OBJECT DETECTION USING HIERARCHICAL GRAPH-BASED SEGMENTATION

Jungho Kim†*, Byeongho Choi*† *and In-So Kweon*‡

†Multimedia IP Research Center, KETI
‡Robotics and Computer Vision Lab., KAIST

ABSTRACT

Object detection in real images or videos is challenging because the shapes and sizes of objects vary significantly according to their poses, camera viewing direction, and partial occlusion. Previous detection methods employ sliding-window-based schemes that scan windows across an image, requiring many differently shaped windows to capture shape and size variation. In order to solve this problem, we propose an object detection method using hierarchical graph-based segmentation: color-consistent parts are obtained by part-level segmentation and category-consistent regions are found using object-level segmentation. Thus we can avoid scanning a lot of windows across whole images by using part-level segmentation and robustly detect the objects of various shapes and sizes by using object-level segmentation. In addition, we evaluate detection performance using various classifiers with our detection approach.

***Index Terms*—** Object detection, Object classification, Graph-based segmentation.

1. INTRODUCTION

Object detection and classification have been long-lasting research topics in computer vision. For instance, the authors of [1] analyzed various feature sets for human detection. Viola and Jones proposed an efficient approach for face detection using image features obtained by the integral image and a technique for feature selection based on AdaBoost [2].

In order to detect the objects of interest, sliding-window-based detection methods have been proposed [1][2]. These methods create differently sized windows that are scanned across whole images in order to compute features as the inputs to a classifier, such as a support vector machine (SVM) or a boosting classifier. However, the shapes and sizes of objects in the images vary significantly depending on their poses and camera viewing direction, and partial occlusion. Fig. 1 shows cars of different sizes in a single image with different poses and partial occlusion. In order to detect each car, we need a lot of differently shaped windows scanned across a whole image.

In contrast to sliding-window-based schemes, approaches for image segmentation have been studied in order to find color-consistent segments in which pixels in each segment are likely to belong to the same object [3]. Because a single object is generally composed of several segments, an additional segmentation approach that assigns the same label to segmented parts belonging to one object is required for object detection.

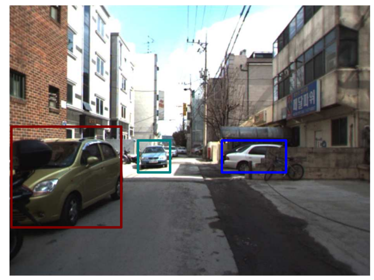


Fig. 1. An example of an image that contains three differently-shaped cars; the differences in shapes are contributed by pose variations and partial occlusion

In [4], the authors proposed category-level object segmentation by using a bag-of-words recognition component to segment objects with a Dirichlet process to determine the number of objects. Zhe and Davis proposed a Bayesian approach to human detection and segmentation combining local part-based and global template-based schemes [5]. In [6], the authors presented a hierarchical model for joint object detection and image segmentation.

In this paper we present an efficient object detection approach based on hierarchical graph-based segmentation, as shown in Fig. 2. We first perform part-level graph-based segmentation from image pixels in order to obtain color-consistent parts using the efficient graph-based segmentation approach proposed in [3]. Using part-level segmentation, we avoid scanning many different windows across all possible locations. However, part-level segmentation provides only color-consistent segments rather than object-level segments. Thus, we hierarchically combine an object-level segmentation approach by constructing a graph whose vertices are segmented parts from part-level segmentation and edges are object-level confidences computed by using a previously trained classifier.

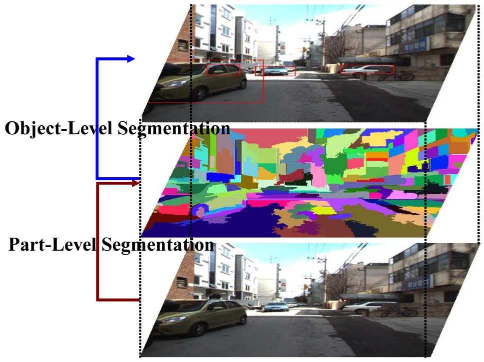


Fig. 2. Our object detection approach using hierarchical graph-based segmentation consisting of part-level and object-level segmentation

2. HIERARCHICAL GRAPH-BASED SEGMENTATION

2.1. Part-Level Segmentation

For part-level segmentation, we use an efficient graph-based segmentation method proposed in [3] and briefly review this segmentation approach here. Given a graph $G = (V,E)$ with vertices $v_i \in V$, a set of elements such as pixels, and edges $(v_i, v_j) \in E$ corresponding to pairs of neighboring vertices, graph-based segmentation is aiming at dividing an image into a set of segments.

For this purpose, we initially construct a graph whose vertices are all pixels and edges are the color differences between two neighboring vertices, v_i and v_j, as $w_p((v_i,v_j)) = \|I(v_i) - I(v_j)\|$. Starting with a segmentation in which each vertex v_i is in its own component C_i, and sorting edges by non-decreasing edge weight, w_p, the graph-based segmentation approach defines the region comparison predicate $D_p(C_1,C_2)$ between two components C_1 and C_2 by checking if the difference between the components $Dif(C_1,C_2)$ is large relative to the internal difference within at least one of components, $Int(C_1)$ and $Int(C_2)$, as shown in Eq. (1). Then we merge two components when $D_p(C_1,C_2)$ is false.

$$D_p(C_1,C_2) = \begin{cases} \text{true} & \text{if } Dif(C_1,C_2) > MInt(C_1,C_2) \\ \text{false} & \text{otherwise} \end{cases} \quad (1)$$

where

$$\begin{aligned} Dif(C_1,C_2) &= \min_{v_i \in C_1, v_j \in C_2, (v_i,v_j) \in E} w_p((v_i,v_j)), \\ MInt(C_1,C_2) &= \min(Int(C_1) + \ (C_1), Int(C_2) + \ (C_2)), \\ Int(C) &= \max_{e \in MST(C,E)} w_p(e), \end{aligned} \quad (2)$$

Where $MST(C,E)$ is the minimum spanning tree of the component C, and $(C) = k/|C|$. $|C|$ is the size of components and k is a constant that controls the merging conditions.

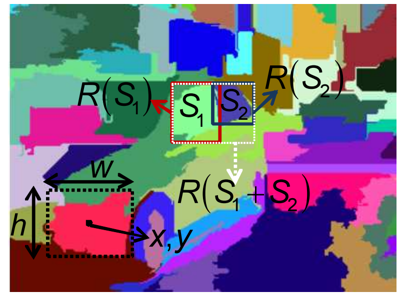


Fig. 3. Object-level confidence estimation from segmented parts produced by part-level segmentation

2.2. Object-Level Segmentation

For object-level segmentation, we construct a graph $G' = (V',E')$ whose vertices $v'_i \in V'$ are the segmented parts produced by part-level segmentation and edges $(v'_i, v'_j) \in E'$ are the weighted means of confidence values computed from pairs of neighboring vertices as $w_o((v'_i,v'_j)) = \ C_f(R(v'_i)) + \ C_f(R(v'_j))$, where C_f is the object confidence value determined by classifiers (see Section 3) and $R(v'_i)$ is a minimum rectangular region that contains all the pixels in the segment v'_i. And and are weighting parameters determined by the sizes of the regions as $= |R(v'_i)|/(|R(v'_i)| + |R(v'_j)|)$ and $= |R(v'_j)|/(|R(v'_i)| + |R(v'_j)|)$, and $|R(v)|$ is the size of the rectangular region for v. After sorting the edge weights in non-deceasing order, we calculate the region comparison predicate $D_o(S_1,S_2)$ between two components S_1 and S_2 by checking if the merged segment has a higher object confidence than the confidence determined by two segments as Eq. (3). Finally, we merge two components when $D_o(S_1,S_2)$ is false.

$$D_o(S_1,S_2) = \begin{cases} \text{true} & \text{if } \hat{C}_f(S_1+S_2) < MCof(S_1,S_2) \\ \text{false} & \text{otherwise} \end{cases} \quad (3)$$

where

$$\begin{aligned} MCof(S_1,S_2) &= \max(\ C_f(R(S_1)) + \ C_f(R(S_2)), C_o) \\ \hat{C}_f(S) &= \max C_f(R(S) + \ ), \quad \sim \text{Uniform}(-B,B) \end{aligned} \quad (4)$$

where $R(S)$ is a minimum rectangular region that contains all the pixels of segments $v'_i \in S$, shown as red and blue boxes in Fig. 3, and $R(S)$ is represented by 4 parameters, x, y, w and h shown around the black-dashed box in Fig. 3. And $= |R(S_1)|/(|R(S_1)| + |R(S_2)|)$ and $= |R(S_2)|/(|R(S_1)| + |R(S_2)|)$.

From these regions, we compute Histogram of Oriented Gradients (HoG) features as inputs to the classifier for object detection as introduced in [1]. The confidence value $C_f(R(S_1+S_2))$ is determined by a minimum rectangular region that includes all the pixels for S_1 and S_2, shown as the white-dashed box in Fig. 3.

Since the object cannot be exactly extracted by part-level segmentation, we generate the several region hypotheses from a merged region $R(S_1+S_2)$ by adding a noise that varies the locations and sizes of the regions, as shown in Eq. (4). For this purpose, we define the lower and upper bounds by B and then uniformly generate the region hypotheses. In our implementation, we set B to $0.2R(S)$, which means that we additionally generate rectangular regions that are 1.2 times larger and 0.8 times smaller than the region $R(S)$ and move the center locations that are uniformly distributed between $x-0.2w$ and $x+0.2w$ along the horizontal axis, and between $y-0.2h$ and $y+0.2h$ along the vertical axis. We pick the best confidence value as $\hat{C}_f(S_1+S_1)$. C_0 is our minimum confidence value that decides the objects of interest.

We summarize our object detection algorithm that hierarchically performs part-level segmentation and object-level graph segmentation as Algorithm 1.

Algorithm 1 Hierarchical Graph-based Segmentation

1) Part-Level Segmentation

Input: A graph $G=(V,E)$ with n pixels and m edges connecting neighboring pixels

- Sort E into $e_1,e_2,\cdots,e_m$ by non-decreasing edge weight
- Start with a segmentation in which each vertex v_i is in its own component C_i
- FOR $q=1,\cdots,m$
  - If $C_i^{q-1}\neq C_j^{q-1}$ and $w_p(e_q)\leq MInt(C_i^{q-1},C_j^{q-1})$, then merge C_i^{q-1} and C_j^{q-1}
- END FOR

Output: Segmented parts $C_1,\cdots,C_{n'}$

2) Object-Level Segmentation

Input: A graph $G'=(V',E')$ with n' segmented parts, $C_1,\cdots,C_{n'}$, and m' edges connecting neighboring parts

- Sort E' into $e'_1,e'_2,\cdots,e'_{m'}$ by non-decreasing edge weight
- Start with a segmentation in which each vertex v'_i is in its own component S_i
- FOR $q=1,\cdots,m'$
  - If $S_i^{q-1}\neq S_j^{q-1}$ and $\hat{C}_f(S_i^{q-1}+S_j^{q-1})\geq MCof(S_i^{q-1},S_j^{q-1})$, then merge S_i^{q-1} and S_j^{q-1}
- END FOR

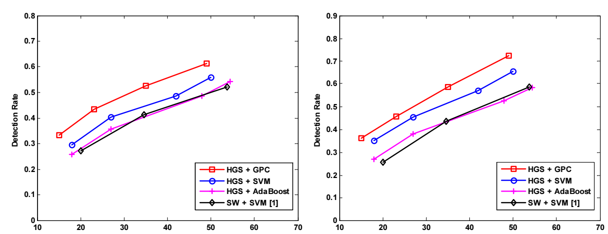
Output: Objects $S_1,\cdots,S_N$ corresponding to the components whose confidence values are larger than C_0

3. EXPERIMENTAL RESULTS

We trained classifiers from manually labeled samples (image patches) consisting of 400 positive samples and 1000 negative samples. We used additional 200 positive samples generated by cutting some positive image patches in half in order to detect objects partially occluded and classify our initial segments from part-level segmentation. We used three different classifiers, a support vector machine (SVM) [7], an AdaBoost classifier [2] and a Gaussian process classifier (GPC) [8][9]. More specifically, we used a linear SVM and the confidence value is determined by the margin value of the SVM classifier. For an AdaBoost classifier, the confidence value is determined by the magnitude of the classifier output. For a GPC, we used a zero mean function and a squared exponential covariance function. We used the logistic function and expectation propagation for likelihood and inference, respectively. More details can be found in [10]. We used the posterior mean as a confidence value because a GPC directly provides probabilistic prediction estimates.

In order to evaluate the performance of vehicle detection under pose variations and partial occlusion, we captured an image sequence with a hand-held camera whose resolution is 640×480 pixels. The ground truth was generated by manually labeling every image. Finally a total of 808 cars from 350 images were determined, and divided into two categories: fully visible objects and partially occluded objects.

We tested 350 images. Fig. 4 shows the comparison result as ROC curves obtained by the proposed hierarchical graph segmentation using three classifiers with the sliding-window-based detection approach [1]. Our approach is more effective for detecting the partially occluded objects, as shown in Fig. 4(b). Fig. 5 shows the detection results by the proposed method with a GPC confirming the detection of objects of various shapes and sizes.



(a) A ROC curve for fully visible objects (b) A ROC curve for partially occluded objects

Fig. 4. ROC curves for vehicle detection using hierarchical graph-based segmentation (HGS) with three different classifiers, a GPC, a SVM and an AdaBoost classifier, and the sliding window (SW)-based approach [1]

For pedestrian detection, we trained two classifiers, the SVM and the GPC, by 400 positive samples and 1000 negative samples obtained from PETS datasets[1]. We evaluated detection performance using 870 images in the i-LIDS dataset[2],

[1] http://www.cvg.rdg.ac.uk/slides/pets.html
[2] http://www.eecs.qmul.ac.uk/ andrea/avss2007d.html

Fig. 5. Results for vehicle detection by using the proposed method with a Gaussian process classifier

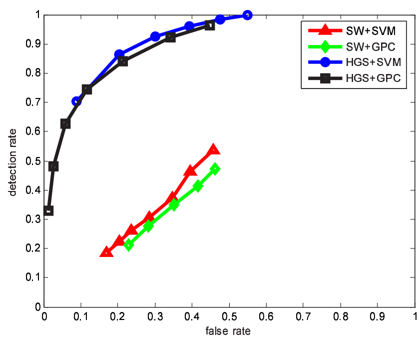


Fig. 6. A ROC curve for pedestrian detection using HGS-based and SW-based methods with two classifiers

which contains many differently sized and shaped objects. Fig. 6 shows comparison results and Fig. 7 shows some results of pedestrian detection using the i-LIDS dataset.

4. CONCLUSION

We have presented an efficient method for object detection using hierarchical graph-based segmentation. Our method hierarchically performs part-level segmentation which is aiming at dividing an image into the color-consistent parts and object-level segmentation to determine the objects using pre-trained classifiers. Our method is able to efficiently detect the objects of various shapes and sizes caused by pose variations and different viewing directions as well as partial occlusion in contrast to sliding-window-based detection approaches.

5. ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF). (No. 2012-0000986) and supported by the IT R&D program of MKE/KEIT. (No. 10040246)

Fig. 7. Results for pedestrian detection by using the proposed method with a support vector machine

6. REFERENCES

[1] N. Dalal and B. Triggs., "Histograms of oriented gradients for human detection," in *CVPR*, 2005.

[2] P. Viola and M. Jones., "Robust real-time obnject detection," *IJCV*, 2001.

[3] P. F. Felzenszwalb and D. P. Huttenlocher., "Efficient graph-based image segmentation," *IJCV*, vol. 59, 2004.

[4] D. Larlus, J. Verbeek, and F. Jurie., "Category level object segmentation by combining bag-of-words models with dirichlet processes and random fields," *IJCV*, vol. 99, pp. 238–253, 2010.

[5] Z. Lin and L. S. Davis., "Shape-based human detection and segmentation via hierarchical part-template matching," *PAMI*, vol. 32, pp. 604–618, 2010.

[6] S. Gould, T. Gao, and D. Koller., "Region-based segmentation and object detection," in *NIPS*, 2009.

[7] R. O. Duda, P. E. Hart, and D. G. Stork., "Pattern classification," in *the Wiley press*, 2001.

[8] C. E. Rasmussen and C. Williams., "Gaussian processes for machine learning," in *the MIT Press*, 2006.

[9] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell., "Active learning with gaussian processes for object categorization.," in *ICCV*, 2007.

[10] C. E. Rasmussen and H. Nichisch., "Gaussian processes for machine learning (gpml) toolbox.," *JMLR*, vol. 11, pp. 3011–3015, 2010.