Edge Fields for Robust Object Proposal

Junseok Kwon, Andrii Grygoriev, Yusun Lim, Youngki Hong, and Hansung Lee

SW R&D Center, Samsung Electronics, Co. Ltd., Suwon, Rep. of Korea
{js7.kwon,a.grigoryev,yusun.lim,youngki2002.hong,hs7108.lee}@samsung.com

Abstract. In this paper, we present a novel object proposal approach that can accurately detect and localize objects in an image. Our method searches for objects with the assumption that an object can be represented by a closed boundary. To search for closed boundaries in an image, the method employs edge features with the proposed Edge Fields (EFs) technique. With the EFs, our method can extract high quality of edges and can obtain good boundaries from the image. EFs consists of blurring and thresholding, in which blurring helps extract high quality of edges and thresholding prevents the method from losing image details during the blurring process. Experimental results demonstrate that our method is competitive with the state-of-the-art object proposal methods on the PASCAL VOC 2007 dataset.

1 Introduction

Object detection aims to determine whether an object exists in an image and to find exact positions and scales of the objects. In conventional approaches, to localize the objects, object classification is performed at every location in an image. Given large size images, however, this sliding window based approaches [10] inevitably undergo the difficulty with the high computational cost.

To solve this problem, a new framework for object detection has been proposed recently. In this framework (e.g. [6]), a set of bounding boxes of object candidates are proposed. Because the number of object proposals is typically very smaller than the number of all locations in an image, a search space for object classification can be drastically reduced, which makes the computational cost lower. With this reason, several state-of-the-art object detection algorithms employ the object proposal based approaches and demonstrate their efficiency on the PASCAL VOC dataset [8].

The human vision exploits a few characteristics of objects during the detection process. To propose object candidates in an image, the aforementioned approach follows the manner of the object being detected by human vision. For example, [1] used a characteristic that objects typically contain a salient region, while objects include a closed boundary in [5, 16]. An approach called EdgeBox [16] generates object bounding box proposals by finding the the closed

* Corresponding author
The closed boundary can be found by searching for circularly connected edges in an image, where edges provide a simplified but useful information on the object shape. However, edges extracted by conventional algorithms [6, 16] are not good enough for the object proposal when there are severe background clutters and occlusions. The edge extraction method has its own objective function, which measures the score on the degree of edgeness of a candidate edge. Then the goal of this 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.1(a)) in optimization landscape. Because the optimization landscape is typically very rough (left curve in Fig.1(a)), however, the method easily gets trapped in local optima (green circles in Fig.1(a)). One of promising solution to this problem is to smooth the optimization landscape like the right curve in Fig.1(a), in which all local optima are removed. This solution also has
To extract robust edges for the object proposal, we propose a novel approach called Edge Fields (EFs), which is based on the distribution fields algorithm [13]. In contrast with traditional blurring that applies to only a single image, EFs applies blurring to multiple images layered by thresholding. By doing this, EFs can smooth optimization landscape without destroying image details, as shown in the bottom of Fig.1(b). Please refer to description in Fig.1 for further explanation.

Our contributions are three-folds. The first contribution is to present a novel edge extraction method called Edge Fields. The second is to develop a new object proposal system that shows the good performance in terms of both the recall rate and the speed. The last is to exhaustively analyze and evaluate our object proposal system in the experiment.

2 Object Proposal Algorithm

2.1 Pipeline

Our object proposal system consists of four steps. Give an input image, our system first extracts multiple features, in which a feature corresponds to a color channel of the image (Fig.2(b)). 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.2(c)). In the next step, thresholded images are blurred separately by using a Gaussian kernel (Fig.2(d)). The combination of previous two steps is called Edge Fields. As a last step, our system retrieves closed boundaries on the thresholded images (Fig.2(e)). The output of our system is object bounding boxes that compactly contain closed boundaries obtained from the previous step (Fig.2(f)).

2.2 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 images. For example, hue, saturation, and value channel images are obtained from the HSV color
space. Luma and two chrominance channel images from the YUV color space. Lightness and two chrominance channel images from the LAB color space. We ignore Y and L channel because they are same with the value channel of the HSV color space. All color channels are then normalized to have a value range from 0 to 255. Because our color channels have complementary relationship each other, as shown in Fig.3 (b)-(h), we use as more channels as possible to get further information on objects.

2.3 Edge Fields Construction

Our method thresholds the aforementioned color channel images at each intensity level, as shown in Fig.4:

$$I_T(i, j, k) = \begin{cases} 1 & \text{if } I(i, j) \geq k; \\ 0 & \text{otherwise}, \end{cases}$$

where $i$ and $j$ indicate $x$ and $y$ positions of the image $I$, and $k$ denotes the possible intensity value of a pixel. For efficiency, we threshold an image at regular intervals other than at all intensity levels, where the interval size is fixed to 15 in the experiment. Notably a set of thresholded images contain exactly the same information as the original image. The difference is that the information in the original image is spread into multiple images layered by thresholding.

In blurring, each thresholded image is convolved with a Gaussian filter $G_\sigma$:

$$I_S(k) = I_T(k) * G_\sigma,$$
where the variance $\sigma$ is set to 3 in the experiment. Because there has been no mixing of pixel values across different thresholded images, all information has been preserved about the original image. Notably the convolution introduces positional uncertainty of a pixel and thus it makes the optimization landscape smooth.

### 2.4 Contour Detection

Our method retrieves closed boundaries on the blurred images obtained by previous section, as shown in Fig.5. To find the closed boundaries, we adopt a contour detection algorithm proposed in [14]. Our method produces bounding boxes as outputs, which compactly contain boundaries found by the aforementioned algorithm.

### 2.5 Filtering and Scoring Processes

The initial outputs contain many duplicated bounding boxes. By filtering out the duplicated ones, our method significantly reduces the number of outputs and leaves only reliable object bounding boxes that need to be further analyzed (e.g. classification). For this, we use a naive filtering algorithm that checks IoU (Intersection of Union) of two bounding boxes and removes one of bounding boxes (i.e. bounding box that has a lower score) if IoU$^1$ is lower than some threshold. In the experiment, the threshold is set to 0.95.

Our method also sorts the bounding boxes in descending order. We assume that an object is not so small and has a square shape. Hence we design a score function $S(B)$, which measures objectness using the size and the aspect ratio of a bounding box:

$$S(B) = \min(B_w, B_h) \times \frac{\min(B_w, B_h)}{\max(B_w, B_h)}$$

where $B_w$ and $B_h$ denote width and height of the bounding box $B$.

---

$^1$ IoU$(B_1, B_2) = \frac{A(B_1 \cap B_2)}{A(B_1) \cup A(B_2)}$, where $A(B_1)$ and $A(B_2)$ denote areas of two bounding boxes, respectively.
3 Experiments

Our method (EdgeFields) is compared with 7 object proposal algorithms, BING [5], CPMC [4], EdgeBoxes [16], MCG [3], Objectness [2, 1], Randomized-Prims [11], and SelectiveSearch [15]. We evaluate all algorithms on PASCAL VOC 2007 dataset [8] with the recall rate metric [9]. The parameters of our method is fixed through all experiments and parameters of other algorithms are adjusted to show the best performance.

3.1 Analysis on Our Method

As explained in previous section, our system processes several steps before proposing object bounding boxes. Hence we analyzed which step most contributes to the whole system.

**Feature Extraction** We evaluated performance of our method as the number of features used in the method varies. As reported in Fig.6(a), recall rate increases as the number of features increases. This is because the method can exploit richer information from more features.

**Thresholding** Fig.6(b) illustrates the performance graph of our method according to different thresholding step sizes. As shown in the figure, a recall rate is high when the thresholding step size is small, which means that more details are preserved with a smaller step size.
Blurring  We analyzed performance of our method with different degrees of blurring. Fig.7(a) shows that the highest recall rate is achieved when we properly blur the image. If we blur the image too much, many image details can be lost, which results in a low recall rate. If we weakly blur the image, however, the optimization landscape is not smooth enough, which also produces a low recall rate. Moreover the weak blurring makes our system slow because the landscape is very rough and thus the system has to spend much time to find good edges, as illustrated in Fig.7(b).

Filtering and Scoring Fig.8(a) illustrates recall rates when we use the filtering function or not. As demonstrated in the figure, even a naive filtering function can dramatically improve the accuracy of an object proposal method. A score function also improves the accuracy of an object proposal method, as shown in Fig.8(b).

3.2 Comparison with Other Methods

Fig.9 includes quantitative results of our method and other object proposal methods. EdgeBox shows the best performance in terms of recall rate. Our method is the second best. In terms of speed, however, our method is faster than EdgeBox. To conclude, our method is competitive with EdgeBox. Fig.9 also contains qualitative results of our method, in which red and blue denote bounding boxes of our method and Ground Truth, respectively. Our method detects objects on input images, even though the objects were very small or occluded by other objects. Because our method uses edge features for the object proposal, the method can accurately localize the objects, as shown in Fig.9.

4 Discussion

We can estimate the upper bound of recall rate. For example, Fig.10(a) shows recall rate of Ground Truth. No object proposal algorithms can produce a higher recall rate than that of Ground Truth. Fig.10(a) also provides useful information...
that proposing a 1 object bounding box yields only 40% recall rate and we need at least 15 object bounding boxes to get 100% recall rate.

Fig.10(b) show the upper bound of recall rate. For example, the best recall rate of our method is about 80% recall rate with 289 bounding boxes. If we design a better scoring function, our method can achieve 80% recall rate even with 100 bounding boxes.

5 Conclusion

In this paper, we propose a novel object proposal algorithm based on Edge Fields. By using Edge Fields, our method robustly extracts object edges and accurately
localized the objects. In the experiments, we achieve the good performance in terms of recall rate and speed.

References