Optik 125 (2014) 2708–2712

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Contour level object detection with top-down information

Huapeng Yu[a,b,c,*], Yongxin Chang[a,b,c], Pei Lu[a,b,c], Zhiyong Xu[a], Chengyu Fu[a], Yafei Wang[b]

[a] Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China
[b] School of Optoelectronic Information, University of Electronic Science and Technology of China, Chengdu 610054, China
[c] Graduate University of Chinese Academy of Sciences, Beijing 100039, China

ARTICLE INFO

Article history:
Received 5 June 2013
Accepted 5 November 2013

Keywords:
Contour
Object detection
Top-down information
Salient closed contour

ABSTRACT

This paper presents a contour level object detection approach. In contrast to conventional bounding box results, we give out the salient closed contour of the object, which provides a possibility of semantic analysis for the object. We get the salient closed contour with Ratio Contour algorithm. The top-down information needed by salient closed contour extraction is based on the well-known Bag-of-Features methodology. Our top-down information based contour extraction and completion is much more efficient and robust than many related approaches lack of the top-down information. We also propose a novel post-processing framework for object detection. With low threshold and a refined binary classifier, we can get stable high performance. We evaluate our approaches on UIUC cars dataset. We show that our approaches apparently improve the performance of object detections under clutter.

1. Introduction

Object detection under clutter is a challenging task for computer vision. Conventional approaches only give out the bounding box of the object, which has little help for in-depth semantic analysis of the object. To give out the contour, especially the closed contour, of the object is an interesting idea for this. Closed contour contains much more information of the object (like shape) than bounding box, which provides a possibility of semantic analysis for the object. However, to extract and complete the closed contour of the object is non-trivial. In fact, it lies in the arena of segmentation, which is a well-known unresolved problem. We believe that top-down information of the object is critical for this task. With the top-down information of the object, like location and scale, the problem of extracting and completing the contour of the object can be reduced into a salient closed contour extraction problem, which can be resolved with the Ratio Contour algorithm [8] in polynomial time. But how can we get the top-down information of the object? Fortunately, Bag-of-Features methodology [9] proposed within recent years provides a good solution. With learned feature patches, we obtain a detection map pyramid of the image, which gives out the possible object location and scale.

Conventional post-processing of object detection involves thresholding and non-maximal suppression (NMS) [1,5,10]. Sequential thresholds are used to obtain the Precision/Recall curve, while NMS gives out the possible object location under a certain scale. NMS verifies the possible object location has the maximal response within the predefined neighborhood size. Although simple, empirically obtaining the best threshold is not a good solution. Also NMS cannot avoid multiple detections for the same object. We propose a novel post-processing framework for object detection. With low threshold, we assure low true negative rate. With sequential detection based on elimination, we can lower the rate of multiple detections for the same object. With a refined binary classifier, we can effectively lower the false positive rate.

2. Related work

BOF scheme [9] is originally inspired by statistical language modeling. At training stage, a feature detector (e.g. SIFT [11]) is applied to the bounding box labeling of each positive training image to get a set of feature points, each of which is described by a feature descriptor. Then these feature points are clustered into N features, which we call a bag-of-features. With BOF we train a linear classifier on the positive and negative training samples. At test stage, the same process is applied to get the feature descriptor of each sub-window of the test image. The final detection result is given out by the linear classifier.

HMAX model [1–4], which is originally a primary vision model, together with a linear classifier is a typical BOF scheme.

* Corresponding author at: The 5th Lab, Institute of Optics and Electronics, Chinese Academy of Sciences, P.O. Box 350, Shuangliu, Chengdu 610209, Sichuan Province, China. Tel.: +86 28 85101485; fax: +86 28 85100966.
E-mail address: yuhuapeng@uestc.edu.cn (H. Yu).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.11.029

In our work, we make use of BOF scheme, particularly the HMAX model together with a linear classifier, to accomplish the bounding box level object detection.

Contour extraction is a traditional research area of computer vision. From the classical Canny [12] to Pb [13], a set of relative mature theories and approaches based on gradient have been set up. An alternative for contour extraction is the Non-classical Receptive Field (NCRF) [14] which is a biological vision model.

On the contrary, contour completion, particularly the closed contour extraction is a more challenging task. Ratio Contour algorithm [8] tries to solve the salient closed contour extraction based on Gestalt grouping principles. Its default low level contour extractor is Canny edge detector. However, pure salient closed contour extraction is insufficient for contour level object detection because for a typical image under clutter the most salient one may not be the object in care. To deal with this problem, Zhiqi Zhang etc. [15] propose combining appearance and shape to do contour level object detection. At first they extract the contours of the whole image with Pb. Then they make use of BOF scheme to weight the loss of each contour segment, based on which Ratio Contour algorithm is used to form the closed contour of the object. Our work also makes use of BOF scheme, however we solve the problem in two stages. At first we obtain the bounding box of the object, which we call the top-down information of the object. Then we do salient closed contour extraction in the bounding box, which naturally is the closed contour of the object. To improve the precision of object closed contour detection, at the second stage, we can also again make use of the detection map of BOF scheme just like what Zhiqi Zhang does.

Classical post-processing of object detection [1,5,10] contains two steps: thresholding and localization. Thresholding the detection map can ultimately obtain the precision/recall curve and the best empirical threshold. Localization can be done with elimination or NMS [5]. Elimination can be done within the detection map or raw input image. For part-based model [5,10], repeated part elimination algorithm can be adopted. In our work, we analyze the limitation of the classical post-processing of object detection and propose a novel post-processing framework based on low threshold and sequential detection algorithm.

3. Contour level object detection

Traditional object detection only gives out the bounding box of the object, which has little help for in-depth processing, like semantic analysis of the object. On contrary, contour level object detection can give out the contour, especially the closed contour, of the object.

However, generic contour extraction and completion encounters performance bottleneck due to lack of top-down information, like the location and scale of the object. We propose a contour level object detection approach based on the top-down information of the object, which is acquired through the well-known Beg-of-Features methodology.

3.1. Formally discussion

Classical sliding window based object detection approach can be formally described by Eqs. (1)–(3).

$$f : X \to Y \tag{1}$$

$$f(x) = w^T x + b \tag{2}$$

$$f(x)_{(i)} = w^T x_{(i)} + b, \quad i = 1, \ldots, n \tag{3}$$

In Eq. (1), X is the input, Y is the binary classifier output, typically $Y \in \{-1,1\}$. Eq. (2) depicts a linear classifier. Note that in Eq. (2), here $f(x)$ represents the real-valued raw output of the binary classifier, namely the detection map. In the detection map, pixel with value greater than 0 is regarded as object point, while less than 0 background point. Eq. (3) depicts the multi-scale version, namely the detection map pyramid. In Eq. (3), n represents the scale number. To train the binary classifier $f : X \to Y$, the well-known Bag-of-Features methodology is a good choice. We will present a typical Bag-of-Features object detection scheme in 3.2.

Traditional object detection gives out the bounding box of the object, which is depicted by Eq. (4).

$$B = arg \max_{i,[row,col]} \{f(x)_{(i)}\}_{i=1,\ldots,n} \tag{4}$$

Within Eq. (4), n represents the scale number, [row, col] represents the location of the maximal response, while B represents the bounding box determined by the scale and location of the maximal response.

Correspondingly, contour level object detection can be depicted by Eq. (5).

$$C = arg \underset{i,[(r1,c1)\ldots]}{thres} \{f(x)_{(i)}\}_{i=1,\ldots,n} \tag{5}$$

Eq. (5) differs from Eq. (4) in several aspects. At first, a thresholding operation replaced the maximal one in Eq. (4). The boundary of the pixels with the value greater than a pre-specified threshold forms the contour, which is represented by [($r1$, $c1$)...] in Eq. (5). Apparently, contour extraction based on a fixed threshold segmentation on detection map is arbitrary. Detection map discards too much information. Also fixed threshold segmentation is inappropriate.

In fact, state-of-the-art contour extractor (like Pb) relies on gradient computation of the raw input image under multiple scales. Based on these work and the top-down information provided by Eq. (4), we are now ready to reconsider the problem of contour level object detection.

Assume we get a set of contours of the bounding box B with contour extractor (like Pb), Ratio Contour algorithm find the most salient closed contour through optimizing a boundary cost function, which encode the Gestalt laws of proximity and continuity. The salient closed contour in the bounding box B is most probably the closed contour of the object. Through encoding the information of the detection map (bounded by B) defined by Eq. (2) or (3) into the contours, we can expect to obtain a more accurate closed contour of the object. Eq. (6) depicts these ideas.

$$C_{opt} = arg\min_{C} \phi(C) \tag{6}$$

Within Eq. (6), C represents all the candidate closed contours, the optimal closed contour C_{opt} is the one with minimal cost. The cost object function $\phi(C)$ may encode both the contour and region (bounded by the contour) information [15].

3.2. A typical BOF scheme

In this paper, we will make use of a typical Bag-of-Features object detection scheme, namely HMAX model [1–4] together with a linear classifier. HMAX model is a hierarchical feature computation model, the ultimate output of which is a feature vector (the same dimensions as the learned bag-of-features). Then the feature vector, namely C2, is used to train a linear classifier depicted by Eq. (2) or (3). During test, the procedure is similar, at first to get C2, then C2 is used by the linear classifier to classify the input image.

Unlike popular BOF scheme which makes use of local feature point extractor (like SIFT) and clustering to get the bag-of-features, typical HMAX implementation just randomly selects feature patches from the training images to form it. Details about this and our improvement could be found in our earlier papers [6,7].