Cognitive Vision for Efficient Scene Processing and Object Categorization in Highly Cluttered Environments

Changhyun Choi and Henrik I. Christensen
Robotics & Intelligent Machines, College of Computing
Georgia Institute of Technology
Atlanta, GA 30332, USA
{cchoi,hic}@cc.gatech.edu

***Abstract*— One of the key competencies required in modern robots is finding objects in complex environments. For the last decade, significant progress in computer vision and machine learning literatures has increased the recognition performance of well localized objects. However, the performance of these techniques is still far from human performance, especially in cluttered environments. We believe that the performance gap between robots and humans is due in part to humans' use of an attention system. According to cognitive psychology, the human visual system uses two stages of visual processing to interpret visual input. The first stage is a pre-attentive process perceiving scenes fast and coarsely to select potentially interesting regions. The second stage is a more complex process analyzing the regions hypothesized in the previous stage. These two stages play an important role in enabling efficient use of the limited cognitive resources available. Inspired by this biological fact, we propose a visual attentional object categorization approach for robots that enables object recognition in real environments under a critical time limitation. We quantitatively evaluate the performance for recognition of objects in highly cluttered scenes without significant loss of detection rates across several experimental settings.**

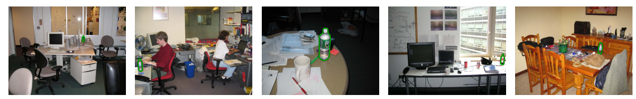
I. Introduction

In computer vision, the object recognition area has experienced significant progress over the last decade. But most of the object recognition systems still require uncluttered scenes or enough resolution of images. It is still hard to recognize objects in extreme situations, such as highly cluttered scenes or too small objects in images. In this paper, we propose an object recognition approach that can handle some of these situations efficiently and robustly.

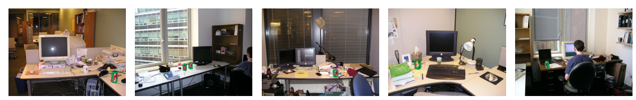
As a first step to understand scenes in a complex world, we need a mechanism to hypothesize important regions. Treisman [1] proposed a theory for object recognition composed of two stages inspired by human visual search strategies. According to the theory, when humans do visual search, in the first stage they select highly salient regions by integrating multiple features, such as shapes, colors, lines, and curves. In the second stage they carefully identify objects within the salient regions. Unfortunately, within the vision community these two stages have been developed separately, and there have been few attempts to combine them.

A. Visual Attention

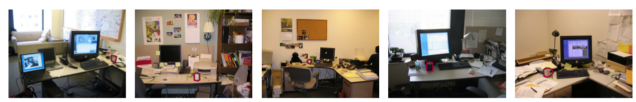
Itti and Koch [2] proposed a model of saliency-based visual attention based on Treisman's feature-integration theory [1]. The bottom-up visual attention model automatically identifies highly salient regions based on color, intensity, and orientation stimuli. Recently, Hou and Zhang [3] proposed a Spectral Residual (SR) approach for fast saliency detection. In this approach, salient regions are selected from spectral residual, which is the difference between the log spectrum and the smoothed log spectrum of an image. Since the approach relies on the Fourier Transform and the Inverse Fourier Transform, it can detect salient regions efficiently, and demonstrates better detection performance than the Itti's model. Wang and Li [4] enhanced the SR approach by using a two-stage approach, but this approach is still limited to the bottom-up saliency detection.



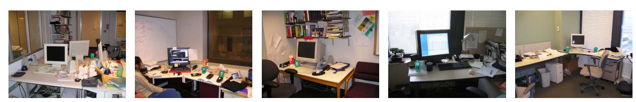
(a) Bottle



(b) Can



(c) Mug



(d) Paper cup

Fig. 1: Example images from four target object classes obtained from the LabelMe dataset [5]. The labels are depicted on the images. Each object is in a complex environment and the size of each object is very small.

B. Object Categorization

Recent work in cognitive science [6] and neuroscience [7] suggest that if salient regions are determined by attention, more detailed visual information of the regions is processed through eye movements, so-called "saccades". The fovea

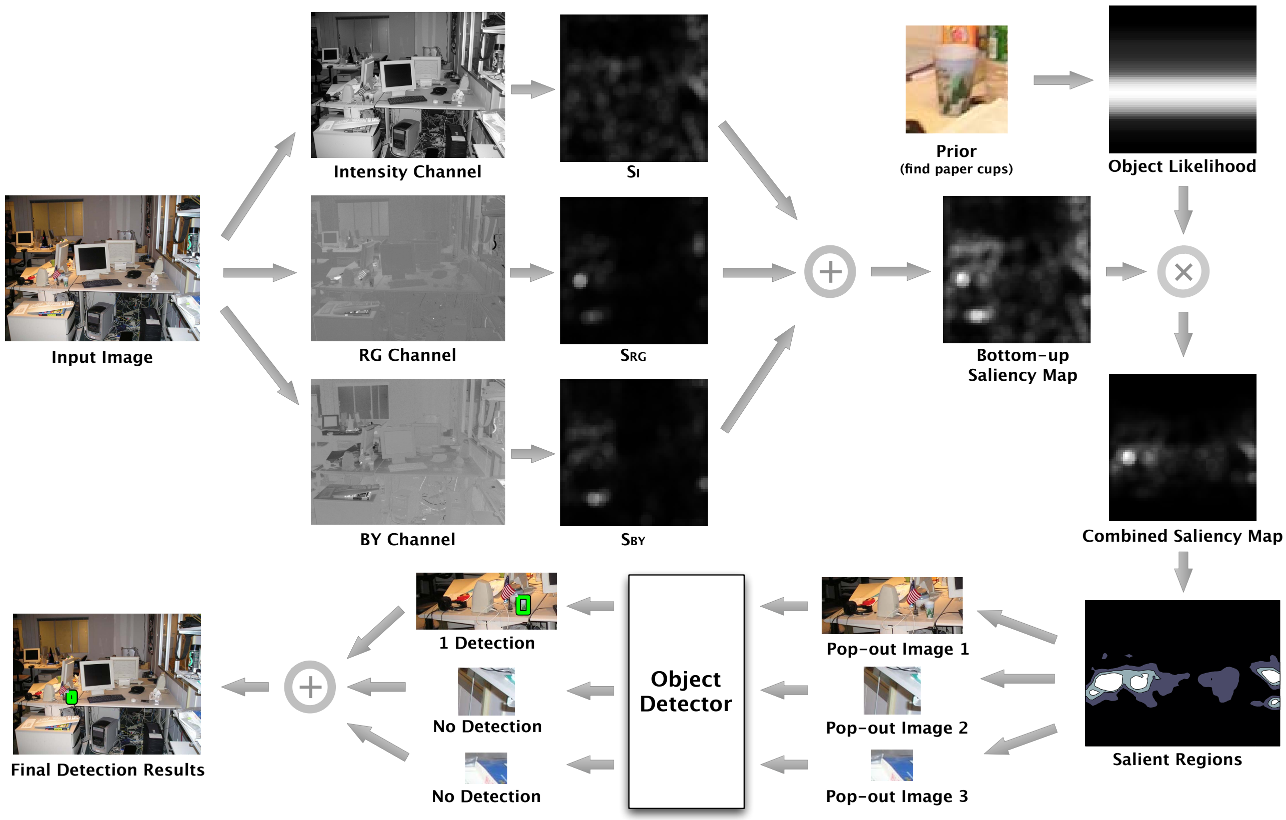


Fig. 2: The overall data flow. A bottom-up saliency map is computed by combining intensity and two color channels. The bottom-up saliency map is then combined with top-down contextual knowledge, which is the object likelihood. The combined saliency map is used to detect salient pop-out regions determined by thresholding. The pop-out regions are examined to detect objects.

subtends less than 2 arc seconds and as such for many structures there is a need to move the fovea to provide detailed coverage of the entire object. During the saccade period, this partial visual information is integrated for higher level processing, and this process is termed trans-saccadic perception. In trans-saccadic perception, the types of visual features and the ways of processing are still a subject of ongoing research, but it implies that parts of an object and their relationship play an important role in human object recognition.

In computer vision, methods based on patches of object appearance have recently been combined with machine learning methods, such as boosting [8] and support vector machines [9], and these combined approaches have shown notable recognition performance. The bag of words models of Fei-Fei and Perona [10] and Sivic *et al.* [11] represent objects as local appearance and have been used in object categorization. While these models ignore local appearance's spatial relationship, the approach of Felzenszwalb and Huttenlocher [12] uses both local appearance and relative position. Similarly, Torralba *et al.* [13], [14] proposed a method for object classification by using gentle boosting with local stumps and their spatial relationship. For the second detailed analysis of attended regions, we apply a boosting algorithm based on Torralba's work.

C. Previous attempts of the two-stage recognition

Torralba *et al.* [15] introduced an approach combining the bottom-up saliency with top-down knowledge from global features. More specifically, they used a set of global features as contextual guidance of attention. This approach is, however, limited to saliency detection, and is not extended to object recognition. One of the main advantages of combining bottom-up attention and high level processing is computational efficiency. Efficiency is one of the key requirements in robotic systems. Frintrop *et al.* [16] proposed a system that determines salient regions from a 3D laser scanner and efficiently detects objects on the selected regions by using the Viola-Jones classifier [8]. To select potentially important areas in images, Vogel *et al.* proposed a gaze planning system composed of bottom-up attention and top-down information [17]. They used a probabilistic framework to determine highly probable regions of target objects by using gist features and a SMLR classifier. Peripheral-foveated vision systems also use bottom-up attention to calculate important regions in low resolution images and do more complex tasks, such as object recognition or active object acquisition, in high resolution images [18], [19], [20]. With Itti's model,

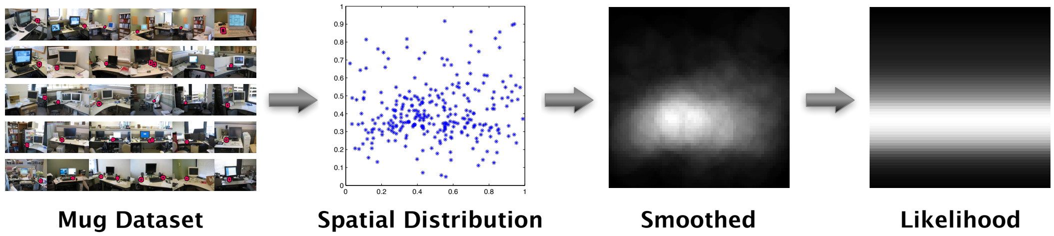


Fig. 3: Generation of an object likelihood model. The fundamental idea behind this is that an object's position follows predictable patterns. To find the spatial distribution we searched for object positions within the LabelMe dataset (in mug case, 287 mug positions were obtained from 174860 objects in the dataset). After smoothing with a 2D Gaussian filter, the values are projected onto the y-axis. The likelihood is simply the projected values with a uniform distribution along the x-axis. The likelihood for each object was computed offline.

Dirk Walther *et al.* [21], [22], [23] tried to analyze the usefulness of attention for object recognition. They showed that applying object recognition based on SIFT descriptors [24] on the attended regions is computationally efficient and suitable for active object learning. However, SIFT descriptor-based recognition is not a good method to apply to pop-out images due to two considerations. First, it is only applicable to objects that have enough discriminative texture or shape. There are still numerous objects which do not meet this requirement. Secondly, the resolution of an object in an image should be large enough to be recognized. In descriptor-based recognition, it is hard to extract distinctive descriptors from low resolution images, which mainly result from the object occupying a relatively small area in an image.

To overcome these shortcomings of descriptor-based recognition in attended regions, we propose an approach using a boosting algorithm, which is capable of recognizing objects having limited or no texture at low image resolutions. In section II, we introduce a way to detect salient regions for visual attention considering object likelihood as well as bottom-up visual saliency. We then explain how boosting can be applied to the pop-out regions selected by visual attention in section III. In section IV, we quantify the performance of our approach. Finally, we conclude and briefly mention future work in section V.

II. The First Stage: Saliency Detection with Color Channels and Object Likelihood

To detect salient regions in an image, our approach is based on the recently proposed Spectral Residual (SR) approach [3]. Although the SR can detect salient regions efficiently with gray-level intensity, it does not directly generalize to color saliency. To address this issue we devise an enhanced approach, which we named Spectral Residual with Color channels (SRC).

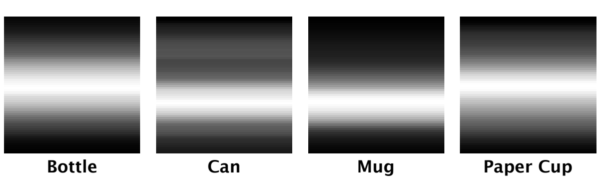


Fig. 4: Object likelihood models for each target object class.

For the SRC, we tried to implement Itti's visual attention model [2] as an SR approach. Itti's model uses intensity, color, and orientation as input channels. Since orientation is well detected in the SR of the intensity image [25], we omitted the orientation channel. From the color channels four broadly-tuned color channels are created as follows:

$$R = r - \frac{1}{2}(g + b) \tag{1}$$

$$G = g - \frac{1}{2}(r + b) \tag{2}$$

$$B = b - \frac{1}{2}(r + g) \tag{3}$$

$$Y = \frac{1}{2}(r + g) - \frac{1}{2}|r - g| - b \tag{4}$$

where r, g, and b are the red, green, and blue channels of the input image. To replicate the effect of color double-opponent cells in human primary visual cortex, the red/green and blue/yellow channels are created (5) and (6), and the intensity channel is simply computed by (7):

$$RG = R - G \tag{5}$$

$$BY = B - Y \tag{6}$$

$$I = \frac{1}{3}(r + g + b) \tag{7}$$

These channels are then converted to separate saliency maps by using the SR approach [3], $\mathcal{SR}(\cdot)$ (8) - (10):

$$S_I = \mathcal{SR}(I) \tag{8}$$

$$S_{RG} = \mathcal{SR}(RG) \tag{9}$$

$$S_{BY} = \mathcal{SR}(BY) \tag{10}$$

Finally, we apply a map normalization operator $\mathcal{N}(\cdot)$ [2], which globally promotes a small number of outstanding peaks and suppresses numerous similar peaks, to each

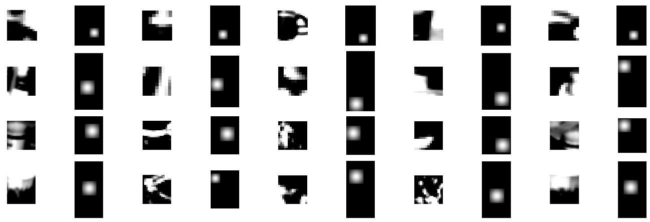


Fig. 5: Visual memory is composed of local stumps (image patch on the left) and spatial masks (on the right). Only five features are depicted per target object classes. (from top to bottom) bottle, can, mug, and paper cup.

saliency map, S_I, S_{RG}, and S_{BY}, and we get a bottom-up saliency map S_{bu}:

$$S_{bu} = \frac{1}{3}(\mathcal{N}(S_I) + \mathcal{N}(S_{RG}) + \mathcal{N}(S_{BY})) \tag{11}$$

To obtain pop-out image regions that are more likely to contain the objects we seek, the bottom-up saliency S_{bu} is combined with an object likelihood model L:

$$S = S_{bu} \otimes L \tag{12}$$

where $\otimes$ is the element-wise multiplication. The reason why we use the object likelihood is that the position of objects has some spatial constraints. For instance, mug – one of our target object classes – is likely to be placed in the middle of images because it is usually located on the table or next to a sink. Of course, this likelihood can be adjusted with respect to robots' viewpoint (tilt), although we leave this enhancement to future work. The generation of an object likelihood model is explained in Fig. 3. Fig. 4 shows the object likelihood models used in this paper. Note that although each likelihood is slightly different, the overall shapes are very close because these objects share similar possible locations, such as tables and desks.

A simple way to obtain pop-out images from the saliency map S is through thresholding. Threshold values were determined experimentally. In addition, we also explored the empirical number of pop-out images through experiments. The experiments will be discussed further in section IV-A.

III. The Second Stage: Object Categorization with Boosting

Once pop-out images are obtained from a saliency map, our approach then executes a recognition algorithm on these images. For this, we use a gentle-boost classifier with local stumps and their spatial masks [14].

A. Training data and target object classes

For tuning of the boosting classifier there is a need to have a training set. Data with four different object classes were selected. The images used contain objects that are frequently encountered in our daily lives. The objects were partly selected to be difficult to detect with standard descriptor based methods. Finally, it was desirable to have data that can be obtained from the LabelMe dataset [5]. Fig. 1 represents some selected images among the dataset of the four target object classes: bottle, can, mug, and paper cup. Note that the objects are in highly cluttered scenes and the proportion of each object in the images is relatively small. In some images it is even hard for a human to find the target objects. During the training phase, local patches and their spatial information are saved into visual memory which will be used during the detection phase (Fig. 5).

B. Boosting on pop-out images

The usual way to detect objects with a boosting classifier is through sliding of different sized windows across an image. Such a brute-force strategy is not very efficient. Since in real scenes there are many uniform regions, applying boosting only to highly salient regions, which are probable regions having target objects, is a better strategy, and it is computationally efficient. The overall procedure including the first stage is systematically represented in Fig. 2. In section IV-B, we show that our approach scanning only pop-out images can significantly save computing time.

IV. Experiments and Evaluations

In this section we document the performance of our approach. In section IV-A, we compare the detection rates of SR, SRC, and SRC+LH (SRC with object likelihood), and we show that our SRC+LH method has better performance. Similarly, we quantify the performance of boosting with and without visual attention by comparing both the recognition result and the computation times in section IV-B.

As we mentioned in section III-A, we gathered a dataset by searching for the four target objects in the LabelMe [5] dataset.

A. Evaluating Saliency Detection

Initially i^{th} pop-out images PO_i are acquired by thresholding of the i^{th} saliency map S_i by (13):

$$PO_i = \begin{cases} 1 & \text{if } S_i > \tau_{pu} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Similarly, the i^{th} positive region P_i can be determined by referring to the label data (14), and the i^{th} negative region N_i is simply the positive region's complement (15):

$$P_i = \begin{cases} 1 & \text{if it is labeled area} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$N_i = \overline{P_i} \tag{15}$$

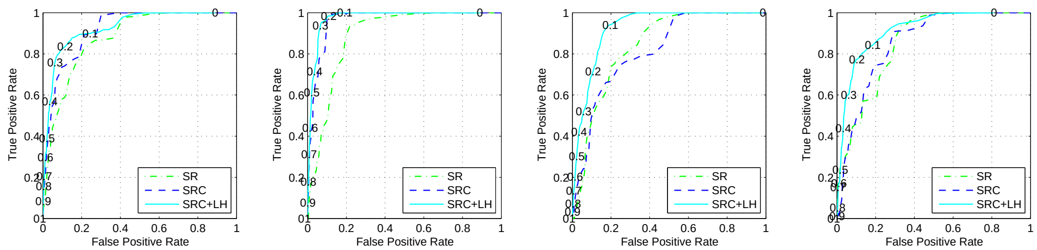
The i^{th} true positive region TP_i and the i^{th} false positive region FP_i are obtained by (16) and (17):

$$TP_i = PO_i \cap P_i \tag{16}$$

$$FP_i = PO_i \cap N_i \tag{17}$$

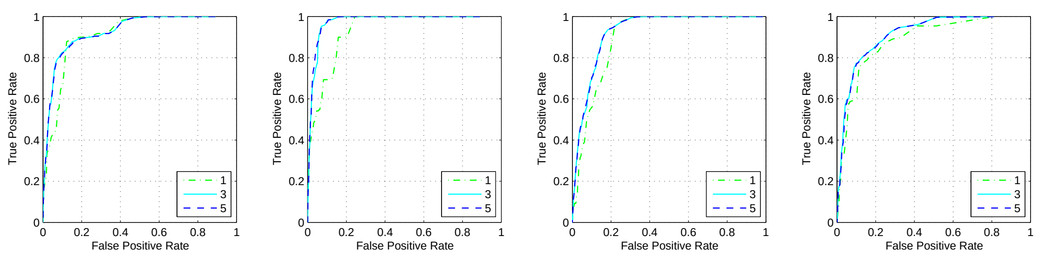
Finally, the total true positive rate TPR and the total false positive rate FPR over a set of a target object class are calculated as follows:

$$TPR = \sum_{i=1}^{K} \frac{TP_i}{P_i} \tag{18}$$



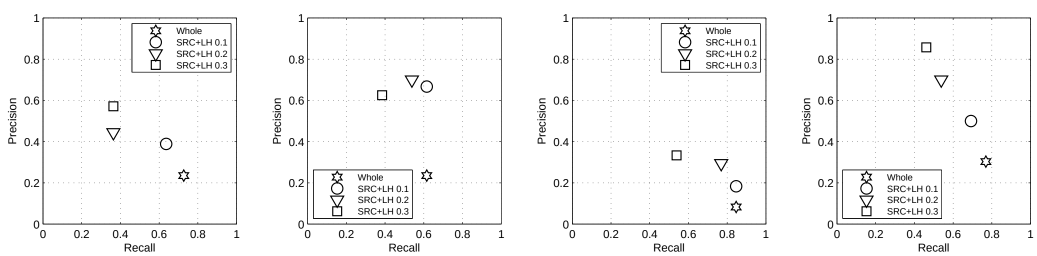
(a) Bottle (b) Can (c) Mug (d) Paper cup

Fig. 6: ROC curves of SR, SRC, and SRC+LH by varying the threshold, τ_{pu}, from 0 to 1. The values of τ_{pu} are depicted on SRC+LH curves. SRC+LH outperforms the other two methods. This ROC curves shows that best threshold value τ_{pu} is between 0.1 and 0.3.



(a) Bottle (b) Can (c) Mug (d) Paper cup

Fig. 7: ROC curves of SRC+LH by varying the number of pop-out image regions, τ_n, from 1 to 5. Note that 3 and 5 are nearly the same and 1 shows good performance as well. This implies that the first pop-out image is highly likely to contain target objects, and fourth or fifth pop-out images are less likely to have the target objects.



(a) Bottle (b) Can (c) Mug (d) Paper cup

Fig. 8: Precision-recall graphs for the four tests. The brute force search strategy (Whole) represents very low precision because of many more false positives. Our approaches (SRC+LH where $\tau_{pu} = 0.1$, 0.2, and 0.3) show much better precision without significant loss of recall.

$$FPR = \sum_{i=1}^{K} \frac{FP_i}{N_i} \tag{19}$$

where K is the number of images of the target object's dataset.

With these criteria, the ROC curves of the four target object classes are obtained by varying τ_{pu} from 0 to 1 as shown in Fig. 6. According to the ROC curves the SRC+LH method is superior to the SR or SRC applied on their own. The threshold value τ_{pu} shows the best performance around 0.2, but varies slightly depending on target objects.

In addition to varying τ_{pu}, we also investigated the effect of the number of pop-out images, τ_n. The ROC curves of SRC+LH with respect to the τ_n are depicted in Fig. 7 by varying τ_n from 1 to 5. Following winner-take-all and inhibition of return [2], we selected pop-out images in order of peak values. In Fig. 7 the detection rates generally converge when $\tau_n = 3$. Considering that the detection rates in $\tau_n = 1$ are high enough, we can guess that the first pop-out image is highly likely to have the target object.

B. Evaluating Object Categorization with and without Visual Attention

In this section, for verifying the advantages of using visual attention, we present an experiment with boosting applied to the original images and pop-out images generated by our visual attention model. Since the size of each image varies, we first resize every image to 1024×768 pixels, and for scale invariance we build up pyramidal images across

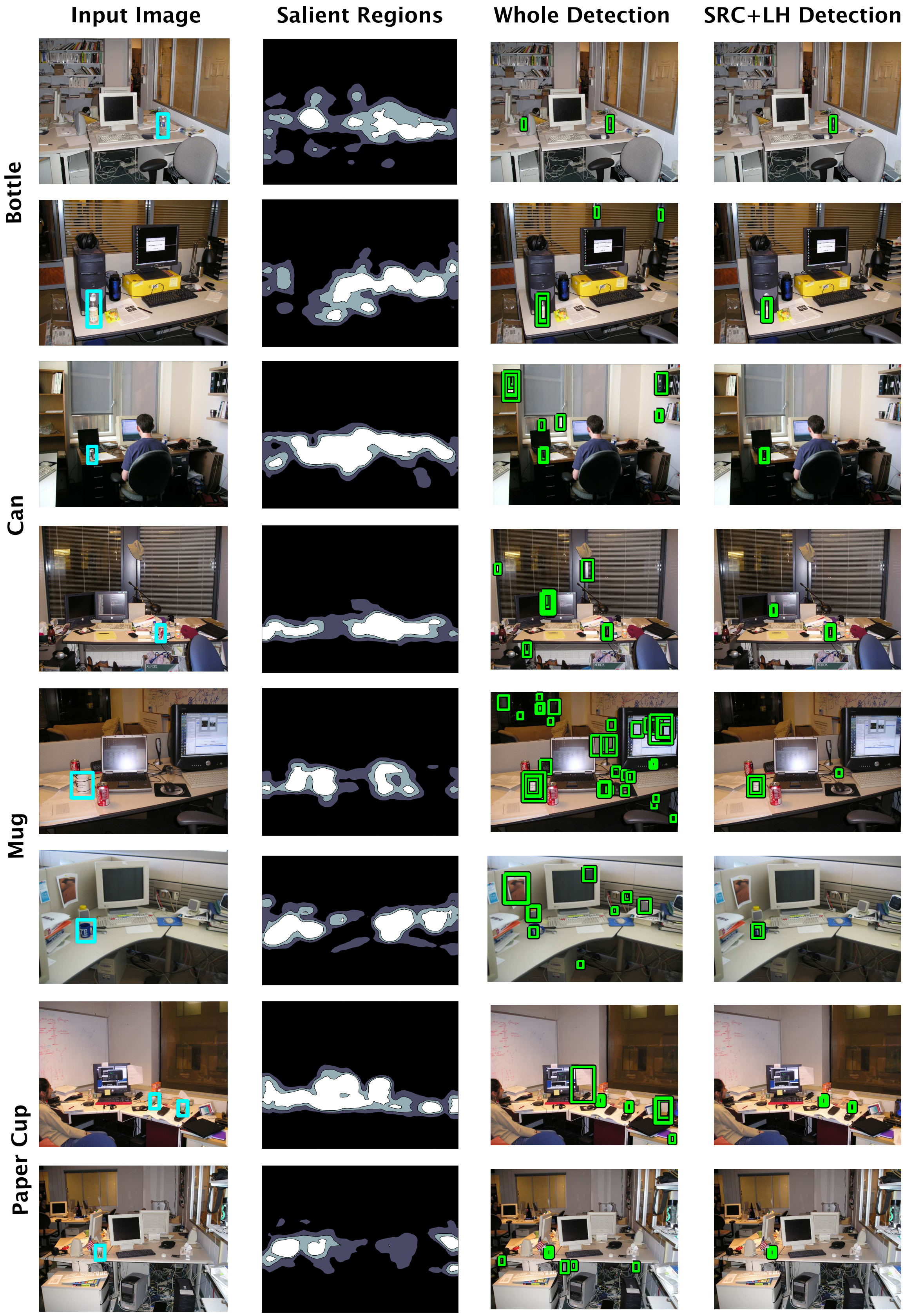


Fig. 9: Selected object detection results. While the whole detection strategy returns an amount of false positives, our strategy can localize target objects with significantly fewer false positives. The ground truth position of target objects are depicted on input images.

Fig. 10: Average execution times of each strategy. Our SRC+LH strategies outperform the whole detection strategy in computing times. The error bars represent the standard deviation of execution times.

TABLE I: F_1-measure for objects detected. Our SRC+LH strategies show higher scores than the whole detection strategy.

	Bottle	Can	Mug	Paper cup
Whole	0.3556	0.3404	0.1507	0.4348
SRC+LH 0.1	**0.4828**	**0.6400**	0.3014	0.5806
SRC+LH 0.2	0.4000	0.6087	**0.4255**	**0.6087**
SRC+LH 0.3	0.4444	0.4762	0.4118	0.6000

five scales in which the scale step is 0.7. Fig. 9 shows representative boosting results of bottle, can, mug, and paper cup.

To count the detection across five scales of an image, we merge the detection regions into one image. We compare results through merging of regions. In the image, when there are more than two areas in a merged region, a comparison is performed. If the merged area is the label area which contains the target object region, it is regarded as a true positive, and if the merged area has less than 50% overlap with the target region, it is regarded as a false positive. In the false positive case, if the merged area is composed of two or more areas across scales, we count the maximum number of false positive areas for each scale. By following these criteria, we plot precision-recall graphs in Fig. 8. According to the plots, the precision increases as τ_{pu} increases, while the recall varies significantly less. The F_1-measure for the detection results is shown in Table I. These results show that our visual attention model SRC+LH performs better than the complete detection strategy.

The advantages from visual attention are not only in terms of fewer false positives, but also in terms of computational benefit. The application of a saliency detector reduces the need for search which in turn reduces complexity. Fig. 10 shows the execution times for our approach. In SRC+LH strategies, the computation times of saliency detection with object likelihood are included in the average execution times, but since our attentional model is efficient enough, the additional cost is negligible. As τ_{pu} increases, the size and the number of pop-out images decrease, hence the execution times decrease significantly.

V. Conclusions and Discussions

Inspired by the two-stage framework from cognitive psychology, we proposed an object class recognition approach using bottom-up visual priming, top-down object likelihood, and a boosting object classifier. Our SRC+LH approach detects more accurate pop-out images than the original SR approach. We also showed that a gentle boosting classifier with visual attention promises better precision as well as more efficient computations. We believe that our combined approach will be an alternative to previous descriptor-based recognition schemes to detect objects in extreme situations.

We anticipate that our approach will be useful in robotic applications, especially, in the service robotics area which requires robust object categorization in highly cluttered environments under some time constraints. Here our approach can provide an efficient and accurate object class recognition solution. When our approach is applied in robotic applications, we expect that the recently proposed spatio-temporal saliency detection [25] will help robots perceive an additional saliency channel, motion saliency, which has been ignored for a long time even though it is very important in human perception [26], [27], [28].

Although we tried to minimize false positives by only focusing on pop-out regions, we think that there is more room to enhance the recognition rates. As Torralba *et al.* [15] indicated, the global scene based context will play an important role as a top-down guidance for adjusting the object likelihood. With the place context we should also consider the camera tilt information of robots in order to obtain a more accurate saliency map. In addition, if we use additional prior knowledge of target objects, such as maps and objects' positions in them obtained from SLAM (Simultaneous Localization And Mapping) which is a major robotics area, we could expect more robust recognition through probabilistic inference.

Last but not least, even though we do not currently adopt the joint boosting proposed by Torralba *et al.* [14], if we try to share local stumps across object classes, we could anticipate not only better recognition performance, but also attain a solution for scalable issue as object's classes increase.

VI. ACKNOWLEDGMENTS

The authors would like to thank Dr. James M. Rehg for fruitful discussions, and acknowledge anonymous reviewers' helpful comments.

REFERENCES

[1] A. M. Treisman and G. Gelade, "A feature-integration theory of attention." *Cognit Psychol*, vol. 12, no. 1, pp. 97–136, January 1980.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[3] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *Proc. CVPR, 2007*, pp. 1–8, June 2007.

[4] Z. Wang and B. Li, "A two-stage approach to saliency detection in images," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 965–968, 31 2008-April 4 2008.

[5] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157–173, 2008.

[6] D. Melcher, "Selective attention and the active remapping of object features in trans-saccadic perception," *Vision Research*, vol. In Press, Corrected Proof, pp. –, 2008.

[7] E. I. Knudsen, "Fundamental components of attention." *Annu Rev Neurosci*, April 2007.

[8] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[9] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *Int. J. Comput. Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[10] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," *Proc. CVPR, 2005*, vol. 2, 2005.

[11] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering object categories in image collections," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2005.

[12] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *Int. J. Comput. Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[13] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, p. 2000, 2000.

[14] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: efficient boosting procedures for multiclass object detection," *Proc. CVPR, 2004*, vol. 2, pp. II–762–II–769 Vol.2, June-2 July 2004.

[15] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search." *Psychological Review*, vol. 113, no. 4, pp. 766–786, October 2006.

[16] S. Frintrop, A. Nuchter, H. Surmann, and J. Hertzberg, "Saliency-based object recognition in 3D data," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3, 2004, pp. 2167–2172 vol.3.

[17] J. Vogel and N. de Freitas, "Target-directed attention: Sequential decision-making for gaze planning," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 2372–2379.

[18] S. Gould, J. Arfvidsson, A. Kaehler, B. Sapp, M. Meissner, G. Bradski, P. Baumstarck, S. Chung, and A. Y. Ng, "Peripheral-foveal vision for real-time object recognition and tracking in video," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2007.

[19] D. Meger, P. Forssn, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe, "Curious george: An attentive semantic robot," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 503–511, June 2008.

[20] P. Forssen, D. Meger, K. Lai, S. Helmer, J. Little, and D. Lowe, "Informed visual search: Combining attention and object recognition," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 935–942.

[21] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition - a gentle way," *Lecture Notes in Computer Science*, vol. 2525, no. 1, pp. 472–479, 2002.

[22] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006.

[23] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "On the usefulness of attention for object recognition," *Workshop on Attention and Performance in Computational Vision at ECCV*, pp. 96–103, 2004.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[25] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *Proc. CVPR, 2008*, pp. 1–8, June 2008.

[26] A. Correa, D. Sanabria, C. Spence, P. Tudela, and J. Lupianez, "Selective temporal attention enhances the temporal resolution of visual perception: Evidence from a temporal order judgment task," *Brain Research*, vol. 1070, no. 1, pp. 202 – 205, 2006.

[27] P. Fraisse, "The psychology of time," *Greenwood Press,Westport, CT*, 1963.

[28] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, January 1985.