Neurocomputing 168 (2015) 55–64

Contents lists available at ScienceDirect

Neurocomputing

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

A weakly supervised geodesic level set framework for interactive image segmentation

Tinghuai Wang [a,*], Huiling Wang [b], Lixin Fan [a]

[a] Nokia Research Center, Nokia Corporation, Visiokatu 3, 33720 Tampere, Finland
[b] Lappeenranta University of Technology, Skinnarilankatu 34, 53850 Lappeenranta, Finland

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form
13 April 2015
Accepted 9 June 2015
Communicated by Yi Zhe Song
Available online 18 June 2015

Keywords:
Weakly supervised learning
Geodesic distance
Image segmentation
Level set methods

ABSTRACT

Interactive image segmentation is growingly useful for selecting objects of interest in images, facilitating spatially localized media manipulation especially on touch screen devices. We present a robust and efficient approach for segmenting image with less and intuitive user interaction. Our approach combines geodesic distance information with the flexibility of level set methods in energy minimization, leveraging the complementary strengths of each to promote accurate boundary placement and strong region connectivity while requiring less user interaction. We harness weakly supervised segment annotation to maximize the user-provided prior knowledge. This leads to a seed generation algorithm which enables image object segmentation without user-provided background seeds. We demonstrate that our approach is less sensitive to seed placement and better at edge localization, whilst requiring less user interaction, compared with the state-of-the-art methods.

1. Introduction

Interactive image segmentation aims to extract the foreground object out of the cluttered background in natural images with the aid of user-provided prior knowledge. This interactive approach is becoming more and more popular especially on touch screen devices, since prior knowledge about the desired object and background can be easily defined with intuitive user interactions such as marking of object boundaries [18,6,41], placing a bounding box around the foreground object [34,23], and loosely drawing scribbles on foreground/background regions [7,19,5,40]. Among these different interaction modes, drawing scribbles is of the most interests for touch screen devices since it provides sufficiently rich information about the desired foreground and background regions while requiring less amount of flexible user interventions (see Fig. 1 for examples). Despite the significant advances delivered in recent years, two open issues prevent the scribble-based approach from being used by massive mobile device end-users. First, to draw scribbles on both the foreground and background objects is too troublesome for small-sized touch screens. The need to switch between foreground and background scribbles further complicates the UI—at least one toggle button needed on the small touch screen. Second, it is cumbersome for novice end users to perform tedious fine-tunings to correct mis-segmentations, especially in the case of noisy boundaries and disjoint regions which may severely affect the quality of the target applications. In order to deal with these two compelling issues, we propose in this paper a novel interactive segmentation method that (a) relieves the burden of drawing scribbles on background regions; (b) automatically promotes accurate boundary placements and fills contiguous, coherent regions without substantial user corrections.

The novelty of this paper lies in two aspects. Firstly, we contribute a seed generation algorithm formulated as a weakly supervised segment annotation problem, initially generating background seeds from the given image border, pruning mis-labelled background seeds depending on their *superpixel geodesic distance* from foreground seeds and generating new background seeds via cascaded classification (see Section 4). Secondly, we propose a geodesic level set framework that combines *geodesic region information* with the flexibility of *the level set methods* in energy minimization. As illustrated in Fig. 1 and the rest of this paper, these two building blocks conspire to improve the boundary localization accuracy and reduce disjoint regions.

1.1. Related work

Geodesic distance has been used for scribble-driven interactive image segmentation methods [5,33], which selectively fill the desired region by expanding from the interior of the selected object outwards without explicitly considering the object boundary. Using geodesic distance makes it advantageous for

* Corresponding author.
E-mail addresses: tinghuai.wang@nokia.com (T. Wang), huiling.wang@lut.fi (H. Wang), lixin.fan@nokia.com (L. Fan).

http://dx.doi.org/10.1016/j.neucom.2015.06.014
0925-2312/

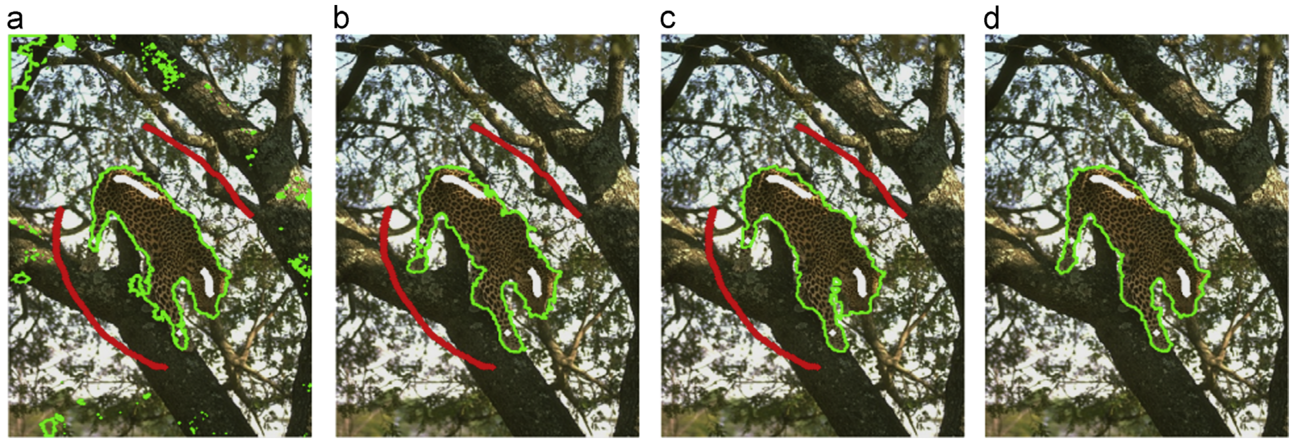

Fig. 1. Geodesic level set framework with seed generation for interactive image segmentation. (a) Geodesic segmentation suffers from the sensitivity to seed placement and may fail to accurately identify the real object boundaries due to the lack of edge information and an optimization framework. (b) Geodesic active regions framework is prone to under- or over-segment objects without encoding spatial information. (c) Geodesic graph cut suffers from short-cutting problem when the color models become less distinct. (d) Geodesic level set delineates the foreground object with more accurate boundary localization and less disjoint regions. In addition, a seed generation algorithm substantially reduces the amount of user interaction.

segmenting objects with complex topologies, whilst it may suffer from a bias that favors shorter paths from the seeds. These methods may also fail to accurately identify the real object boundaries due to the lack of an explicit optimization framework taking into account of edge contrast, something at which level set or graph cut based methods generally excel.

An important category of seeded segmentation algorithms is those performing a *Graph Cut* optimization. Boykov and Jolly [7] addressed object segmentation via max-flow/min-cut energy minimization, in which the regional cost is based on individual pixel color (or intensity) values and the boundary cost is derived from the dissimilarities between two neighboring pixels in terms of intensity values and/or spatial distances.

Notably the user-specified scribbles serve as hard constraints and also provide statistical information. Despite the improvements reported by the approaches combining geodesic distance information and graph cut [20,32], these approaches are limited by the inherent bias of graph cut towards shorter paths, i.e. small segments as the optimization sums over the boundaries of segmented regions. While an iterative graph cut method might be used to improve the segmentation [34], additional user interaction is needed. Anh et al. [2] proposed a continuous-domain convex active contour model combining geodesic distance based probability with color distance based probability. However, the geodesic distance based probability in [2] essentially relies on the difference between geodesic distances from foreground (FG) and background (BG) seeds. This difference generally accounts for the sensitivity to seeds placement and disjointed regions in geodesic segmentation.

Level set methods [44] neatly enable the minimization of energy functionals such as those proposed by Mumford-Shah [25] or Zhu-Yuille [45]. Caselles et al. [9] and Kichenassamy et al. [22] proposed the edge-based geodesic active contour model for image segmentation as a geometric alternative for snakes, to obtain the optimal image partition by finding the set of minimal length geodesic curves that are attracted by the real region boundaries. This approach depends primarily upon image gradient and is sensitive to clutter and image noise, and consequently, often mistakenly converges to local minima that do not well describe the intended region shape. More robust approaches that encode region information were proposed later by Paragios et al. [28]. Paragios et al. [29] proposed a region-based active contour model extending the previous geodesic active contour formulation, finding the geodesic curve of minimal length that is in accordance with the desired image characteristics. Chan and Vese [10] proposed an active contour model based on Mumford-Shah segmentation techniques and the level set method, partitioning objects via intensity distributions modeled with different variances.

Higher level prior knowledge such as geometric shape priors have been introduced to level set framework [36,31,14,8,39,35,11,13,42]. Rousson and Paragios [36] proposed the variational integration of the shape prior based on the assumption of a Gaussian distribution. Paragios et al. [31] introduced implicit representations and distance transforms to represent shape in a higher dimensional framework for optimization under level sets. Tsai et al. [39] proposed a very efficient implementation of shape-driven level set segmentation by directly optimizing in the linear subspace spanned by the principal components. The use of nonparametric density estimation to model larger classes of level set based shape distributions was proposed by Cremers et al. [11] and Rousson and Cremers [35]. Paragios [27] proposes to integrate user interactive constraints to level set framework by taking into account distance from seeds. Yezzi et al. have explored the fusion of multiple cues, including geometry, in their segmentation frameworks [15,21]. A more complete survey of previous works integrating multiple priors to level set framework, the reader is refereed to [12].

Several recent efforts have been exploiting novel approaches to learning from user inputs for segmentation. Straehle et al. [37] introduce a weakly supervised segmentation method by learning an edge model from sparse region scribbles. However, the method is not specifically tackling interactive FG/BG segmentation problems on natural images, which requires substantial amount of FG and BG user scribbles to achieve a moderate level of accuracy. Wu et al. [43] formulate the interactive segmentation problem as a multiple instance learning (MIL) task by generating positive bags from pixels of sweeping lines within a user-provided bounding box. Nieuwenhuis et al. [26] propose an algorithm based on texture and color information, which leverages the co-sparse analysis model for image segmentation within a convex multilabel optimization framework. CRF based algorithms to tolerate user input errors in scribble-based interactive segmentation have been proposed in [38,4]. Nevertheless, these approaches are relative demanding in the amount of FG and BG scribbles in order to 'correct' the errors in the user input.

2. Level set methods revisited

The basic idea of active contour models implemented via level set methods is that a contour $\mathcal{C}$ in a domain Ω can be represented