iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 Issue: III Month of publication: March 2015

DOI:

A Survey on Efficient Object Localization and Image Classification for Thumbnail Browsing

P.S.Korhale[1], P.S.Deshpande[2]
[1]PG Student, [2]Asst. Professor
Department of E&TC, STES'S, SMT.Kashibai Navale College of Engineering, Pune, India

Abstract— Nowadays, with an increasing demand of advanced intelligent systems, computer vision applications dealing with foreground objects are becoming more challenging. The common assumption taken into consideration in most of the cases is that an image contains more than one object which produces undesired results when noticeable objects do not appear in the image. Our visual attention model originates from a well-known property of the human visual system that the human visual perception is highly adaptive and sensitive to structural information in images rather than nonstructural information. In this paper, we not only address the problem of ascertaining the existence of objects in an image but also discuss various methods for object detection. The input image is divided into non-overlapping patches, then the patches are categorized into different classes, like natural, man-made, and object to estimate the existence of the object. This paper provides a systematic review of algorithms and performance measures and assesses their effectiveness via metrics.
Index Terms— Existence of objects, Object detection, non-overlapping patches, human visual system (HVS)

I. INTRODUCTION

Over the last years, object class detection has become a major research area. There is a variety of methods but most state-of-the-art method based on sliding window based method. A classifier is trained to distinguish windows containing instances of a given class from all other available windows. The classifier is applied to score every window in a test image.
To define the existence of object, we argue that any object has at least one of three different characteristics:
A. Well-defined closed boundary
B. Different characteristics from its surroundings
C. Some time it is unique within the image and stands out as salient

Many objects have several of these characteristics at the same time. Although objects can appear at a variety of locations and sizes in an image, some windows cover objects without analyzing the pixel patterns inside them. A very large window in an image corner is less probable a priori than a square window in the middle of it[1].
Detecting humans in images is a challenging task owing to their variable appearance and the wide range of poses that human can adopt. The primary need is a robust feature set which allows the human form to be discriminated cleanly, even in cluttered backgrounds under unsuitable illumination. We study the different issues of feature sets for human detection, which shows that normalized Histogram of Oriented Gradient (HOG)[2] descriptors provide excellent performance relative to other existing feature sets including wavelets. The descriptors are reminiscent of edge orientation histograms, SIFT descriptors and shape contexts, but they are computed on a dense grid of uniformly spaced cells and they use overlapping local contrast normalizations for improved performance. A detailed study of the effects of various implementation choices on detector performance, taking "pedestrian detection" which means the detection of mostly visible people in more or less upright poses as a test case.
The rest paper is organized as follows: Section II is about object segmentation. Section III discusses on the different approaches of detection of object in an image. Section IV is about proposed design. Section V concludes the paper.

II. OBJECT SEGMENTATION

Object segmentation is one of the most important and challenging issues in image analysis. It facilitates a number of high-level applications, like object recognition, image editing, scene reconstruction, image retrieval etc. Most existing object segmentation

International Journal for Research in Applied Science & Engineering
Technology (IJRASET)

systems adopt *interaction-based paradigms* [3], i.e., users are asked to provide segmentation cues manually and carefully. Although the interaction-based methods are promising, they all have a critical problem in which they need the users' semantic intention. Such manual labeling is time consuming and often infeasible. Moreover, the segmentation performance heavily depends on the user-specified seed locations. Thus, additional interactions are necessary when the seeds are not accurately provided.

For this reason, developing a sophisticated fully automatic object segmentation method has been strongly demanded. The human brain and visual system can effortlessly grasp certain salient regions in cluttered scenes. By observing the fact that, under most circumstances, the salient parts [4] of an image are usually consistent with interesting objects to be segmented, we first attempt to estimate salient regions. In contrast with existing interaction- based approaches that specify the object and background seeds by manual labeling, our method determines the seed locations based on the visual attention model.

III. LITERATURE SURVEY

There is an extensive literature on object detection. In this section, various techniques that can be used for the purpose of detection of objects in an image are described which shows effectiveness of each method. Fig.1 shows classification for the methods of object detection.

A. *Salient Object Detection for Searched Web Images via Global Saliency*

In recent years, many saliency detection methods have been designed because of its broad applications. However, localizing salient objects is still a very challenging problem.

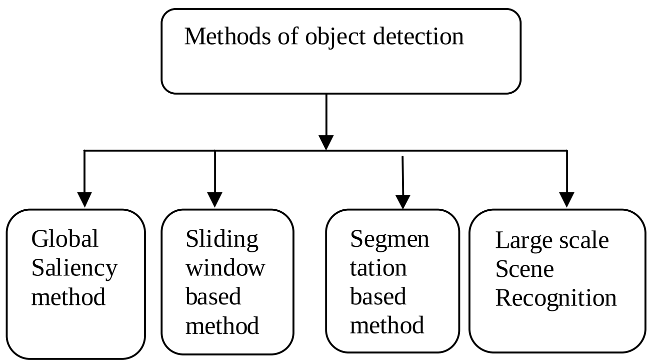

Fig.1: Methods of object detection

There are many practical applications for salient object detection, such as image cropping, adaptive image display on mobile devices, extracting dominant colors on the object of interest for web image filter, removing the images that do not contain an object of interest in image search, and so on [6].

There are several challenges in detecting salient objects. On the one hand, objects have various visual characteristics, which make it hard to differentiate salient objects from the background according to appearance only. Additionally, thumbnail images have a low resolution (e.g., 130 × 130), which is enough for a human to recognize the salient object but makes it difficult to get a reliable segmentation that some previous salient object detection methods rely on.

B. *Sliding window-based method*

Sliding windows detect salient objects by combining local cues. Given a window on the image, the system evaluates the probability of the window containing an object. Heuristic methods that evaluate windows on a single saliency map are efficient. The detection accuracy, however, is not guaranteed. Alexe et al. propose an "Objectness" measure to localize objects in an image. They combine various "Objectness" cues, such as multi-scale saliency,color feachers superpixel straddle, and edge density into a Bayesian framework [7]. One later work instead proposes a limited number of object bounding box candidates. Compared to "Objectness", this approach adopts more robust visual cues and uses Structured SVM for ranking the candidates.

A feature cascade scheme is then used for acceleration. Although the cues provided by the previous methods are effective, for the local characteristics around a single window, the produced bounding box might not be the best globally. Feng et al.compute the window saliency based on super-pixels. They use all the super-pixels outside the window to compose the inside ones, thus the global image context is combined. Although higher precision is achieved compared with "Objectness", the mono scale super-pixel segmentation they use sometimes performs poorly on thumbnail browsing.

C. Segmentation-based method

Alternative approaches generate a salient object bounding box through segmenting the salient object based on the saliency maps. Marchesotti et al. proposed to retrieve similar images, and they separately model the object and background based on those retrieved images. The final saliency regions are segmented via graph cut optimization [8]. However, appearance-based retrieval depends strongly on the database, which limits the generalization of the system. The algorithm of Liu et al. learns to optimally find weights by incorporating various saliency cues from the image.

A binary segmentation step is then applied to find the salient object, but there is a problem of noisy regions yielded from bounding box based training images. To avoid this defect and achieve global optimization, Chen et al. apply Grab-cut to iteratively refine the segmentation based on their proposed saliency maps. Wang et al. integrate Auto-context into the saliency cut for combining context information. Nevertheless, they train the classifier on the pixel level within each iteration, which slows down the progress. Under our scenario, we focus on proposing an algorithm that efficiently generates one global optimal bounding box for object localization. This is based on the consideration that most thumbnail images on the web contain a single salient object. Nevertheless, sliding window-based schemes always propose too many candidate bounding boxes, thus it is perplexing to choose the best one for mentioned applications.

Segmentation-based methods generally propose one global salient object region, but iterative approaches like make the algorithms inefficient in practical usage. More importantly, detecting the existence of salient objects has not been attempted before localization by previous arts. This may lead to unexpected results for background images with possesses the pattern which are repeating, as shown in Fig. 2

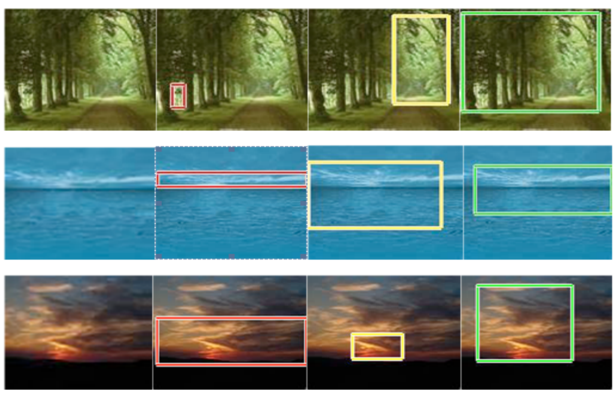

Fig.2: Background object bounding boxes

D. Large-scale Scene Recognition

The fields of computer vision and cognitive science have developed several databases to organize knowledge about object categories, a comprehensive database of real world scenes does not exist. A "scene" is a place in which a human can act within, or a place to which a human being could navigate. How many types of scenes are there? How can the information about environmental scenes be organized? How do the current state-of-art scene models perform on more realistic and ill-controlled environments, and how does this compare to human performance? To date, computational work on scene and place recognition has classified natural images within a limited number of semantic categories, representing typical indoor and outdoor settings [9]. However, any restricted set of categories fails to capture the richness and diversity of environments that make up our daily experience.

Like objects, scenes are related with specific functions and behaviors, such as eating in a restaurant, reading in a library, and sleeping in a bedroom. Scenes, and their associated functions, are very closely related to the visual features that structure the space. The function of environments can be defined by their shape and size (a narrow corridor is for walking, an expansive arena is for public events), by their constituent materials (snow, grass, water, wood, plastic, rubber), or by embedded objects (table and chairs, displays of jewelry, laboratory equipment). The spatial layout and structure of a place often constrain human activities, for instance, a long open corridor is for walking or running, a classroom is for seating.

Previous attempts to characterize environmental scenes have capitalized on uncovering a manageable set of dimensions, features or objects that are correlated with the semantic of purpose of a space [10].

There are three main contributions of object detection. We introduce each of these ideas briefly below.

The first contribution of this paper is a new image representation called an *integral image* that allows for very fast feature calculation. By the work of Papageorgiou et al. our detection system does not work directly with image intensities [11]. Like these authors we use a set of features which is of Haar basis function. To compute these features rapidly at many scales we apply the integral image representation. The integral image can be calculated from an image using a few operations per pixel. Once the integral image is computed, then Harr features can be computed at any scale or location in constant time.

The second contribution of this paper is a method for constructing a classifier by selecting a small number of important features using AdaBoost [12]. Within any image sub window the total number of Harr-like features are large, which are larger than the number of pixels in an image. For the fast classification, the learning process must not take a large majority of the available features, and concentrate on a small set of critical features. By the work of Tieu and Viola, feature selection is performed through a simple modification of the AdaBoost procedure: the weak learner is constrained so that each weak classifier returned can depend on only a one feature. So each stage of the boosting process, which selects a new weak classifier, can be viewed as a feature selection process. AdaBoost gives an effective algorithm and strong bounds on generalization performance.

The third major contribution of this paper is a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image. The impulse behind focus of attention methods is that it is often possible to rapidly determine where in an image an object might occur. More complex processing is reserved only for these sign regions. The important measure of such approach is the "false negative" rate of the process. It must be the case that almost all, object instances are selected by the attentional filter. In general, most existing image thumbnailing methods rely on saliency detection techniques in order to crop the ROI of the given image[13].

IV. PROPOSED DESIGN

In proposed work following steps will be carried out.

A. Segmentation

It is the process of identifying components in the image. Image segmentation has operations such as boundary detection, connected component, thresholding, labeling, etc. Boundary detection helps to find out edges in the image. Differential operator is used for detection of boundary [14]. Thresholding is the process of reducing the grey levels in the image. There are many algorithms exist for thresholding. Segmentation could be used for object recognition, dentistry boundary estimation, image compression or image editing.

B. Foreground extraction

As the name suggests this is the process of separating the foreground and background of the image. We assume that foreground contains the objects of interest. We are using the method for foreground extraction is:

Use of difference images -In this method we use subtraction of images in order to find objects in an image. The result of the subtraction is used as another grey image called difference image. Different types of accumulative difference images are defined [15].

Absolute

$f(x,y) = f(x,y) +1 \ldots\ldots\ldots\ldots\ldots\ldots \text{if } |g(x,y,ti+1) - g(x,y,ti)| > T$

Positive

$f(x,y) = f(x,y) +1 \ldots\ldots\ldots\ldots\ldots\ldots \text{if } g(x,y,ti+1) - g(x,y,ti) > T$

Negative

$f(x,y) = f(x,y) +1 \ldots\ldots\ldots\ldots\ldots\ldots \text{if } g(x,y,ti) - g(x,y,ti+1) > T$

C. Background extraction

Once foreground is extracted a simple subtraction operation can be used to extract the background.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

D. Feature extraction

The next step is to extract useful features from the frames sequence. On the basis of algorithm, definition of feature may vary. The features are defined as follows:

Centroid = (cx, cy)

where, cx = Σ(pi,j * i)/ Σ(p i,j)

cy = Σ(pi,j * j)/ Σ(p i,j)

Dispersion = (Σ √((I – cx)2 + (j – cy)2)*pi,j) / (Σpi,j)

Edge orientation: We take edge orientation feature to distinguish properly between man-made and object classes.

Man-made architectures tend to contain straight lines, while objects tend to have curvature. Therefore, we use the HOG feature to describe the behavior of the edge orientation in patch[12].

V. CONCLUSIONS

We proposed new approach for determination of existence of object in an image. A general comparison is performed between the different methods of the detection of object in an image. We have considered the problem of identifying the existence of objects in an image. Solving this situation, a new approach can be designed which can help to get rid of difficulties observed during this survey of object detection techniques. This paper brings together new algorithms and representations which are quite generic and may well have broader application in computer vision and image processing. For future research, we can demonstrate the effectiveness of this review in the application of image thumbnailing and recovery of the image.

VI. ACKNOWLEDGMENT

I am indeed thankful to my guide **Prof. P. S. Deshpande** for her valuable guidance and assistance to complete this paper; otherwise it would not have been accomplished. I extend my special thanks to Head of Department of Electronics & Telecommunication, **Dr. S. K. Shah** who extended the preparatory steps of this paper-work. I am also thankful to the head & Principle of STES'S, SMT. Kashibai Navale College of Engineering, **Dr. A.V. Deshpande** for his valued support and faith on me.

REFERENCES

[1] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," IEEE Trans. Patt. Anal. Mach. Intell., vol. 34, no. 11, pp. 2189–2202, 2012.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Int. Conf. Comput. Vis. and Pattern Recognit, 2005, vol. 1, pp. 886–893.

[3] E. N. Mortensen and W. A. Barrett, "Intelligent scissors for image composition," in Proc. ACM SIGGRAPH, 1995, pp. 191–198.

[4] C. Jung and C. Kim, "A unified spectral-domain approach for saliency detection and its application to automatic object segmentation," IEEE Trans. Image Process., vol. 21, no. 3, pp. 1272–1283, 2012.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in IEEE Int. Conf. Comput. Vis. and Pattern Recognit, 2001, vol. 1, pp. I–511.

[6] P. Wang, J. Wang, G. Zeng, J. Feng, H. Zha, and S. Li, "Salient object detection for searched web images via global saliency," in IEEE Conf. Comput. Vis. Pattern Recognit, pp. 3194–320, 2012.

[7] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. Salient object detection by composition. In ICCV, pages 1028–1035, 2011.

[8] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen. Gaze-based interaction for semi-automatic photo cropping. In Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, volume 1 of Collecting and editing photos, pages 771–780, 2006.

[9] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in IEEE Conf. Comput. Vis. Pattern Recognit, 2010, pp. 3485–3492.

[10] B. Suh, H. Ling, B. B. Bederson, and D.W. Jacobs, "Automatic thumbnail cropping and its effectiveness," in Proc. ACM Symp. User Interface Software and Technology, 2003, pp. 95–104.

[11] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In International Conference on Computer Wsion, 1998

[12] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt '95,pages 23-37. Springer-Verlag, 1995.

[13] J.Choi, C.Jung, J.Lee, and C.Kim, "Determining the existance of objects in an image and its application to image thumbnailing", IEEE sig. procc. Letters, vol.21, no.8, pp.957-961, Aug.2014.

[14] H. M. Dee, and S. A. Velastin, ―How close are we to solving the problem of automated visual surveillance?,‖ Machine Vision Applications, vol. 19, no. 5-6, , pp 329-343, Sep. 2008

[15] T. Celik, and H. Kusetogullari, ―Solar-Powered automated road surveillance system for speed violation detection," IEEE Trans. Ind. Electron., vol. 57, no. 9, pp 3216-3227, Sep. 2010.