Knowledge-Based Systems 95 (2016) 58–70

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Computing force field-based directional maps in subquadratic time

Iker Gondra[a], Iván Cabria[b,*]

[a] Department of Mathematics, Statistics and Computer Science, St. Francis Xavier University, Antigonish, NS, Canada
[b] Departamento de Física Teórica, Universidad de Valladolid, Valladolid, Spain

ARTICLE INFO

Article history:
Received 22 August 2015
Revised 21 October 2015
Accepted 7 December 2015
Available online 18 December 2015

Keywords:
Spatial knowledge
Spatial relationships
Pattern recognition
Computer vision
Image segmentation
Directional maps
Force fields

ABSTRACT

Spatial knowledge, i.e., knowledge about configurations among distinct spatial entities, plays a key role in everyday life and has wide applications in numerous application domains. In particular, computer vision systems exploit spatial knowledge to enhance their object recognition capabilities. The focus of this article is on the *representation* of spatial knowledge. Specifically, we concentrate on the representation of spatial relationships among objects in an image, a fundamental problem in pattern recognition and computer vision. In the study of spatial relationships, significant progress has been made by using directional maps. However, the time complexity of map generation is $O(N^2)$ and thus not practical. We present an improvement of a state-of-the-art approximation method that is based on an analogy with the force exerted by an object on another object. The improvement is based on the use of a family of forces whose dependence on the pixel–pixel distance r is of the type $1/r^{l+1}$ with $l > 1$. Analysis of time complexity shows that the improved method is $O(N^x)$, with $x = 1 + 3/(l+2)$. Application of this method to both artificial and real images validates the time complexity analysis and shows that the proposed method is much faster than the original method. Furthermore, for the special case of small objects relative to the image size, the proposed method generates more accurate maps and, in other cases, the difference in accuracy is very small.

1. Introduction

Spatial knowledge, i.e., knowledge about configurations among distinct spatial entities, plays a key role in everyday life and has wide applications in numerous application domains. In particular, computer vision systems exploit spatial knowledge to enhance their object recognition capabilities. Computer vision started with the goal of building machines that emulate the human visual system by using image-capturing equipment as the "eyes" and algorithms as the "brain processes" to perform perception for robots. The field of robotics itself has the potential to profoundly change our lives by, e.g., providing assistance or performing medical surgery. However, computer vision is now much broader. Applications, such as industrial automation and inspection, image search, gesture recognition for human-computer interaction, driver assistance, biological imaging and diagnosis, aids for the visually impaired, security and biometrics, just to name a few, keep arising.

A fundamental task of the human visual system is the ability to identify objects in an image, i.e., partitioning image pixels into semantically meaningful regions. As humans, we seem to do this effortlessly but it is a hard problem for computers that requires both low-level and high-level *object-specific knowledge*. The object segmentation problem is very closely related to the object recognition problem in the sense that, in order to segment an object, at least a partial recognition of the object is needed, which in turn requires segmentation. Indeed, object segmentation is a necessary first step in any recognition framework, even though most current research normally assumes that the object has already been segmented and focuses on classification.

Conventional segmentation extracts regions that satisfy uniformity criteria, e.g., uniformly colored. It tends to perform well in narrow domains, e.g., medical images, where visual variability is limited. See Fig. 1 for an illustration of this. Unfortunately, most real objects are quite heterogeneous and, thus, tend to be oversegmented into multiple regions. In essence, the conventional approach focuses on the understanding of the segmentation process from raw low-level visual data alone. Object segmentation can be tackled by using object-specific knowledge. Formally, given an image containing one or more objects and a set of labels corresponding to a set of object models known to the system, the task is to assign labels to regions in the image. The limiting factor is the amount of work that is required to incorporate each model into the knowledge base of object classes that the system can recognize.

* Corresponding author. Tel.: +34 983423141; fax: +34 983423013.
E-mail addresses: igondra@stfx.ca (I. Gondra), cabria@fta.uva.es (I. Cabria).

http://dx.doi.org/10.1016/j.knosys.2015.12.003
0950-7051/

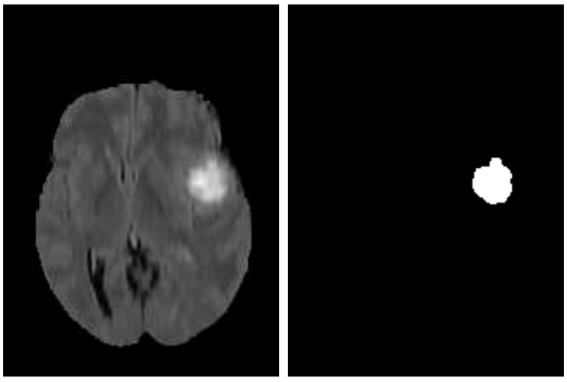

Fig. 1. Object recognition/segmentation in the medical imaging domain: (a) a brain MRI containing a tumor (the object of interest) composed of a single region with limited visual variability; (b) segmentation of the tumor using only the visual class-specific knowledge that a tumor appears as a bright region in an MRI. A conventional and simple region-growing segmentation algorithm was used to generate the segmentation.

However, people do not require that such models be provided in advance. As humans, we need to see only a few examples of a new object in order to learn a model of it and recognize new examples at a later time. This human ability poses a central challenge to the fields of computer vision, cognitive science and, in particular, machine learning. A number of learning-based approaches, e.g., [1–12], have been proposed.

A central issue is object modeling: what object-specific knowledge should be included and how it should be represented. An object is usually composed of several regions, e.g., the tree in Fig. 2(a) consists of a green (leaves) and a brown (trunk) fragment. However, the visual characteristics of the individual regions may not be discriminating enough to achieve individual recognition of the object's regions because regions that belong to a different object may have similar visual characteristics. To illustrate this (with a simplistic image and model), a tree model that includes only visual knowledge about the regions of a tree may result in the segmentation in Fig. 2(b) that includes the grass region as part of the tree. Fortunately, despite natural within-class variation in visual appearance, there are relationships among the regions that are generally shared by objects of the same class and thus support discrimination from objects in different classes. In particular, spatial knowledge about configurations among regions is fundamental in the human process of similarity comparison [13]. Published results [3,5–7,14–19] confirm the hypothesis that modeling spatial relationships can significantly improve recognition performance. For example, a richer tree model that includes both visual and spatial knowledge about the regions of a tree may result in the more accurate segmentation in Fig. 2(c). In this way, spatial knowledge helps in solving ambiguities by providing structural information about the spatial arrangement of the object components. Notice that, for real objects, the spatial arrangement of the object's components, i.e., spatial knowledge, is much less susceptible to variability than other types of knowledge such as, e.g., color or size.

In this paper we focus on *spatial knowledge representation*. Specifically, we concentrate on the representation of spatial relationships among objects in an image, a fundamental problem in pattern recognition and computer vision. Several methods [20–32] have been proposed to represent spatial knowledge on the spatial relationships among objects. In particular, significant progress has been made by using directional maps [23] (also known as fuzzy landscapes [24], spatial templates [25], applicability structures [27] and potential fields [28]). However, their computational cost (even when using an approximation scheme) remains very high. We present an improvement of the original force field-based method to compute directional maps [33]. The original method is an $O(NK\sqrt{N})$ approximation which, as far as we know, is the most efficient state-of-the-art approximation method.

Our contribution in this paper is two fold. First, we show that a good approximation requires a value of K that is proportional to $\sqrt{N}$ and thus the original force field-based method becomes $O(N^2)$. Second, we propose an improved force field-based method that is $O(N^x)$, $x = 1 + 3/(l+2)$, with $l > 1$, and show that it is a better alternative.

The rest of this paper is organized as follows. Directional maps, force fields and the original force field-based method are reviewed in Section 2. In Section 3 we show that, for a good approximation, the original force field-based method is $O(N^2)$. The improved force field-based method is presented in Section 4. Experiments, in Section 5, using both artificial and real images, validate the theoretical analysis and demonstrate the efficiency and accuracy of the improved method. Concluding remarks are given in Section 6.

2. Directional maps and force fields

Spatial relations among objects or, more generally regions, play a very important role in higher-level vision processes. Let us illustrate the motivation behind directional maps as a representation of spatial relations with the simple example in Fig. 2. Suppose we have a rule-based object segmentation/recognition system with a set of object models known to the system. The "tree" model might contain a rule such as: If two regions are adjacent AND the first region is somewhat green and circular AND the second region is somewhat brown and rectangular AND the first region is

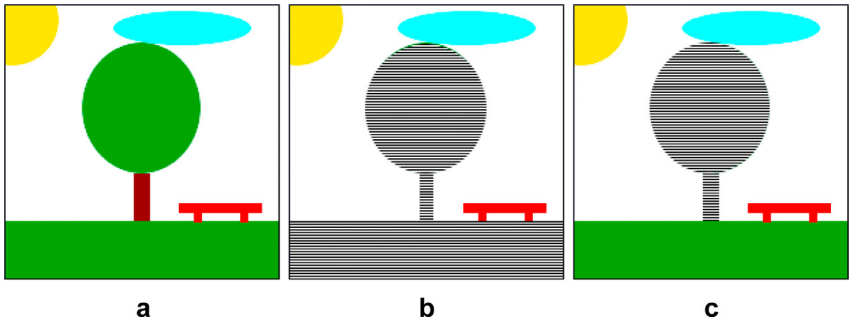

Fig. 2. Segmentation of an object of interest using different types of class-specific knowledge: (a) an image containing a tree (the object of interest) composed of a green region (the leaves) and a brown region (the trunk); (b) segmentation of the tree using only the visual class-specific knowledge: *A tree contains adjacent green and brown regions*. Notice that the grass, which is green, is also included as part of the tree; (c) segmentation of the tree using the visual and the spatial class-specific knowledge: *A tree contains adjacent green and brown regions AND the green regions are "on top" of the brown regions*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)