A Review on Effect of Segmentation and Feature Selection in the Object Detection

Preeti Sinha[1] Toran Verma[2]
[1,2]Department of Computer Science & Engineering
[1,2]Rungta College of Engineering & Technology Bhilai, Chhattisgarh

Abstract— Object Detection and segmentation are the fundamental task in vision and very much interacted with each other. The effect of image segmentation as pre-processing for object recognition is not well understood. One factor hindering the utility of segmentation for recognition is the unsatisfactory quality of image segmentation algorithms. In this work we review different segmentation approaches in the object recognition task and tries to find out best approach. We extend the review to feature selection for the object detection.

Key words: Object detection, Segmentation, Classification

I. Introduction

The interaction between image segmentation and object detection has been an important analysis for many decades in computer vision as well as psychology. Combining object model and initial low level segmentation has been shown to boost segmentation accuracy. The influence of segmentation on the object detection and classification is still not clear.

Realizing global structure of the foreground is the main part of image segmentation. For example, image segmentation methods depends on spectral clustering continued by computing local measurements about every pixel followed by a partitioning [2, 19, 27]. In this type of methods Global descriptors are well represented by set of partition vector and group membership. Many good recognition methods, however, are merely based on local feature descriptors. However in contrast, the principle of global precedence recommends that global image arrangement and formations rule local feature processing in humanoid pattern observation and recognition [8, 18].

Many efforts have been taken to control the segmentation process based on prior knowledge of foreground object to improve the classification in [20]. Segmentation of object of interest, the noise introduced about the object can be minimized. Up till now, approaches of unsupervised segmentation have not been widely held as preprocessing for recognition and classification. It may be due to unacceptable quality of segmentation algorithms. It is generally hard to find segmentations that capture all correct object boundaries in images of real world scenes. If the segments were satisfactory, an ideal segmentation based recognition system would resemble the sketch in Figure 1. After perfect segmentation, each segment (representing an object) is labeled by the recognition engine.

Segment boundaries are used for localization and the scene category label is inferred from the individual object labels. Existing recognition algorithms that advocate the use of segmentation appear to work well if strong initial object hypotheses are built into the segmentation engine [11, 30]. For the task of detecting and recognizing objects in still images without object knowledge, however, the recognition capability is still very weak, perhaps due to the segmentation performance. For example, the approach of Martin et al. [15] attempts to integrate all necessary visual cues together to produce one "best" segmentation. The work of Mori et al. [17] acknowledges that an erroneous segment boundary will degrade recognition accuracy, and thus proposes to over segment an image into super-pixels to increase the potential quality of a single merged segmentation. Alternatively, works such as Viola and Jones [29] suggest that attempts to calculate a segmentation for an input image are likely to introduce more harm than good, and that a bounding box, at every possible location and scale in the image, must be considered as an object outline for satisfactory object recognition and categorization performance.

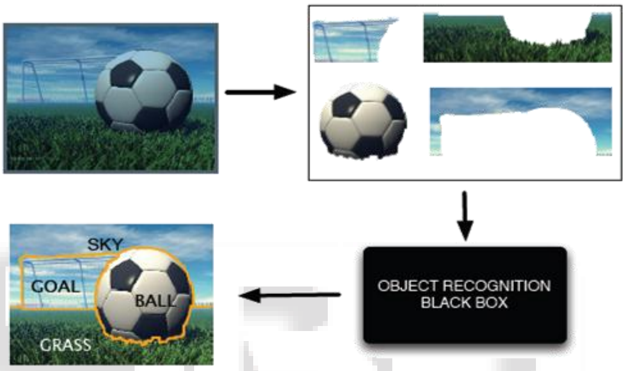

Fig. 1: Illustration of a segmentation-based object detection system.

A reason for the inadequate performance of image segmentation is the ambiguity of image representation, model parameterization, and the task itself. As described in [23], in general there does not exist a single correct segmentation of an image, but rather a shortlist of meaningful image partitionings.

Thus, unlike the above mentioned approaches of using a single segmentation or all possible bounding boxes, the idea of using several segmentations has recently emerged [23, 25, 26]. A handful of segmentations is chosen in hope that a collection of all segments from these few segmentations will result in adequate object boundaries. Russelet al. rely on a collection of random segments to perform object detection, while we use stability as a predictor of "goodness" of a particular set of parameters, cue weightings and model order, as done in [9, 23] to perform object recognition and categorization. Only the most stable segmentations that depict various aspects of the image are chosen to describe object boundaries. In this regard, the segmentations we use go beyond what is available via a simple over segmentation or superpixel representation in terms of capturing salient image structure.

Partitioning images into segments has been proposed for learning the joint distribution of image regions and words for image region annotation [1]. Recently, a work by Rothand Ommer [25] suggested using multiple segmentations for object recognition. They build a segmentation based recognition system and report competitive results. However, they do not show the performance of their system without segmentation. Thus the effects of segmentation on object categorization remain unclear. Also they do not leverage segmentation for object localization and multi-class object recognition.

In this paper we evaluate the benefits of image segmentation, as pre-processing, for object categorization on the Caltech and PASCAL databases.

Finally, we investigate the importance of image segmentation and feature selection for object categorization, and answer the following questions:

- Can segmenting an image improve object recognition?
- How does the number of segments affect recognition accuracy?
- Does the quality of segmentation affect recognition accuracy?
- Is it beneficial to perform localization and multi-class recognition using segmentation?
- What are good diversification strategies for adapting segmentation as a selective search strategy?
- How effective is selective search in creating a small set of high quality locations within an image?
- Can we use selective search to employ more powerful classifiers and appearance models for object recognition.

II. SEGMENTATION FOR DETECTION

To understand the effects of image segmentation on object recognition and categorization, we consider the stability based image segmentation framework and the BoF object recognition model. Although our results are influenced by these choices, we believe that the conclusions will carry over to other object based recognition models.

The goal of an unsupervised clustering algorithm is to partition the data based on some criterion that, by definition, does not use labelled examples. Open problems in this area include choosing the appropriate grouping criterion (cue selection and combination) and the number of clusters (model order). Recent advances in stability based clustering algorithms have shown promising results for choosing these parameters.

Segmentation, which aims for a unique partitioning of the image through a generic algorithm, where there is one part for all object silhouettes in the image. Research on this topic has yielded tremendous progress over the past years [3, 6, 13, 26]. But images are intrinsically hierarchical: In Figure 2a the salad and spoons are inside the salad bowl, which in turn stands on the table. Furthermore, depending on the context the term table in this picture can refer to only the wood or include everything on the table. Therefore both the nature of images and the different uses of an object category are hierarchical. This prohibits the unique partitioning of objects for all but the most specific purposes.

Hence for most tasks multiple scales in segmentation are a necessity. This is most naturally addressed by using a hierarchical partitioning, as done for example by Arbelaez et al. [3]. Besides that a segmentation should be hierarchical, a generic solution for segmentation using a single strategy may not exist at all.

There are many conflicting reasons why a region should be grouped together: In Figure 2b the cats can be separated using colour, but their texture is the same. Conversely, in Figure 2c the chameleon is similar to its surrounding leaves in terms of colour, yet its texture differs. Finally, in Figure 2d, the wheels are wildly different from the car in terms of both colour and texture, yet are enclosed by the car. Individual visual features therefore cannot resolve the ambiguity of segmentation.

And, finally, there is a more fundamental problem. Regions with very different characteristics, such as a face over a sweater, can only be combined into one object after it has been established that the object at hand is a human. Hence without prior recognition it is hard to decide that a face and a sweater are part of one object [29].

This has led to the opposite of the traditional approach: to do localisation through the identification of an object. This recent approach in object recognition has made enormous progress in less than a decade [8, 12, 16, 35].

III. FEATURE SELECTION

In object recognition task robust feature set is very important to discriminate object clearly, even in occlusion and uneven illumination. Various types of features are introduced till now to be used for object detection and recognition task.

A. HoG(Histogram of Oriented Gradient):

The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. Inpractice this is implemented by dividing the image window into small spatial regions (cells), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation. For better invariance to illumination, shadowing, etc., it is also useful to contrast-normalize the local responses before using them.

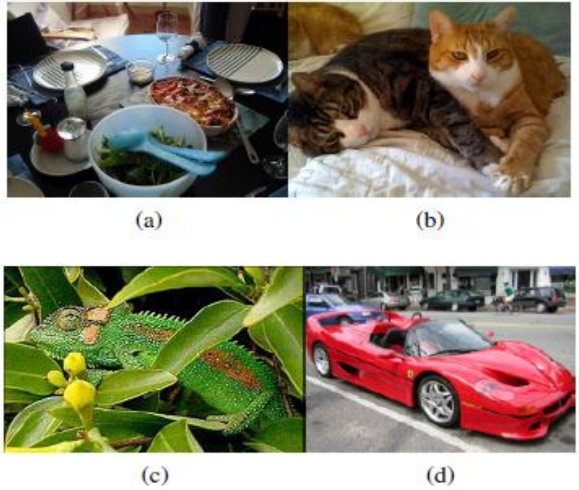

Fig. 2: There are a large number of reasons that an image region makes an object. In (b) the cats can be differentiated by color, not texture. In (c) the chameleon can be differentiated from the surrounding leaves by texture, not color.In (d) the wheels can be part of the car because they are enclosed, not because they are similar in texture or color.

This can be done by accumulating a measure of local histogram energy over somewhat larger spatial regions (blocks) and using the results to normalize all of the cells in the block. We will refer to the normalized descriptor blocks as Histogram of Oriented Gradient (HOG) descriptors[20].

B. Bag of Features:

In this work we utilize the BoF object recognition framework [7, 22] due to its popularity and simplicity. This method consists of four steps: (i) images are decomposed into a collection of "features" (image patches); (ii) features are mapped to a finite vocabulary of "visual words" based on their appearance; (iii) a statistic, or signature, of such visual words is computed; (iv) the signatures are fed into a classifier for labeling. All four steps can be implemented in a variety of ways. Here we adopt the implementation and default parameter settings provided by [28].

C. SIFT Color Descriptors:

Changes in the illumination of a scene can greatly affect the performance of object and scene type recognition if the descriptors used are not robust to these changes. To increase photometric invariance and discriminative power, color descriptors have been proposed which are robust against certain photometric changes [41].

The SIFT descriptor proposed by Lowe [9] describes the local shape of a region using edge orientation histograms. The gradient of an image is shift-invariant: taking the derivative cancels out offsets (Section 2.2). Under light intensity changes, i.e., a scaling of the intensity channel, the gradient direction and the relative gradient magnitude remain the same. Because the SIFT descriptor is normalized, the gradient magnitude changes have no effect on the final descriptor. The SIFT descriptor is not invariant to light color changes because the intensity channel is a combination of the R, G, and B channels. To compute SIFT descriptors, the version described by Lowe [9] is used.

IV. DISCUSSION

Often importance of segmentation is neglected in the object detection and recognition task. Obtaining the segmentation specific to the object of interest is a difficult task. In this paper segmentation strategy is discussed. In object recognition it is important to be consider three things about segmentation First is Diversification of grouping strategy, second is grouping strategy should be fast to compute and third is Scale invariant regions.

Feature Selection for object recognition is an important factor to get good recognition accuracy. The use of orientation histograms has many precursors [13,4,5], but it only reached better grouping strategy and it can be full filled by using maturity when combined with local spatial histogramming and normalization in Lowe's Scale Invariant Feature Transformation (SIFT) approach to wide baseline image matching [12], in which it provides the underlying image patch descriptor for matching scale invariant key points.

V. CONCLUSION

In this paper we study the problem related to the segmentation procedures in object detection and recognition task. Hieratical segmentation presented in [20] can be used to improve the object detection procedure. If hieratical segmentation will be combined with diversified grouping strategy which could cover texture, color, shape etc; then it will improve object detection and recognition accuracy. Object recognition can also be improved using appropriate feature-set and this can be achieved using the combination of bag of features and SIFT descriptors.

REFERENCES

[1] K. Barnard, P. Duygulu, R. Guru, P. Gabbur, and D. Forsyth. The effects of segmentation and feature choice in a translation model of object recognition. CVPR, 2003.

[2] F. Benezit, T. Cour, and J. Shi. Spectral segmentation with multi-scale graph decomposition. In CVPR, 2005.

[3] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. European Conference on Computer Vision, 2, 2002.

[4] T. Cour, F. Benezit, and J. Shi. http://www.seas.upenn.edu/timothee/software/ ncut multiscale/ncut multiscale.html.

[5] G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. Workshop on Statistical Learning in Computer Vision, ECCV, 2004.

[6] M. Everingham et al. The 2005 pascal visual object classes challenge. In In Proc. of PASCAL Challenge Workshop, LNAI,, 2006.

[7] R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. CVPR, 2003.

[8] H. Hughes, G. Nozawa, and F. Kittler. Global precedence, spatial frequence channels, and the statistics of naturals scenes. J. of Cog. Neuroscience, 8(3):197–230, May 1996.

[9] T. Lange, V. Roth, M. Braun, and J. Buhmann. Stabilitybased validation of clustering solutions. In NIPS, 2002.

[10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Computer Vision and Pattern Recognition, 2006.

[11] B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. Workshop on Statistical Learning in Computer Vision, ECCV, 2004.

[12] A. Levin and Y.Weiss. Learning to Combine Bottom-Up and Top-Down Segmentation. ECCV, 2006.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60(2):91–110, 2004.

[14] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In Proc. 7th Int'l. Conf. Computer Vision, pages 918–925, 1999.

[15] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. PAMI, 26(5):530–549, May 2004.

[16] K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. CVPR, 2006.

[17] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: combining segmentation and recognition. In CVPR, 2004.

[18] D. Navon. Forest before trees: The precedence of global features in visual perception. Congitive Psychology, 1977.

[19] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. IEEE transactions on Pattern Analysis and Machine Intelligence, 2012. 3, 8, 10, 13

[20] P. Arbel´aez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. TPAMI, 2011.

[21] J. Carreira and C. Sminchisescu. Constrained parametric mincuts for automatic object segmentation. In CVPR, 2010. 2, 3,

[22] O. Chum and A. Zisserman. An exemplar model for learning object classes. In CVPR, 2007.

[23] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. TPAMI, 24:603–619, 2002.

[24] G. Csurka, C. R. Dance, L. Fan, J.Willamowski, and C. Bray. Visual categorization with bags of keypoints. In ECCV Statistical Learning in Computer Vision, 2004.

[25] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

[26] I. Endres and D. Hoiem. Category independent object proposals. In ECCV, 2010.

[27] M. Everingham, L. V. Gool, C.Williams, J.Winn, and A. Zisserman. Overview and results of the detection challenge. The Pascal Visual Object Classes Challenge Workshop, 2011.

[28] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 88:303–338, 2010.

[29] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. TPAMI, 32:1627–1645, 2010.

[30] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Graph- Based Image Segmentation. IJCV, 59:167–181, 2004.

[31] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. TPAMI, 23:1338–1350, 2001.

[32] C. Gu, J. J. Lim, P. Arbel´aez, and J.Malik. Recognition using regions. In CVPR, 2009.

[33] H. Harzallah, F. Jurie, and C. Schmid. Combining efficient object localization and image classification. In ICCV, 2009.

[34] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Efficient subwindow search: A branch and bound framework for object localization. TPAMI, 31:2129–2142, 2009. 2, 5

[35] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006.

[36] F. Li, J. Carreira, and C. Sminchisescu. Object recognition as ranking holistic figure-ground hypotheses. In CVPR, 2010.

[37] C. Liu, L. Sharan, E.H. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. In Computer Vision and Pattern Recognition 2010. IEEE, 2010. 4 [21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60:91–110, 2004.

[38] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In CVPR, 2008.

[39] S. Maji and J. Malik. Object detection using a max-margin hough transform. In CVPR, 2009. 3

[40] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intel-ligence, IEEE Transactions on, 24(7):971–987, 2002.

[41] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. TPAMI, 32:1582–1596, 2010.