Image and Vision Computing 69 (2018) 22–32

Contents lists available at ScienceDirect

Image and Vision Computing

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

Visual tracking based on edge field with object proposal association ☆

Junseok Kwon [a,*], Hansung Lee [b,*]

[a] School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea
[b] SW R&D Center, Samsung Electronics Co. Ltd., Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 22 November 2016
Received in revised form 10 July 2017
Accepted 10 November 2017
Available online 21 November 2017

Keywords:
Visual tracking
Object proposal

ABSTRACT

In this paper, we present a novel tracking system based on edge-based object proposal and data association called object proposal association. Our object proposal method accurately detects and localizes objects in an image by searching for object-like regions, with the assumption that an object is represented by a closed boundary. To search for closed boundaries in an image, we present a new Edge Fields (EFs) technique. Using this technique, our method can extract high-quality edges and can obtain accurate boundaries from the image. The EFs technique consists of blurring and thresholding steps, where the former helps extract high-quality edges and the latter prevents the method from losing image details while blurring. After the method extracts object-like regions, we associate the regions in the previous frame with those in the current frame. For this purpose, using the Markov chain Monte Carlo data association (MCMCDA) algorithm, we can find pairs of similar regions across two frames. Experimental results demonstrate that our object proposal method is competitive with state-of-the-art object proposal methods on the PASCAL VOC 2007 dataset. Our tracking method is also competitive with state-of-the-art tracking methods on Object Tracking Benchmark dataset.

1. Introduction

The goal of object detection is to determine whether an object exists in an image and, if it exists, to find the exact positions and scales of the objects. To accurately localize the objects, conventional object detection algorithms perform object classification at every location in an image. Recently, Dollar and Zitnick [1] consider only a set of bounding boxes of the object candidates instead of all locations in an image.

In this paper, to propose objects in an image, we follow the philosophy of Zitnick and Dollar [2] wherein an object is represented by a closed boundary. Closed boundaries can be found by searching for circularly connected edges in an image, where the edges provide simplified but useful information about the object shape. To extract accurate edges for a good object proposal, we propose a novel approach called Edge Fields (EFs) technique, based on the distribution fields algorithm [3]. In contrast with traditional blurring, which is applied only to a single image, our EFs technique first divides a single image into multiple ones by thresholding and then applies blurring to these multiple images separately. By doing this, the EFs technique can smooth the distribution of an energy function for edge extraction without destroying image details, as illustrated in the bottom of Fig. 2 (b).

Using the aforementioned object proposal technique, our visual tracker tracks a target accurately in real-world environments [4–8]. We associate similar object regions across consecutive frames by using the Markov chain Monte Carlo data association (MCMCDA) algorithm [9]. In Oh et al. [9], MCMCDA is used for solving a multi-target tracking problem. In this paper, however, we use MCMCDA for solving a single-target tracking problem. The idea of using MCMCDA for single-target tracking is to associate not only similar object regions but also similar background clutters across frames. By associating similar background clutters in advance, we can prevent an object region from being associated with the background clutter, This makes our visual tracker more robust with respect to background clutter that has a similar appearance to the target object. Please refer to the description in Fig. 6 for further explanation.

Our contribution is threefold:

- We present a novel edge-extraction method called the EFs technique.
- We develop a new object proposal system that shows a good performance in terms of both recall rate and speed. We exhaustively analyze and evaluate our object proposal method in the experiment.

☆ This paper has been recommended for acceptance by Ming-Hsuan Yang.
* Corresponding authors.
E-mail addresses: jskwon@cau.ac.kr (J. Kwon), hs7108.lee@samsung.com (H. Lee).

https://doi.org/10.1016/j.imavis.2017.11.004
0262-8856/

- We develop a new single-target tracking system that uses EFs and data association called Object Proposal Association (OPA).

The remainder of this paper is organized as follows: Related work is presented in Section 2. Section 3 describes the whole pipeline of our method. In Section 4, we explain how to extract object-like regions. The OPA algorithm is introduced in Section 5. Section 6 presents the proposed single-target tracking algorithm. The object proposal and tracking results are shown in Section 7. Finally, Section 8 concludes this study with discussion.

2. Related work

2.1. Object proposal

The human vision exploits a few characteristics of an object during the detection process. To extract object-like regions in an image, the aforementioned approach follows the characteristics of an object being detected by human vision. For example, Alexe et al. [10] used a characteristic of an object that describes it as typically containing a salient region, whereas in Cheng et al. and Zitnick and Dollar [2,11], the characteristic is that it is represented by a closed boundary. An approach called EdgeBox [2] generates object bounding boxes that include closed boundaries using the edge information. The approach called BING [11] learns the boundary shape of objects using training images and detects object boundaries with a machine learning technique. Based on predefined object characteristics, these two methods measure the objectness score of the found regions and leave only the reliable ones with a high score. In other approaches [12–16], object-like regions are found, which have high color contrast, by segmenting the foreground-background regions. Please refer to Hosang et al. [17] for a further survey of the object proposal methods.

This paper is different from our conference paper [18]. We extend the object proposal method in Kwon et al. [18] to the visual tracking method. To track a target, we combine the object proposal method with a data association algorithm. Moreover, we explain the method in Kwon et al. [18] in more technical details and extensively evaluate it with more test sequences.

2.2. Visual tracking

In tracking methods that use a data association technique, the Joint probabilistic Data Association (JPDA) method [19] was proposed to associate detected objects using a joint probabilistic score. The multi-hypothesis tracker (MHT) [20] is a data association method based on the Bayesian framework. JPDA and MHT suffer from combinatorial complexity as the number of targets increases. To handle this problem, Oh et al. [9] presented the MCMCDA algorithm, which employs a more principled JPDA approximation. All of the aforementioned data association methods are used for multi-target tracking problems. However, our method uses MCMCDA for single-target tracking problems.

Tracking by detection methods learn the discriminative appearance model of a target object at an initial frame and use it to detect the target objects in subsequent frames and to track the target. For example, the Struck method [21] detects the target with a structured support vector machine. The tracking-learning-detection (TLD) algorithm [22] can re-initialize the tracker when it fails by using the detector. Conventional tracking by detection methods, however, does not detect all object-like regions for visual tracking. On the other hand, our method detects all object-like regions in an image.

Song [23] proposed an efficient tracking algorithm that utilizes high-dimensional informative features. The features are represented by multi-scale and spatio-color vectors. Hence, the tracker is robust to scale variations and deformation of a target. Song et al. [24] also presented a novel tracker that adopts self-similarity learning. The self-similarity information among the local features boosted visual tracking accuracy.

Recently, an object tracking method with object proposals [25] was introduced. The method measures objectness scores and then selects the best object regions. The disadvantage of this method is that it does not associate object regions across frames. Thus, there is no explicit mechanism for ignoring background clutter. Zhu et al. [26] utilized Edge Box proposals and, based on the proposals, searched randomly moving objects globally across the entire frame. Liang et al. [27] proposed a novel tracker that uses BING objectness as a prior knowledge. They adapted the objectness measure to handles various tracking environments. Similar to the aforementioned trackers, our method also adopts the object proposals for visual tracking. Contrary to the trackers, however, we use a more robust object proposal algorithm, Edge Field, proposed by ours. Moreover, to deal with multiple object proposals efficiently, we employ data association techniques for a single target tracking problem, which are typically utilized in fields of multi-target tracking.

3. Pipeline of the proposed system

Our object tracking system consists of three steps, namely, object proposal step, data association step, and object tracking step.

The object proposal step, which is illustrated in Fig. 1 (a), in turn, contains four steps. In the object proposal step, given an input image, our system first extracts multiple features, in which a feature corresponds to a color channel of the image (Fig. 1 (a-1)). To get several color channels, we convert the image to several color spaces. Then, the system thresholds each color channel image at each intensity level (Fig. 1 (a-2)). In the next step, the thresholded images are blurred separately by using a Gaussian kernel (Fig. 1 (a-3)). The combination of the previous two steps is called Edge Fields. As a last step, our system retrieves the closed boundaries on the thresholded images (Fig. 1 (a-4)). The output of the object proposal step is object bounding boxes that compactly contain the closed boundaries obtained from the previous step.

In the data association step, our system finds pairs of similar objects across two consecutive frames, as shown in Fig. 1 (b). The object tracking step (Fig. 1 (c)) is required to determine which pair is for the target object among multiple pairs made by the data association step. For this, the system compares the image patches in each pair with that of the reference target and selects the best pair.

4. Object proposal step

An edge-extraction method has its own objective function, which measures a score on the degree of edgeness of a candidate edge. In this context, the goal of the method is to find the best edge that gives the highest score. This goal can be achieved by optimizing the objective function and finding the global optimum (red circles in Fig. 2 (a)) in an energy landscape related to the objective function. Because the landscape is typically very rough (left curve in Fig. 2 (a)), however, the method easily gets trapped in local optima (green circles in Fig. 2 (a)). One promising solution to this problem is to smooth the landscape from the left curve to the right one in Fig. 2 (a), where all the local optima are removed. However, this solution has a problem in that a blurred image significantly loses the details of the object, as shown in the top of Fig. 2 (b).

4.1. Feature extraction

Fig. 3 shows color channel images, which are used as features. We employ HSV, YUV, and LAB color spaces to get several color channel