J. Vis. Commun. Image R. 31 (2015) 54–63

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Robust tracking based on local structural cell graph ☆

Heng Fan [a], Jinhai Xiang [b,*], Honghong Liao [c], Xiaoping Du [d]

[a] College of Engineering, Huazhong Agricultural University, Wuhan 430070, China
[b] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[c] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[d] Key Laboratory of Digital Earth, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

ARTICLE INFO

Article history:
Received 30 October 2014
Accepted 26 May 2015
Available online 5 June 2015

Keywords:
Visual tracking
Local structural cell (LSC)
Local structural cell graph (LSCG)
Bayesian framework
Graph matching
Superpixel
Local based representation
Appearance model

ABSTRACT

Structure information has been increasingly incorporated into computer vision, however most trackers have ignored the inner spatial structure of the object. In this paper, we develop a simple yet robust tracking algorithm based on local structural cell graph (LSCG). This approach exploits both partial and spatial information of the target via representing the object with local structural cells (LSCs) and constructing a graph to model the spatial structure between the inner parts of the object. The tracking is formulated as matching LSCG, whose nodes are target parts and edges are the interaction between two parts. Within the Bayesian framework, we achieve object tracking by matching graphs between the reference and candidates. Eventually, the candidate with the highest similarity is the target. In addition, an updating strategy is adopted to help our tracker adapt to the fast time-varying object appearance. Experimental results demonstrate that the proposed method outperforms several state-of-the-art trackers.

1. Introduction

Visual tracking is an essential component of many applications in computer vision, such as surveillance, human–computer interaction and robotics [1]. For robust object tracking, many different methods have been proposed. Despite reasonably good results from these approaches, some common challenges remain for tracking targets through complex scenes, e.g., when objects undergo significant pose variations or other severe deformations, i.e., object pose variations accompanied with long-term object partial occlusions or object intersections. In order to handle these problems, a wide range of appearance models for tracking have been presented by researchers [2]. Roughly speaking, these appearance models can be categorized into three types: based on visual representation such as global feature-based representations [3–5,12,15] and local feature-based representations [6–8,16]; based on statistical modeling containing generative models [3,9–11,17] and discriminative models [12,13,15,16]; based on structure information including [14,18–20]. Although structure information has drawn increasing interest in object recognition [20] and object detection [21,22], much less attention is paid to it in visual tracking.

In this paper, we exploit effective and efficient structure information for object tracking. Object is firstly segmented into superpixels. Then we construct the structural appearance model based on the superpixel map of object. The structure information is generated by superimposing a rectangular grid on top of the superpixel map as shown in Fig.1(d). In the rectangular grid, each grid is represented by its center point, which is annotated with the feature vector of the covered superpixel. Thus we obtain an undirected graph **G**, whose nodes are the grid center points and edges are the interactions between the grid points. Further, taking the local relationship of the inner object parts into consideration, a novel approach is proposed to construct the local structural cell (LSC) for each grid point. We employ these local structural cells (LSCs) to substitute the grid points in **G** and obtain a new local structural cell graph $\mathcal{G}$, whose nodes are the LSCs and edges are the interaction between LSCs. Therefore, the appearances of object parts and their relations are embedded into the local structural cell graph (LSCG), and the tracking is viewed as matching LSCG in the subsequent frames. Within the Bayesian inference framework, we can track the object by the similarities of the local structural cell graphs (LSCGs) between the reference and candidate targets, and select the candidate with maximal similarity as the object. Meanwhile, an online updating mechanism is used to adapt our tracker to occlusions and deformations.

The contributions of this work are summarized as follows. Firstly, we propose a novel appearance model LSCG to represent

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author. Fax: +86 27 87286876.
E-mail addresses: hfan@webmail.hzau.edu.cn (H. Fan), jimmy_xiang@163.com (J. Xiang), hustliaohh@gmail.com (H. Liao), duxp@radi.ac.cn (X. Du).

http://dx.doi.org/10.1016/j.jvcir.2015.05.011
1047-3203/

the target, in which both the spatial and structural relationship is considered in the inner object parts. Secondly, a tracking method based on LSCG is proposed and implemented via the Bayesian framework. Finally, an intuitive and effective updating mechanism is introduced to improve robustness of the tracker in presence of appearance changes.

2. Related work

General tracking approaches can be categorized into either generative or discriminative models [2], however, only a few trackers take structure information into consideration. The discriminative methods regard tracking as a classification problem which aims to best separate the object from the ever-changing background. These methods employ both the foreground and background information. Avidan [23] proposes an ensemble tracker which treats tracking as a pixel-based binary classification problem. This method can distinguish target from background, however the pixel-based representation needs more computational resources and thereby limits its performance. In [7], Grabner et al. present an online boosting tracker to update discriminative features and further in [24] a semi-online method is proposed to handle drifting problem. Kalal et al. [25] introduce a P-N learning algorithm to learn effective features from positive and negative samples for object tracking. This tracking method nevertheless is prone to induce drifting problem when structure variations of object occur. Babenko et al. [12] utilize the multiple instance learning (MIL) method for visual tracking, which can alleviate drift to some extent. Whereas the MIL tracker may detect the positive sample that is less important because it does not consider the sample importance in its learning process. Further in [26], Zhang et al. propose the online weighted multiple instance learning (WMIL) by assigning weight to different samples in the process of training classifier. Nevertheless, the above methods undermine the robustness to occlusion and non-rigid distortion ignoring structure information.

The generative models formulate the tracking problem as searching for regions most similar to object. These methods are based on either subspace models [3] or templates [4,17]. To solve the problem of appearance variations caused by illumination or deformation, the appearance model is updated dynamically. In [3], the incremental visual tracking method suggests an online approach for efficiently learning and updating a low dimensional PCA subspace representation for the object. However, this PCA subspace based representation scheme is sensitive to partial occlusion. Adam et al. [4] present a fragment-based template model for visual tracking. Kwon et al. [17] decompose the appearance model into multiple basic observation models to cover a wide range of illumination and deformation. Similarly, the ignorance of structure information results in bad performance in deformation and occlusion.

Recently, several part-based tracking methods have been proposed [4,18,19]. In [4], the object is segmented into several fragments to construct appearance model, while this model does not consider the geometry relations between parts. Kwon et al. [18] use SIFT descriptor to generate parts, which is very unstable, and the tracking results usually consist of bad tracked parts. The tracker in [19] generates parts by oversegmenting the target into superpixels. This method formulates tracking task as figure/ground segmentation and models superpixels correspondence with CRF during the process of segmentation, whereas the high complexity of CRF constraints its further application in object tracking.

Another work similar to ours is [6], in which only a probability map of superpixels is constructed to distinguish the target from background without consideration of structure information, which is easy to cause tracking drift in color-similar background. In our work, we take structure information of target appearance into account, which enables our tracker to robustly track object in presence of occlusions and deformations.

3. Structural appearance model

In this section, the structural appearance model for tracking is presented. Section 3.1 introduces the process of constructing LSCG and the LSCG method is illustrated in Section 3.2.

3.1. Local structural cell graph (LSCG)

For the rectangular region R (target region) in one frame (See Fig. 1(a)), we firstly extract the surrounding region[1] of the target and segment it into superpixels via SLIC [28] (see Fig. 1(b) and (c)). A rectangular grid $\boldsymbol{\Theta}$ is then utilized to superimpose on the superpixel map of the target area $\mathbf{M}$ (see Fig. 1(d)), which segments object into several uniform blocks. In the rectangle grid, each grid block is denoted by its center point, which is related to the feature vector of the covered superpixel. Assume that the rectangular grid $\boldsymbol{\Theta}$ consists of $m \times n$ grid points, and the collection of these points is represented by $\boldsymbol{\Omega} = \{g_{ij} | i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\}$, where g_{ij} denotes grid point. Taking spatial relation between the grid points into account, we can construct an undirected graph $\mathbf{G} = (\boldsymbol{\Omega}, \mathbf{E})$, where $\boldsymbol{\Omega}$ is the node set and $\mathbf{E}$ represents the edge set between adjacent nodes respectively (see Fig. 1(e)). The $\mathbf{E}$ is defined as $\mathbf{E} = \left\{e_{(ij)(kl)} | (i-k)^2 + (j-l)^2 \leqslant r^2\right\}$, where r represents the distance (i.e., the radius of the cell node in local structural cell graph). For each node g_{ij} in $\mathbf{G}$, we select the associated superpixel from the superpixel map $\mathbf{M}$. Specifically, let $sp_{ij} = \mathbf{M}(p_{ij})$ represent the covered superpixel of the grid g_{ij} located at position p_{ij}. For g_{ij}, the feature vector f_{ij} of the superpixel sp_{ij} is used to represent its feature vector. The collection of these feature vectors is specified by $\mathbf{F} = \{f_{ij} | i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\}$.

In order to further exploit local spatial relationship between inner object parts, we define the local structural cell (LSC). For each node g_{ij} in $\mathbf{G}$, its LSC c_{ij} is represented by its node members $g^c_{ij} = \{g_{ij}\} \cup \{g_{hk} | (h-i)^2 + (k-j)^2 \leqslant r^2\}$, node member features $f^c_{ij} = \{f_{ij}\} \cup \{f_{hk} | (h-i)^2 + (k-j)^2 \leqslant r^2\}$ and local interactions of the node members (see Fig. 1(f)), where r is radius of the cell node. We utilize LSCs to replace the nodes in $\mathbf{G}$, and a new LSCG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is obtained (see Fig. 1(g)), where $\mathcal{V} = \left\{g^c_{ij} | i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\right\}$ is the cell node set and $\mathcal{E}$ is the edge set representing the interactions between cell nodes in $\mathcal{G}$. The spatial structure information of the target is thus encoded in the proposed appearance model through $\mathcal{G}$.

3.2. Similarity of local structural cell graph

In graph matching, a key aspect is how to measure the similarity between two undirected graphs. Two crucial problems exist in this process: (1) The instability of undirected graph and (2) the criteria for evaluating the correspondence between undirected graphs. In general, the object parts are represented by nodes in graph. As the target is non-rigid and deformable, the number of nodes is changeable. Even worse, some nodes in the graph may disappear, other new nodes may be generated and the spatial structure relation between nodes are mutable due to object appearance variations, which greatly constrains the application of graph model in computer vision. In addition, there are lack of effective criterias and methods to measure the similarity between graphs.

To solve the problems, we exploit both partial and spatial

[1] The surrounding region is a square area centered at the location of target X^c_t, and its side length is equal to $\lambda_s [X^t_s]^{\frac{1}{2}}$, where X^c_t represents the center location of target region X_t and X^s_t denotes its size. The parameter λ_s is a constant variable, which determines the size of this surrounding region.