satellite images?

Guray Erus [a,*], Nicolas Loménie [a,b]

[a] Department of Mathematics and Informatics, LIPADE Laboratory, SIP Team, Paris Descartes University, 45 rue des Saints-Pères, 75006 Paris, France
[b] CNRS – IPAL – UMI 2955, Singapore

ARTICLE INFO

Article history:
Available online 15 January 2010

Keywords:
Object recognition
Structural analysis
Satellite images
CBIR

ABSTRACT

With the new generation of satellite systems, very high resolution satellite images will be available daily at a high delivery rate. The exploitation of such a huge amount of data will be made possible by the design of high performance analysis algorithms for decision making systems. In particular, the detection and recognition of complex man-made objects is a new challenge coming with this new level of resolution. In this study, we develop a system that recognizes such structured and compact objects like bridges or roundabouts. The original contribution of this work is the use of structural shape attributes in an appearance-based statistical learning method framework leading to valuable recognition and false alarm rates. This hybrid structural/statistical approach aims to construct an intermediate step between the low-level image characteristics and high-level semantic concepts.

1. Introduction

The new satellite systems like PLEIADES or QuickBird will provide satellite images up to 70 cm per pixel spatial resolution with a high delivery rate. Thus, new challenges for automatic interpretation of these valuable images are coming up. Much work has been done so far about the segmentation/recognition of textured areas like urban ones at low resolution. The detection of transport networks like roads or rivers considered as lineic structure networks in images has also been intensely studied and efficient algorithms have been proposed so far like in (Tupin et al., 1998; Mayer et al., 2006). However, few results have been presented for the recognition of complex, cartographic structures like bridges or roundabouts. These versatile objects do not lend themselves to classical statistical approaches for which a set of quantitative measures about the intensity distribution within a normalized window feeds a global statistical classifier used to discriminate between different objects. In the case of roundabouts for instance, the spatial configurations of object's parts are so versatile that a global statistical approach is not applicable. While the decomposition of the objects into parts is almost unavoidable, a purely structural part-based approach (Erus and Lomenie, 2005), is not appropriate to be applied on a real detection task, mainly due to the difficulty of segmenting the target objects to well-defined primitive shapes.

We propose in this study a hybrid approach based on learning the spatial configuration of structural primitives constituting an object, according to their statistical distribution in a labeled training set. Our method is evaluated on a cartographic object recognition task defined in the framework of a national program called Technovision that aims to assess the state of the art in object recognition (project ROBIN). For that academic-industrial joint study, The French Space Agency (CNES) prepared a database of satellite images containing cartographic objects. The recognition task consists of classifying images of size 100 × 100 pixels into one of the categories, like bridges, roundabouts, crossroads or isolated buildings.

In the object category recognition domain, the need for more structural analysis is progressively growing. In particular, low-level as well as high-level spatial relationships of object components are used as part of the object model definitions (Ferrari et al., 2008). However, the handling of such spatial relationships is not obvious since symbolic/linguistic reasoning is more or less involved (Erus and Lomenie, 2005) and the concept of spatial ontology is not straightforwardly usable in current image processing lines (Hudelot, 2005). However preliminary examples of the use of spatial relations in a recognition task can be found for example in (Colliot et al., 2006; Cao et al., 2009).

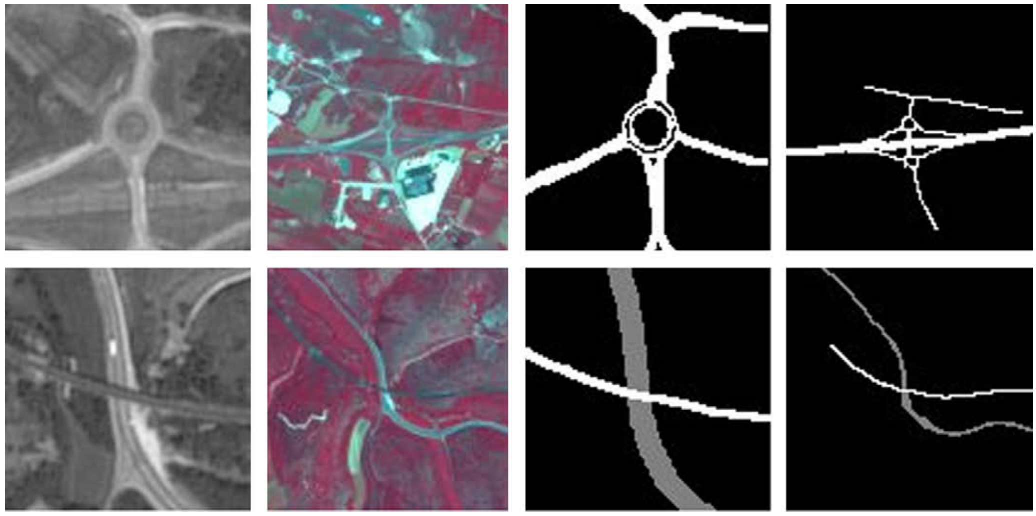
2. The image database

In the frame of the ORFEO program,[1] CNES prepared a database of high resolution cartographic object images. These images are simulations of the PLEIADES acquisition system based on SPOT5 images. In fact, the actual PLEIADES images will be available around 2010.

* Corresponding author. Tel.: +1 334 787 75 11; fax: +1 215 614 02 66.
E-mail address: guray.erus@uphs.upenn.edu (G. Erus).

[1] The research program set up to prepare, accompany and promote the use and the exploitation of the images derived from future Pleiades satellites.

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.01.013



(a) (b) (c) (d)

Fig. 1. SPOT5 images of cartographic objects exhibiting structures, a roundabout and a bridge: (a) the panchromatic images, (b) the multi-spectral images, (c,d) the corresponding manually segmented images.

The images are 100 × 100 pixels windows within which the object of interest is approximatively centered. In each of the 10 defined categories, there are nearly 100 such sample images. For each object instance in each category, this database is made of (see Fig. 1):

(1) the panchromatic image of the object;
(2) the multi-spectral image of the object;
(3) the manually segmented image as a mask drawn by an expert on both previous images.

The panchromatic images are acquired at the resolution of 2.5 m. per pixel and the multi-spectral ones at 10 m. per pixel. The challenge of that study however is to restrict oneself to the panchromatic images to handle the case for which only this channel is available due to transmission rate issues for instance, but also to be as generic as possible. Obviously, the use of such complementary information, if available, should improve the performance of the operational recognition system dramatically.

In this study, we focus our attention to the classification of two cartographic object categories, "*Roundabouts (RA)*" and "*Bridges (BR)*", which have a highly variable and compact structure. The long-term objective is the automatic detection of cartographic objects on very large satellite images (e.g. 24,000 × 24,000 pixels). However, we restrict ourself to the classification task proposed by the CNES, where the goal is to determine the categories of objects on manually extracted and labeled image patches. The basic premise is that a pre-processing module will focus the attention on the regions that may contain the target objects, and extract candidate regions. This initial coarse detection module can be constructed by edge density analysis or using one of the recent interest point detection algorithms for instance. The global recognition task (detection and classification) requires a sophisticated strategy and is still an active research issue. Fig. 2 presents sample panchromatic images belonging to different object categories.

3. State of the art

Few works deal with the recognition task of such specific, highly structured objects in satellite or even in aerial images (Trias-Sanz and Loménie, 2003). Iqbal and Aggarwal (2002) proposed to use the perceptual grouping of edge features for classification. The main idea here is to count evidences of the target class in the image. The evidences are detected starting with the line segments and applying grouping rules to obtain, hierarchically, co-terminations, L and U junctions, parallel lines and groups, and closed polygons. From these, three scalar features are calculated and the images are labeled as *Structure*, *Non-structure* or *Intermediate* using a nearest neighbor classifier. The important limitation of this method is that the low dimension of the final feature vector is only adequate for a very coarse classification.

Inglada (2007) developed a method to classify cartographic objects in satellite images. Low level and high-level geometric descriptors are calculated from the image for learning an SVM classifier. Our method follows a similar approach in that our feature vector is obtained from a set of structural primitives and a supervised learning approach is used for the classification.

In recent years, *appearance-based* approaches have been widely and successfully used in generic object detection problems (Agarwal and Roth, 2002; Csurka et al., 2004; Dorkó and Schmid, 2003; Li and Perona, 2005; Felzenszwalb and Huttenlocher, 2005; Fergus et al., 2003; Heisele et al., 2001; Leibe et al., 2004; Opelt et al., 2006; Shotton et al., 2007). In this approach, the object is considered to be represented by its appearances from different viewpoints, and this representation is learned from a set of object images. As the global appearance of an object is highly variable, the recent methods use in general local descriptors which remain stable on different instances of the object. The object consists of a collection of its parts, and the objective of the learning problem is to learn the characteristic parts of an object and their spatial organization. This approach is also conform to the neuro-psychophysical hypothesis of (Biederman, 1987) (and more recently Biederman, 2007), according to which an object is composed of a small set of elementary geometric shapes called *geons* and the first stage of human object recognition is the recognition of *geons*. The definition of the basic components and the modeling of their spatial arrangement is performed using various methods, giving birth to several detection algorithms.

A very common approach in extracting object parts is to detect local interest points on the image using detectors like Harris, maximally stable regions, difference of Gaussians, or the popular SIFT method. In (Viola and Jones, 2004) a very large number of simple Haar-like rectangular regions are extracted as features. Shotton et al. (2007) proposed to use contour fragments from the outer boundary of the objects. In (Ferrari et al., 2008) contour segments are used as descriptors.