Object Detection using Contour and Structure Information

Guangwei Wang, Zenggang Xiong*, Conghuan Ye, Xuemin Zhang, Suo Hu

School of Computer and Information Science
Hubei Engineering University
Xiaogan, Hubei, China

Abstract — In this paper, we propose an approach of object part contour features combination associated with global spatial structure for object detection. A set of distinctive object parts are constructed from a set of sample images of the object class; objects are then represented using object parts, together with global spatial relations observed among the parts. The object part contour is described using object contour direction histogram method to model the variant characters in different shapes. In detection process, a group of object part classifiers are trained to detect the object parts. Then, we apply a generalized Hough voting scheme based on global spatial structure to generate object locations and scales. We evaluate the proposed approach on ETHZ object test set. Experimental results show that the proposed approach is efficient and robust in object detection.

Keywords - object detection; contour features; object structure; feature descriptor

I. INTRODUCTION

In this paper we consider the problem of detecting and localizing object of a generic category, such as horse or apple in static images. This is a difficult problem because objects in a category can vary greatly in shape and appearance. Variation arise not only from changes in illumination, occlusion, background clutter and view point, but also due to non-rigid deformations, and intra-class variation in shape and other visual properties among objects in a rich category.

How do we deal with the variation, especial the intra-class and pose variability of object? There are several possibilities to represent object classes. A star shape model [1, 2] can be easily trained and evaluated in contrast to the constellation model or complex graphical models [3, 4, 5, 6]. It allows using as many parts as required, since the complexity scales linearly. Moreover, this model is flexible enough to deal with large variations in object shape and appearance of rigid and articulated structures.

Another way is to develop a feature set [4, 7, 8] that was robust to the variable local shape and the wide range of poses. Lowe proposed a scale invariant feature transform (SIFT), which combined a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions [9, 10, 11]. Geometric histogram and shape context computed a histogram describing the edge distribution in a region. These descriptors were successfully used, for example, for shape recognition of drawings for which edges were reliable features. HOG [5] was similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but its difference with the others in that it was based on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. Shatton et al. and Opelt et al. [12, 15] used boundary fragments to represent an object and they used boundary matching method to detect an object.

Our approach has two methods to deal with the variation of object, both local and global. Firstly, multi-cues of object part have been incorporated as a key component of local features. In this paper, we introduce EDH descriptor [10, 11] for shape information to describe the object part. Secondly, we present a hierarchical object representation method to describe object structure. In addition, a voting method based on object part to detect object is presented. The voting method codes the global geometry of generic visual object categories with spatial relations linking object center to object part.

The proposed framework can be applied to any object that consists of distinguishable parts arranged in a relatively fixed spatial configuration. Our experiments are performed on images of ETHZ object set will be used as a running example throughout the paper to illustrate the ideas and techniques involved.

The rest of the paper is organized as follow. Section 2 presents the hierarchical object structure, object part classifiers and voting detection algorithm. In section 3, experiments on real images show that the feature descriptor is effective for object categorization.

II. OBJECT REPRESENTATION AND DETECTION

A. Hierarchical Object Structure

The object representation method is a hierarchical model that codes the global geometry and local appearance of generic visual object categories with spatial relations linking object (top level) to parts (second level), and local feature cues linking sub-pattern (second level) and local feature class (third level). See Figure 1 (a), the object is at the top level, the second and third level is the sub-patterns and local feature descriptors of object parts respectively.
The spatial relations between object parts can be described by global spatial structure [17, 18, 19, 20]. The main idea of global spatial structure is to represent the object as a collection of parts and have each part cast votes in a discrete voting-space [7]. Here, we use an object structure $G$ to describe an object category. The object that consists of $n$ part $p_i$ can be defined by the equation (1).

$$G = \bigcup_{i=1}^{n} p_i(x, r, dc, w, \phi(x, r))$$

(1)

Where $x$ is a two-dimensional vector specifying an “anchor” position for object part $p_i$ relative to the object position; $r$ represent the scale of the object part; $dc$ is a deviation vector; $w$ shows the discriminative weight of the object part. $\phi(x, r)$ denote a feature vector for the object part $p_i$.

B. Object Part Classifier

We use contour shape information as a key component for object detection. Since edge points are related to shape information closely, edge direction histogram (EDH) is a very simple and direct way to characterize shape information of an object. EDH is computed by grouping the edge pixels which fall into edge directions and counting the number of pixels in each direction.

Here, we adapt the methods of [6] and [10] for describing object part contour features. An object part is subdivided evenly its bounding box into a $n \times n$ grid. Each grid encodes information only inside the part. We capture different cues from the cells, and each type of cues is encoded by concatenating grid signal into a histogram. In this paper, we consider contour shape cues.

We firstly train the Contour Feature SVM classifiers $C_f = \{f_1, f_2, \ldots, f_j\}$ for object parts $\{p_1, p_2, \ldots, p_n\}$. Take a classifier for example, given a set of training image windows labeled as positive (object) or negative (nonobjective), each image window is converted into a feature vector as described above. These vectors are then fed as input to a supervised learning algorithm that learns to classify an image window as member or nonmember of the object pattern. In our experiments, we chose linear SVM as classifier [16].

Sliding window classification [4] is a simple, effective technique for object detection. The detection problem is to determine whether the query image contains object part instances and where it is. The classifier $C_f$ is applied to fix-sized windows at various locations in the feature pyramid, each window being represented as a feature vector $f(x, l)$, where $x$ specifies the position of the window in the image, and $l$ specifies the level of the image in pyramid. The equation (2) represents the classifier $C_f$ at one of the sliding windows.

$$t_{p,j} = C_f(f(x, l))$$

(2)

If $t_{p,j} > q$, then $h_{p,j} = (x, l)$, and it is a hypothesis position of object part $p_i$. Otherwise, it is not. Let $H = \{H_{p_1}, H_{p_2}, \ldots, H_{p_n}\}$ denote the collection of all parts hypothesis position. The architecture of object detection is presented in Figure. 2. We use the classifier $C_f$ to build a hypothesis set $H$.

C. Voting for Object Detection

The goal of this section is to generate bounding boxes of that category in the image. To achieve it, we combine the hypothesis location set $H$ and global spatial structure $G$.

The object is identified by the conglomeration of votes in a small neighborhood of the voting space $V$. $V$ is typically defined as the sum of independent votes from each part. The final Hough image is computed by combining the votes from all parts for the image in question.

The architecture of object detection is presented in Fig. 2. The first step is to generate the voting score $s_i$ at a centroid circle location of the object structure $G$ by applying a transformation $T(\cdot)$ to $h_{p_i}$ of $G$. The transformation is a voting procedure, which is characterized by equation (3).

$$s_i(c_i) = T(h_{p_i}, G)$$

(3)
Where $c_i$ is the voting location of part $p_i$. The transformation $T$ exploits the rough localization provided by the spatial relationship between the parts and the object.

Next, we want to find the local max voting score $S_G$ in the Hough image. The object location is obtained by choosing local maxim in the Hough image that has responses above a certain threshold. The $S_{det}(G)$ can be obtained equation (4).

$$S_{det}(G) = \sum_{i=1}^{n} (s_i(c_i) \cdot w_i - d_i)$$

(4)

The overall detection score $S_{det}(G)$ for object $G$ is a combination of the detected part voting score in the domain $D_0$. Where $w_i$ is the discriminative weight of part $p_i$, $d_i$ is the deviation of part from the optimal position. If $S_{det}(G) > h$, $D_0$ is a centroid of detected object structure $\hat{G}$. Otherwise, it is not.

<table>
<thead>
<tr>
<th>Apple</th>
<th>Bottle</th>
<th>Giraffe</th>
<th>Mug</th>
<th>Swan</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>88.6</td>
<td>83.4</td>
<td>83.9</td>
<td>83.5</td>
<td>87.5</td>
<td>85.4</td>
</tr>
<tr>
<td>Fritz</td>
<td>89.9</td>
<td>76.8</td>
<td>90.5</td>
<td>82.7</td>
<td>84.0</td>
</tr>
<tr>
<td>Ferrari</td>
<td>83.2</td>
<td>83.2</td>
<td>58.5</td>
<td>83.6</td>
<td>75.4</td>
</tr>
<tr>
<td>85.4</td>
<td>84.8</td>
<td>76.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The performance of comparison of detection performance with [8] on the ETHZ shape data set at 0.4 FPPI.

We compare to [8] and [9] on the ETHZ shape database. Experiments are conducted in 5-fold cross-validation. We split the entire set into half training and half test for each category, and average performance from 5 random splits is reported. This is consistent with the implementation in [8, 9] which reported the state-of-the art detection performance on this data set. Table 1 shows our comparison to [8] on each of the categories. Average over all categories we improve the performance of [8] by 8.6% to 85.4%. On apple logos, giraffes and swans, we improve the performance by 5.4, 25.3 and 12.1% respectively. On bottles and mugs, our approach performs comparable. We account the performance on the bottles and mugs to the shape which is less discriminant with respect to the background. As the data set was designed to test shape-based approaches, the improvements obtained by our approach underlines the versatility and adaptively of the hierarchical representation.

IV. Conclusion

In this paper, we propose an approach of contour features combination associated with object structure information for object detection. This method effectively describes the object part shape and structure features, and trains a group of object part SVM classifier. Based on the feature descriptors and part detection classifier, Hough voting method was used to detect object. Extensive object detection experiments show high detection rates with relatively low numbers of false detections. These results illustrate the high discriminant power of the shape and structure information.

The proposed framework can be applied to any object category that consists of distinguishable parts arranged in a relatively fixed spatial configuration. In summary, the results show that the object representation using hierarchical model and contour features is general to different kinds of object classes, and the contour feature descriptor are efficient to extract informative shape features for object detection.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Funds of China (No. 61370092), Hubei Provincial Department of Education Outstanding Youth Scientific Innovation Team Support Foundation (T201410), and Hubei Engineering University Teaching Research Project (Grant No. 2015A38).

REFERENCES


