J. Vis. Commun. Image R. 38 (2016) 745–752

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Spatially constrained level-set tracking and segmentation of non-rigid objects ☆

Xuan Cheng [a], Ming Zeng [b], Xinguo Liu [a,*]

[a] State Key Lab of CAD&CG, Zhejiang University, China
[b] Software School of Xiamen University, China

ARTICLE INFO

Article history:
Received 28 November 2014
Revised 3 April 2016
Accepted 8 April 2016
Available online 27 April 2016

Keywords:
Segmentation-based tracking
Level-set
Hough voting
Back-projection

ABSTRACT

Level-set is a widely used technique in segmentation-based tracking due to its flexibility in handling 2D topological changes and computational efficiency. Most existing level-set models aim at grouping pixels that have similar features into a region, without consideration of the spatial relationship of these pixels. In this paper, we present a novel level-set tracking method that incorporates spatial information to improve the robustness and accuracy of tracking non-rigid objects. Both tracking and segmentation are performed in a unified probabilistic framework, with additional spatial constraints from a part-based model—the Hough Forests. In the stage of tracking, the rigid motion of the target object is estimated by rigid registration in both the color space and the Hough voting space. Then in the stage of segmentation, some support points are obtained from back-projection, and guide the level-set evolution to capture the shape deformation. We conduct quantitative evaluation on two recently proposed public benchmarks: a non-rigid object tracking dataset and the CVPR2013 online tracking benchmark, involving 61 sequences in total. The experimental results demonstrate that our tracking method performs comparably to the state-of-the-arts in the CVPR2013 benchmark, while shows significantly improved performance in tracking non-rigid objects.

1. Introduction

Recently segmentation-based tracking have attracted great attention in the field of object tracking. It could provide a more accurate foreground/background separation, compared with classical tracking methods which often use a bounding box to represent the target object. This is particularly important for tracking non-rigid objects, such as hands and pedestrians, because segmentation would introduce less undesirable background information and help avoid the drifting problem in a certain degree.

Level-set is a widely used technique in segmentation-based tracking, due to its flexibility in handing complex topological changes. Several methods have been proposed to track non-rigid or deformable objects in a level set framework. A representative work is the pixel-wise posteriors tracking [1], which performs level-set evolution and warping iteratively to track an object's contour and achieve some promising results. Then this work is extended to other tracking tasks, including multi-object tracking [2], pedestrians tracking [3,4] and 3D objects tracking [5]. Most of these methods rely on a global appearance model, such as color histograms, because of its convenience to describe the arbitrary shape of general object and computational efficiency. When it comes to tracking highly non-rigid objects in front of complex and cluttered backgrounds, the ability of a single global appearance model is relatively limited.

To capture the local property of objects, part-based representation is often used. For instance, the Deformable Part-based Model [6] is a prominent method in the domain of object detection. Some implicit part-based models can be obtained using the generalized Hough-transform [7], where each part of the object is mapped into a voting space in the Hough Forests framework. Furthermore, Godec et al. [8] extend the idea of Hough Forests to the online domain, and propose a Hough-based object tracking method.

The motivation of this paper is to integrate these part-based concepts into the level-set tracking. Compared with simple color histograms, part-based model such as Hough Forests, can provide mid-level information including the texture in local parts and the spatial constraints between these parts. According to previous literature, this information can benefit both the location [7] and the segmentation [8] of target object. However, it's not straightforward to utilize the part-based model in a level-set formalism, since

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: chengxuan90@gmail.com (X. Cheng), xgliu@cad.zju.edu.cn (X. Liu).

http://dx.doi.org/10.1016/j.jvcir.2016.04.009
1047-3203/

level-set usually operates on pixels directly while part-based model tends to handle local patches.

To address this issue, we present a probabilistic level-set framework, which uses Hough voting and support points provided by Hough Forests as the spatial constraints for tracking and segmentation. Begin with a generative model that utilizes color histograms and Hough Forests as the appearance model, we derive the formulation of tracking and segmentation respectively. In the stage of tracking, rigid registration is performed to make the new frame match the old frame in both color space and Hough voting space. Then in the stage of segmentation, we use back-projection to find the support points which have high confidence in belonging to the target object. These support points act as soft constraints for the subsequent level-set evolution. Finally, color histograms and Hough Forests are updated according to the segmentation results. Fig. 1 shows an example, and Fig. 2 depicts the overall procedure of our method.

The rest of this paper is organized as follows: Section 2 reviews some related works; Section 3 derives a probabilistic framework from a generative model; Section 4 outlines the tracking process; Section 5 shows the level-set segmentation; Section 6 presents online updating process; Section 7 shows some experimental results; Section 8 concludes this paper.

2. Related work

Object tracking methods could be categorized into two main classes, namely, bounding-box-based tracking and segmentation-based tracking. For bounding-box-based tracking, we refer readers to a comprehensive survey [9] and a recent benchmark [10]. Here we briefly review some representative works of segmentation-based tracking. Nejhum et al. [11] proposed to track articulated objects with a set of adaptively rectangular blocks, followed by a refine step using graph-cut segmentation. Fan et al. [12] introduced image matting into a tracking process, where the coarse tracking results provide suitable scribbles for matting, and both tracking and matting model are updated in closed-loop manner. Belagiannis et al. [13] combined tracking and segmentation in another way, where segmentation is used for sampling in the particle filtering framework. Godec et al. [8] extended the Hough Forest [7] into the online domain, which use a voting scheme to find the center of the object and back-project the pixels that voted for the object center to initialize the GrabCut segmentation. Later, Duffner and Garcia [14] proposed to use pixel-based descriptors instead of patch-based descriptors in the same online Hough voting scheme, and thus achieve faster tracking speed. Level-set technique, which implicitly represents the contours as the zero level-set of a higher dimensional function, is widely used in segmentation-based tracking. For example, Cremers [15] proposed a Bayesian level-set framework to track the contour of object and learn the dynamical shape priors simultaneously. Sun et al. [16] utilize the online boosting method as a detector to find the position of the object, and then obtain the contour with level-set. Bibby and Reid [1] derived a probabilistic level-set framework based on the pixel-wise posteriors. Their method comprises a rigid registration between frames, a segmentation and online appearance learning. Horbert et al. [4] improved on this work by using additional level-sets to enforce the spatial constraints of different parts of the object, in the context of pedestrians tracking. Our work is also built on [1], but we incorporate spatial constraints from Hough Forests to improve tracking performance.

3. A probabilistic level-set framework

Similar to [1], we present a probabilistic level-set framework for combined tracking and segmentation of an object. In this section, we firstly collect the notations used throughout this paper, and then emphasize two features of the Hough Forests model. We also describe a generative model that set the foundation of our proposed framework, and make some inferences to derive the formulations of tracking and segmentation.

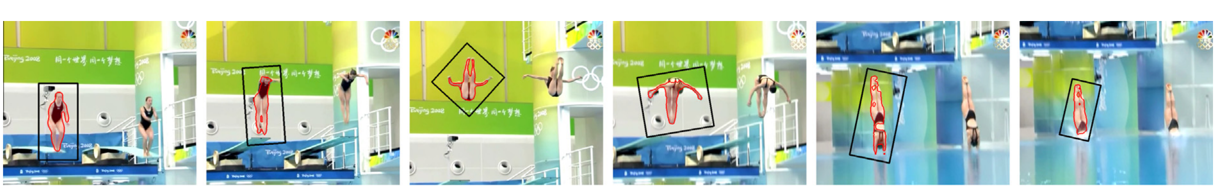
3.1. Notations

Let $\mathbf{I}$ denote the image frame, and $\mathbf{I}_o$ denote the object frame (the black bounding box as shown in Fig. 3(a)). Let $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ denote the set of pixel locations in the object frame coordinate, and $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ denote the set of corresponding pixel values. The object being tracked is represented by its shape $\mathbf{C}$, its position in the image $\mathbf{W}(\mathbf{x}, \mathbf{p})$, the color histograms M and the Hough Forests F. The shape is represented by the zero level-set $\mathbf{C} = \{\mathbf{x} | \mathbf{\Phi}(\mathbf{x}) = 0\}$ of an embedding function $\mathbf{\Phi}(\mathbf{x})$. The position is described by a warp $\mathbf{W}(\mathbf{x}, \mathbf{p})$ which warps a pixel location $\mathbf{x}$ in the object frame coordinate into the image frame coordinate according to parameters $\mathbf{p}$. The color histograms $M = \{M_f, M_b\}$ are built on foreground pixels and the nearby background pixels, with 32 bins per channel. The Hough Forests F contain two components that serve as spatial constraints, with F_v representing Hough voting map and F_s representing a set of support points.

3.2. Hough forests

Hough Forests have been proposed by Gall et al. [7] in the context of object detection. Because of the speed and robustness to noisy training data, Hough Forests inspired a series of extensions and applications in computer vision, such as human pose estimation from depth [17] and facial feature points detection [18]. Hough Forests are in fact a variant of the Implicit Shape Model [19], and thus use a star shaped model to represent the object, where each part of the object is connected to a centroid point through Hough voting procedure.

A training sample for Hough Forests is an image patch that consists of three elements: the feature of the patch, the foreground/background label of the patch and the offset vector pointing to expected object center. In the training process, several randomized tree structures are built to optimize the class impurity or the offset



(a) (b) (c) (d) (e) (f)

Fig. 1. Simultaneously tracking and segmentation of non-rigid object using our method. In tracking, a bounding box (black) is obtained by image warping with translation and rotation. In segmentation, a contour (red) is calculated by level-set evolution. The size of the bounding box is slightly updated to make it tighter to the target object. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)