ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

Exact evaluation of targeted stochastic watershed cuts

Filip Malmberg *, Cris L. Luengo Hendriks, Robin Strand

Centre for Image Analysis, Department of Information Technology, Uppsala University, Sweden

ARTICLE INFO

Article history:
Received 26 January 2015
Received in revised form 17 December 2015
Accepted 5 January 2016
Available online xxxx

Keywords:
Image segmentation
Stochastic watershed
Watershed cut
Minimum spanning forest

ABSTRACT

Seeded segmentation with minimum spanning forests, also known as segmentation by *watershed cuts*, is a powerful method for supervised image segmentation. Given that correct segmentation labels are provided for a small set of image elements, called *seeds*, the watershed cut method completes the labeling for all image elements so that the boundaries between different labels are optimally aligned with salient edges in the image. Here, a randomized version of watershed segmentation, the *targeted stochastic watershed*, is proposed for performing multi-label targeted image segmentation with stochastic seed input. The input to the algorithm is a set of probability density functions (PDFs), one for each segmentation label, defined over the pixels of the image. For each pixel, we calculate the probability that the pixel is assigned a given segmentation label in seeded watershed segmentation with seeds drawn from the input PDFs. We propose an efficient algorithm (quasi-linear with respect to the number of image elements) for calculating the desired probabilities exactly.

1. Introduction

Image segmentation has traditionally been defined as the problem of dividing an image into smaller regions relative to image content, e.g., finding regions that are homogeneous with respect to color, intensity, texture, or some other quantity. In contrast, *targeted* image segmentation is the problem of finding one or more specific objects, for which additional information is available that determines which objects are being segmented [12,14]. This additional information, the *target specification*, can take many forms.

In *seeded* image segmentation, target specification is provided in the form of a partial segmentation. Given an image where a small subset of the pixels (called *seeds*) has been assigned correct segmentation labels (e.g., object or background), the segmentation method completes the labeling for all pixels. The seeded segmentation paradigm is often used in interactive segmentation applications, where the user provides the seeds. Several methods have been proposed for extending the segmentation labels from the seeds to all pixels. Leading methods include *minimal graph cuts* [5], *shortest path forests* [10], *random walks* [13], and *minimum spanning forests* [8,9]. Couprie et al. [7] established a strong theoretical connection between these methods and showed that in fact, they can all be formulated as special cases of a single general seeded segmentation framework. Here, we restrict our attention to segmentation by minimum spanning forests, also known as the *watershed cut* approach.

The key issue in seeded segmentation is to find appropriate seeds. Devising an automatic method for finding correct seeds is a very challenging problem in all but the most trivial cases. The labels of the seeds are typically preserved by the segmentation method, and thus a single erroneous seed can drastically reduce the quality of the segmentation.

* Corresponding author. Tel.: +46 18 471 78 49; fax: +46 18 51 19 25.
E-mail address: filip.malmberg@it.uu.se (F. Malmberg).

http://dx.doi.org/10.1016/j.dam.2016.01.006
0166-218X/

Please cite this article in press as: F. Malmberg, et al., Exact evaluation of targeted stochastic watershed cuts, Discrete Applied Mathematics (2016), http://dx.doi.org/10.1016/j.dam.2016.01.006

Here, we address the problem of finding suitable seeds by formulating a stochastic version of seeded segmentation. Instead of requiring a fixed set of seeds, our method asks for target specification in the form of a number probability density functions (PDFs) – one for each segmentation label– defined over the pixels of the image. These PDFs indicate, for each pixel, the probability of placing a seed with the corresponding label at that pixel. For segmentation with two labels (here denoted *object* and *background*) the proposed method can be defined as follows. Consider the following stochastic segmentation process:

1. Randomly select two seeds, one representing the object and one representing the background, according to the given PDFs.
2. Label each pixel as either object or background by computing the watershed cut relative to the randomly selected seeds.

The proposed method resolves the question: What is the probability that the pixel is labeled as object or background, respectively, in a single execution of the above procedure? These probabilities are transformed to *segmentation scores* by a simple re-scaling procedure. A final segmentation is derived by assigning to each pixel the label with the highest score. The input PDFs may be given interactively, or generated by an automatic method. Such automatic methods might for example be based on the output of a segmentation atlas, or other prior knowledge of the approximate location, color distribution, or other features, of the objects of interest. The precise method used for generating the input PDFs is of course highly dependent on the segmentation problem at hand, and is not addressed here. We do, however, argue that devising automatic algorithms for generating a PDF for seed placement is an easier task than finding appropriate fixed seeds, as required by the original watershed cut method. With fixed seeds, all seeds must be placed inside the corresponding object in order to obtain a correct segmentation. In contrast, the proposed method may still generate an appropriate segmentation even if the input PDFs are non-zero outside of their corresponding objects.

The idea of using watershed segmentation with randomized seeds for targeted segmentation is in itself not new. This idea was previously explored, e.g., by Angulo and Velasco-Forero [2] and Bernard et al. [4], who demonstrated good results in practical applications. In both these papers, the probability of assigning a pixel to a given class was estimated by Monte Carlo sampling, i.e., repeatedly selecting random seeds according to the given distributions and counting the frequency with which each pixel is assigned the given label. To obtain a good estimate, however, a large number of iterations are needed. Thus the Monte Carlo approach is computationally expensive.

Our main contribution is to show that the desired probabilities can be computed exactly, without resorting to Monte Carlo sampling. We present an efficient method, with quasi-linear time complexity with respect to the number of pixels, for calculating the exact probabilities. Additionally, we show how the two-label segmentation method described above can be extended to perform segmentation with $k \geq 2$ labels. A probabilistic interpretation of the multi-label segmentation algorithm, similar to that described above for the two-label case, is given. We demonstrate empirically that the proposed method is faster than the Monte Carlo approach by orders of magnitude, while giving exact results.

The proposed computation method relies on the property that, under certain conditions, it is possible to compactly represent all possible watershed segmentations of an image in a minimum spanning tree defined over the image elements [19,17]. This compact representation reveals that while the number of possible binary segmentations of an image is extremely large, the number of watershed segmentations is much smaller—in fact it is in the same order as the number of image elements. Moreover, we can give an exact specification on how the seeds must be configured in order to produce a given watershed segmentation, and determine the probability of such a configuration occurring.

2. Related work

The idea of randomly selected seeds in conjunction with watershed segmentation was first explored by Angulo and Jeulin [1], who proposed a method for identifying salient contours in an image. Their approach relies on estimating, by Monte Carlo sampling, the probability that a given piece of contour in the image belongs to the segmentation boundary in seeded watershed segmentation with a fixed number of random seeds independently drawn from a uniform distribution. Contours that appear with high probability are assumed to be more important. They also propose to use the resulting probability map as a starting point for unsupervised segmentation. Meyer and Stawiaski [19] showed that the probability map can be calculated exactly, without Monte Carlo sampling. Their work was later extended by Malmberg and Luengo [17], who proposed an efficient (quasi-linear with respect to the number of image elements) algorithm for computing the exact probability map.

The idea of using stochastic watersheds for supervised, or targeted, segmentation was previously explored by Angulo and Velasco-Forero [2] and Bernard et al. [4], who proposed the use of non-uniform PDFs for seed generation to identify specific objects in an image. Again, their method relied on Monte Carlo sampling for estimating the resulting probability maps. The method proposed by Angulo and Velasco-Forero also differs from the method proposed here in that their method calculates, for each pixel, the probability of belonging to the *boundary* of a specific object. In contrast, our method calculates, for each pixel, the probability of being assigned a given segmentation label.

Couprie et al. [7] proposed to include unary terms in watershed cut segmentation, thereby facilitating "soft" target specification similar to the probabilistic input proposed here.

Please cite this article in press as: F. Malmberg, et al., Exact evaluation of targeted stochastic watershed cuts, Discrete Applied Mathematics (2016), http://dx.doi.org/10.1016/j.dam.2016.01.006