Computers & Geosciences 85 (2015) 74–85

Contents lists available at ScienceDirect

Computers & Geosciences

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

A seed expanding cluster algorithm for deriving upwelling areas on sea surface temperature images

Susana Nascimento [a,b,*], Sérgio Casca [b], Boris Mirkin [c,d]

[a] Department of Computer Science, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[b] NOVA Laboratory for Computer Science and Informatics (NOVA LINCS), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2819-516 Caparica, Portugal
[c] Department of Data Analysis and Machine Intelligence, National Research University Higher School of Economics, Moscow, Russian Federation
[d] Department of Computer Science, Birkbeck University of London, UK

ARTICLE INFO

Article history:
Received 2 October 2014
Received in revised form
4 March 2015
Accepted 8 June 2015
Available online 10 June 2015

Keywords:
Seeded region growing
Approximate clustering
Homogeneity criterion
SST images
Upwelling

ABSTRACT

In this paper a novel clustering algorithm is proposed as a version of the seeded region growing (SRG) approach for the automatic recognition of coastal upwelling from sea surface temperature (SST) images.

The new algorithm, one seed expanding cluster (SEC), takes advantage of the concept of approximate clustering due to Mirkin (1996, 2013) to derive a homogeneity criterion in the format of a product rather than the conventional difference between a pixel value and the mean of values over the region of interest. It involves a boundary-oriented pixel labeling so that the cluster growing is performed by expanding its boundary iteratively. The starting point is a cluster consisting of just one seed, the pixel with the coldest temperature. The baseline version of the SEC algorithm uses Otsu's thresholding method to fine-tune the homogeneity threshold. Unfortunately, this method does not always lead to a satisfactory solution. Therefore, we introduce a self-tuning version of the algorithm in which the homogeneity threshold is locally derived from the approximation criterion over a window around the pixel under consideration. The window serves as a boundary regularizer.

These two unsupervised versions of the algorithm have been applied to a set of 28 SST images of the western coast of mainland Portugal, and compared against a supervised version fine-tuned by maximizing the *F*-measure with respect to manually labeled ground-truth maps. The areas built by the unsupervised versions of the SEC algorithm are significantly coincident over the ground-truth regions in the cases at which the upwelling areas consist of a single continuous fragment of the SST map.

1. Introduction

The coastal ocean of Portugal is under the influence of northerly winds which are favorable to the occurrence of upwelling during the summer. This is a phenomenon that manifests at the surface by cold, less salty and nutrient-rich waters over the whole shelf, and by filaments of upwelled waters penetrating into the open ocean. Therefore, coastal upwelling is a phenomenon of ocean dynamics whose study is fundamental to the development of climate models and resulting applications relevant to fisheries, coastal monitoring and detection of pollutants. The identification and continuing monitoring of upwelling is thus a crucial part for the study of the dynamics of the oceans, which involves analyzing large volumes of data such as remote sensing images in the infrared range. The identification of upwelling is frequently analyzed with sea surface temperature (SST) images because of the temperature contrast between the cold7 upwelling waters and the warmer near-shore oceanic waters. These SST maps are typically obtained with the thermal infrared channels of the advanced very high resolution radiometer (AVHRR) sensor on board NOAA-n satellite series.

Several approaches have been proposed for automatic upwelling detection, each involving rather complex computations to achieve satisfactory segmentation results. For example, the works by Kriebel et al. (1998), Arriaza et al. (2003) and Chaudhari et al. (2008), using artificial neural networks, require the extraction of various morphological features followed by a region building using an iterative thresholding process. Marcello et al. (2005) used an extensive comparison of automatic thresholding techniques followed by a combination of watershed transform and region growing approaches. Nieto et al. (2012) developed an upwelling front detection system based on edge detection using a combination of multiple windows followed by a thorough process of

* Corresponding author. Fax: +351 21 294 8541.
E-mail address: snt@fct.unl.pt (S. Nascimento).

http://dx.doi.org/10.1016/j.cageo.2015.06.002
0098-3004/

recovering missing segments.

In our previous work (Nascimento and Franco, 2009; Nascimento et al., 2012), we characterized and automated the process of delineation of upwelling regions and boundaries using a fuzzy clustering method supplemented with what is referred to as anomalous cluster initialization process. Our system, FuzzyUPWELL, has proven to work rather reliably on the available collections of oceanographic images. Yet the system has an evident empiric flavor. It operates over the temperature data only, and uses no geographic information. Also, the system needs some expert driven knowledge to fine-tune thresholds which is not always easy to operationalize.

In this paper, we tackle the problem of determination of upwelling regions by using pixels and their spatial arrangement on temperature maps using a model inspired on the development of the upwelling as a process of step-by-step adding pixels according to the similarity of their temperatures to the temperatures of those already in the region. To this end, we adopt and considerably modify the popular seeded region growing (SRG) method introduced by Adams and Bischof (1994) for region based segmentation (see also e.g. Mehnert and Jackway, 1997; Freixenet et al., 2002; Fan et al., 2005; Shih and Cheng, 2005; Verma et al., 2011). This method tries to grow a region whenever its interior is homogeneous according to a certain feature such as intensity, color or texture, called the *feature of interest* (FoI). The algorithm follows the strategy based on the growth of a region, starting from one or several 'seeds' and by adding similar neighboring pixels. The growth is controlled by using a homogeneity criterion so that the merging decision is generally taken based only on the contrast between the evaluated pixel and the region. However, it is not always easy to decide when this difference is small (or large) enough to make a reasonable decision (Freixenet et al., 2002).

The seeded region growing image segmentation approach has been widely used in various medical image applications like magnetic resonance image analysis and unsupervised image retrieval in clinical databases (Mancas et al., 2005; Whitney et al. (2006); Wu et al., 2009; Harikrishna-Rai and Gopalakrishnan-Nair, 2010; Zanaty, 2013). The approach has been also successfully applied in color image segmentation with applications in medical imaging, content-based image retrieval, and video (Fan et al., 2001; Shih and Cheng, 2005; Ugarriza et al., 2009; Verma et al., 2011), and yet in remote sensing image analysis (Wang et al., 2010; Byun et al., 2011; Zhang et al., 2013).

Main challenging issues that arise with SRG methods are as follows:

(i) How to select the initial seeds in practice, and how critical is the seed selection to getting a good segmentation?
(ii) How to choose the region homogeneity criterion and how to specify its threshold?
(iii) How to organize the pixel labeling procedure efficiently?

Most approaches to SRG involve homogeneity criteria in the format of difference of the feature of interest between that at the pixel to be labeled and the mean value at the region of interest (Adams and Bischof, 1994; Zhou et al., 2004; Fan et al., 2005; Shih and Cheng, 2005; Ibrahim et al., 2010; Verma et al., 2011). A weak point of these algorithms is the definition of the non-homogeneity threshold at which the pixels under consideration are considered as failing the homogeneity test and, therefore, cannot be added to the region. Such a definition is either expert driven or supervised in most of the currently available algorithms (Zhou et al., 2004; Fan et al., 2005; Mat-Isa et al., 2005; Shih and Cheng, 2005; Ibrahim et al., 2010; Verma et al., 2011). Another issue of the SRG methods is the pixel ordering for testing them on joining the growth region and labeling (Mehnert and Jackway, 1997). Many SRG algorithms grow the regions using a sequential list sorted according to the dissimilarity of unlabeled pixels to the growth region (Adams and Bischof, 1994; Grinias and Tziritas, 2001; Harikrishna-Rai and Gopalakrishnan-Nair, 2010; Verma et al., 2011). The disadvantage is that the segmentation results are very much sensitive to this order.

To meaningfully overcome these issues, we apply the concept of approximate cluster from Mirkin (1996, 2013) to the SRG framework. This approximation approach leads us to accept a mathematically equivalent, though somewhat unusual, homogeneity criterion, in the format of a product rather than the conventional difference between the pixel and the mean of the region of interest. Specifically, to segment an SST map, we first subtract the average temperature value from all the temperature values. After this, a pixel p under consideration is added to the cluster if and only if $c \times t(p) \geq \pi$ where c is the average temperature of the pixels already in the cluster, $t(p)$ the temperature of the pixel p, and π a threshold defined by the maximum difference between cold water pixels and the rest. This process is moderated via usage of the concept of window of a pre-specified size around the pixels under consideration: only those within the window are involved in the comparison processes. This provides for the spatial homogeneity and smoothness of the growing region. Indeed, only borderline pixels are subject to joining in because the windows around remote pixels just do not overlap the growing region. Therefore, there is no need in specifying the order of testing for labeling among pixels: all those borderline pixels can be considered and decided upon simultaneously. The process starts from a cluster consisting of just one pixel, the coldest one, according to the approximation clustering criterion. The preprocessed temperature of this pixel is negative with a relatively large absolute value. Since the temperature is a rather conservative characteristic and would not change sharply with respect to space, there must be a bulk of pixels within the window centered at the starting point that will have similarly cold temperatures so that the products of those by the preprocessed temperature at the seed have large positive values and thus have to join in the growing region. This shows that our region growing process initializes with a fragment of the coldest pixels, which is rather robust. Using the window as a constraint brings forth another desirable property. The growing region in our process normally would cover a continuous fragment of the ocean. To warrant this, we tried an additional homogeneity condition: the density of the cluster being built should be greater than or equal to a threshold α. The density is defined locally as the fraction of pixels belonging to the cluster in the window surrounding the pixel under consideration. Yet this proved unnecessary in our experimental computations: the spatial continuity of the region has been warranted by the step-by-step procedure of the growing region to reach out of the boundary in all directions. The proposed method, one seed expanding cluster (SEC), holds out of the issue of dependence on the pixel sorting order. Moreover, the simultaneous borderline labeling considerably speeds up the SRG procedure.

To specify the similarity threshold to judge whether the pixel under consideration should be added to the growing region or not, we take Otsu's (1979) thresholding method. This method fine-tunes the similarity threshold by finding the maximum inter-class variance that splits between warm and cold waters. This defines what we refer to as the baseline version of the SEC algorithm. Also, we introduce a self-tuning version of the algorithm, SelfT-SEC, at which the homogeneity threshold parameter is derived from the cluster approximation criterion.

These two unsupervised versions of the SEC algorithm have been applied to a set of 28 SST images of the western coast of mainland Portugal, and compared against a supervised version fine-tuned by maximizing the *F*-measure with respect to ground-truth maps annotated by expert oceanographers.

The rest of the paper is organized as follows. Section 2 describes the original SRG method. In Section 3 we describe the new seed expanding cluster algorithm, derive its optimal homogeneity