Recognition of Visual OPS using Distributed Clustering

[1]Ms. Kalyani Ugale , [2]Dr. A.D Gawande
[1]PG student at Sipna COET Amravati (MS) India, [2]Professor& Head CSE at Sipna COET Amravati (MS) India
[1]Comuter science & Engineering,
Sipna COET Amravati (MS) India

Abstract : Everyday objects or stores by using image based mobile interactions through their devices. For example, as users walk on the street, they might simply point the mobile camera to a store on the street to quickly access its related information, inquire special offers, and make reservations through their mobile devices without physically entering the store. Users depend on mobile device to maintain information and update. OPSs (On Premises Signs) shows great visual diversity accompanied with complex environmental conditions. Consider an example, user walk on the street and he simply point his mobile camera to a store to quickly access its related information, inquire special offers, and make reservations through his mobile without physically entering on that store. Street view scenes are commonly captured by customers devices and they have more real-world characteristics lacking in most existing image datasets, e.g. perspective distortion, foreground and background clutter, etc. To learn a reliable OPS model for recognizing OPSs, a labeled dataset with a huge amount of real-scene images is required. However, precisely labeling OPS categories and regions, i.e., generating strong labels for Learning involves a significant amount of human labor, and thereby is usually not feasible for training a real-scene OPS model

***IndexTerms* – OPSs, Dataset, Cluster**

I. INTRODUCTION

The intimate presence of mobile devices in our daily life has dramatically changed the way we connect with the world around us. Users depend on mobile devices to maintain information and updates related to all general purpose applications. OPSs (On Premise Sign) are commonly erected on buildings in modern cities. These OPSs exhibit great visual diversity. Everyday objects or stores by using image based mobile interactions through their devices. For example, as users walk on the street, they might simply point the mobile camera to a store on the street to quickly access its related information, inquire special offers, and make reservations through their mobile devices without physically entering the store.Users depend on mobile device to maintain information and update.

OPSs (On Premises Signs) shows great visual diversity accompanied with complex environmental conditions. Consider an example, user walk on the street and he simply point his mobile camera to a store to quickly access its related information, inquire special offers, and make reservations through his mobile without physically entering on that store. Street view scenes are commonly captured by customers devices and they have more real-world characteristics lacking in most existing image datasets, e.g. perspective distortion, foreground and background clutter, etc. To learn a reliable OPS model for recognizing OPSs, a labeled dataset with a huge amount of real-scene images is required. However, precisely labeling OPS categories and regions, i.e., generating strong labels for Learning involves a significant amount of human labor, and thereby is usually not feasible for training a real-scene OPS model.

To learn an OPS model, our approach first exploits the visual saliency analysis to filter out predominant visual words of the background. After obtaining filtered code words, our approach further chooses the most discriminative visual words of each OPS category to enable real-world OPS recognition.

II. LITERATURE REVIEW AND RELATED WORK

The goal of learning and recognition of OPSs in real-world images can be shows as a problem of object recognition and localization. Firstly, understand the value of OPS, the importance of signage in the business community. D. Conroy studied and explained the role of the signage in businesses and organizations ji et al[1]

Bag-of-features technique used for content based image classification gaining to their less complexity and good performance. E. Nwark shows that for random sampling gives equal or better classifiers than the sophisticated multi-scale interest operators that are in common use for a representative selection of commonly used test databases and for moderate to large numbers of samples Chandrasekhar et al[2].

W. H. Cheng.et.al[3] introduced a novel framework for video adaptation based on content re-composition. The objective was to provide effective small size videos which intentioned the important aspects of a scene while faithfully retaining the background context. A generic video attention model was developed to extract user-interest objects, in which a high-level combination strategy was proposed for fusing the adopted three types of visual attention features: intensity, color, and motion Girod et al[4].

J.Harel et al[5] introduced a new visual model named as Graph-Based Visual Saliency (GBVS). It contains of two steps, first is creating activation maps on certain feature channels, and second is normalizing them in a manner which highlights clearness or brightness and to recognize combination with other maps. This model is less complex and biologically acceptable as it is naturally parallelized Zhanget al[6].

To discover association rules at multiple resolutions in order to identify frequent spatial configurations of local features that correspond to classes of logos appearing in real world scenes spatial pyramid mining technique is used Conroy et al[7].

To the best of our knowledge, due to all the learning and recognition of OPSs 1049 real-world characteristics of OPSs, none of these solutions can accurately recognize and localize real-world OPSs. Learning based approaches are then adopted as promising solutions Fergus et al[9]. Supervised learning is the mainstream paradigm. In modeling visual objects for recognition and localization Wuet al[10]. Yeh*et al*[11] addressed the problem of concurrent object recognition and localization according to the data-dependent region hypothesis. Wuetal[10] presented a semantics-preserving bag-of-words model by learning a distance metric to minimize the distance between the visual features with the same semantics. However, the prerequisite of all recognizable objects in the training data to be labeled with pixel-level precision often prevents it from practical applications. In contrast, unsupervised learning infers object models by using clustering techniques without any manual labeling, but a known limitation is the assumption of strong visual homogeneity on objects of the same category kim et al[12].

The task of recognizing and localizing OPSs in real-worldscenes can be viewed as a problem of real-world visual object recognition [8], The visual template based matching techniques exploit pre-defined patterns to discover the correspondences in given images [8], [10], [19]. For example, various researchers [8], [9], [11] proposed approaches to detect business logos/trademarks in real world scenes. To speed up the recognition operations, Romberg *et al.* [10] further developed a scalable recognition framework. Since there are hundreds and thousands of different OPSs in use nowadays [5], it is infeasible to collect all the visual templates in advance.

In response to this problem, several research projects devised approaches to detect texts [12]contained in objects (e.g., OPSs or products) to associate an OPS category with the identified corporate image. However, as the viewing angle of a camera changes, the texts might be significantly changed in their shapes due to perspective distortion or partially (or completely) occluded, and thereby cannot be well recognized by those existing approaches. Moreover, OPSs might also exhibit great diversity of visual appearance as shown in Figure 1(a), such as the variations in color, font style, and OPS size, which makes defining basic templates for an OPS unpractical as the number of OPSs scales up. To the best of our knowledge, due to all the real-world characteristics of OPSs (mentioned in Section I), none of these solutions can accurately recognize and localize real-world OPSs. Learning based approaches are then adopted as promising solutions. Supervised learning is the mainstream paradigm in modeling visual objects for recognition and localization [20]. Yeh *et al.* addressed the problem of concurrent object recognition and localization according to the data-dependent region hypothesis. Hoi *et al.* presented a semantics-preserving bag-of-words model by learning a distance metric to minimize the distance between the visualfeatures with the same semantics. However, the prerequisite of all recognizable objects in the training data to be labeled with pixel-level precision often prevents it from practical applications. In contrast, unsupervised learning infers object models by using clustering techniques without any manual labeling, but a known limitation is the assumption of strong visual homogeneity on objects of the same category.

III. PROBLEM STATEMENT:

Here it is proposed to address our problem basedon the communication theory. From a communication system perspective, an image can be regarded as a visual signal (a set of code words) to transfer information from its source (encoder) to a destination (decoder). It is generated by an OPS encoder (i.e. a random code word selector) and modulated by its corresponding OPS category *Ci*.

- **Definition 1** (**Visual Word Set)**. We formulate the problem of recognizing and localizing OPSs as a superpixel classification problem. A superpixel is a perceptually meaningful atomic region and can be represented by a visual code word which is discriminative with code words of other OPS categories.
- **Definition 2** (**OPS Codebook Generation**). For each OPS category *Ci* , the category alphabet (i.e. a codebook) _*Ci* used by the associated images of *Ci* can be constituted by the subset. However, as shown in Fig. 2, we can observe that not every code word in _*Ci* is discriminative of the category *Ci*. For example, code words in the intersection of the three category alphabets are common elements and unable to distinguish any category. In OPS images, these code words _*BG* often come from the common objects appearing in street view scenes, such as buildings, roads, pedestrians, and vehicles. To obtain a discriminative OPS model, the analysis tools of visual saliency, i.e., graph-based visual saliency (GBVS) [29], are employed to help possibly filter out the background (non-OPS) regions for reducing the number of noisy visual words extracted.
- **Definition 3** (**Discriminative OPS Model**). Given a set of *t* visual words {*v*1, *v*2, …, *vt* } extracted from *n* weakly-labeledstreet view images {*I*1, *I*2, …, *In*}, a set of discriminativecode words _+*Ci* for each OPS category *Ci* can be computedusing the distributional clustering [30]. After obtaining _+*Ci* for each OPS category *Ci* , the discriminative OPS model.

- **Definition 4** (**Superpixel Labeling**). Given an input image over-segmented into k superpixels Pi , a testing set $T = \{\{P1, l1\}, \{P2, l2\}, \ldots, \{Pi , li\} \ldots, \{Pk , lk\}\}$ can be arranged, where $li \in COPS = \{C1, C2, \ldots, Cm\}$ is thecorresponding OPS category of the superpixel Pi Further, this decision making function D should be consistent for superpixels of labeled or unseen images. In other words, this system classifies each superpixel Pi as being one of OPS categories Ci through a simple but effective decision-making mechanism based on the learnt discriminative OPS model M.

IV. PROPOSED WORK AND OBJECTIVES

The main motive of this research is to detecting the problem of real world OPS learning and recognition of weakly labeled street view images. Here, to tackle this problem by developing a framework based on distributional clustering, in which Tsung-hung tsai developed a framework which is used to exploit distributional information of each visual feature. OPS- 62 dataset demonstrated the performance of given approach over a latent semantic analysis model for more accurate recognition and also improvement in average recognition rate.

- **Probabilistic Frame work For Learning &Recognizing OPS:**

To implement the proposed approach two main algorithms are used:

1) Visual saliency based codebook generation of OPS categories.
2) OPS modeling and recognition using distributional clustering.

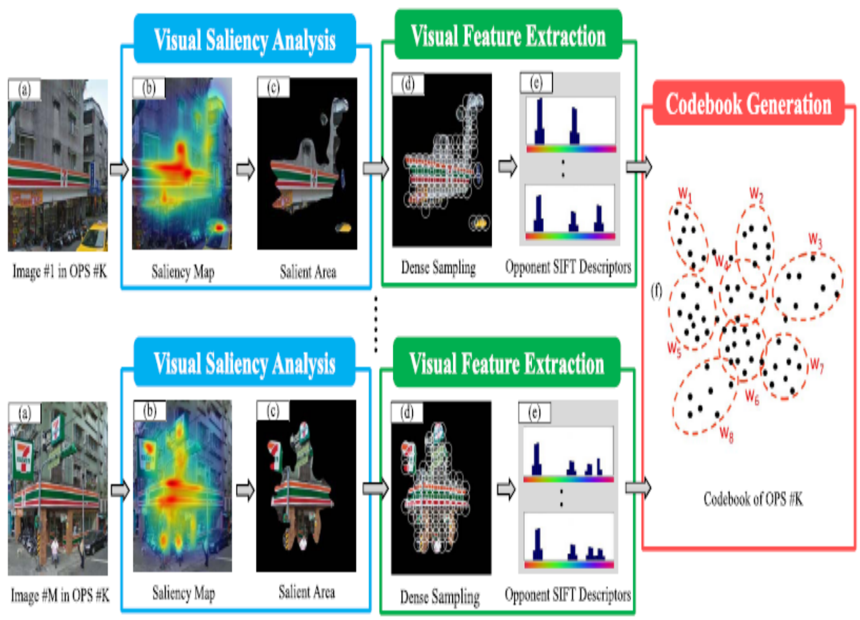

Fig. 1. The illustrative flowchart of visual saliency based codebook generation of an OPS category.

- **System Architecture of a Proposed Model:**

In this system, user gives input as image which is captured through mobile cameras. System gives input images and perform the actual proposed framework on given input image. In this framework two basic algorithms are used: first is visual saliency based codebook generation of OPS categories. In this algorithm first, filter out the background region for minimizing the number of noisy visual word using visual saliency analysis. After removing the background noise, visual feature are extracted using dense sampling strategy and

Opponent SIFT descriptor for codebook Generation. After acquiring a codebook for each OPS category compute a discriminative subset and apply a distributional clustering to collecting of all the code words in OPS categories into two disjoint clusters. Then allow the concurrent OPS recognition and localization in super-pixel level using obtained OPS and background models. Second algorithm is OPS modeling and recognition using distributional Clustering here, super pixel segmentation performed on input image.

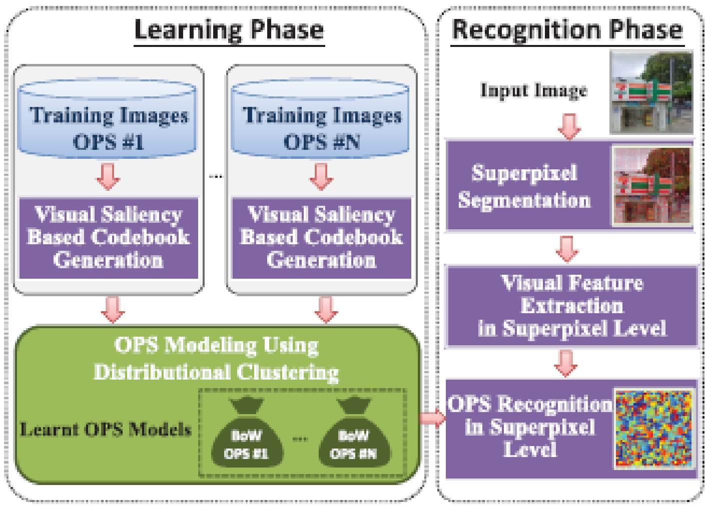

Fig. 2: System Architecture

After the segmentation visual feature extraction in super pixel level then recognized the OPS image with learned dataset. We now describe the re ranking of the returned images based on text and metadata alone. Here, we follow and extend the method proposed by Fergus et al[9] in using a set of textual attributes whose presence is a strong indication of the image content.

V. Conclusion:

Finally we conclude that our aim of this research is to detecting the problem of real world OPS learning and recognition of weakly labeled street view images,for precisely recognized feature of images Here, to tackle this problem by developing a framework based on distributional clustering, in which Tsung-hung tsai developed a framework which is used to exploit distributional information of each visual feature. OPS- 62 dataset demonstrated the performance of given approach over a latent semantic analysis model for more accurate recognition and also improvement in average recognition rate

References

[1] R. Ji, L.-Y. Duan, J. Chen, S. Yang, T. Huang, H. Yao, et al., "PKUBench: A context rich mobile visual search benchmark," in Proc. IEEE ICIP, Jun. 2012, pp. 2545–2548.

[2] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, et al., "The stanford mobile visual search data set," in Proc. 2nd ACM Conf. Multimedia Syst., Feb. 2011, pp. 117–122.

[3] B. Girod, V. Chandrasekhar, N.-M. C. David M. Chen, R. Grzeszczuk, Y. Reznik, et al., "Mobile visual search: Linking the virtual and physical worlds," IEEE Signal Process. Mag., vol. 28, no. 4, pp. 61–76, Jul. 2011.

[4] Y. Zhang, L. Wang, R. Hartley, and H. Li, “Where’s the weet-bix?” in Proc. 8th ACCV, 2007, pp. 800–810. [5] D. Conroy, What’s Your Signage (How On-Premise Signs Help Small Businesses Tap Into a Hidden Profit Center). New York, NY, USA: State Small Bus. Develop. Center, 2004.

[6] (2013). Social Shopping [Online]. Available: http://en.wikipedia.org/wiki/Social_shopping

[7] C.-W. You, W.-H. Cheng, A. W. Tsui, T.-H. Tsai, and A. Campbell, “MobileQueue: An image-based queue card retrieving system through augmented reality phones,” in Proc. 14th ACM Int. Conf. Ubiquitous Comput., 2012, pp. 1–2.

[8] J. Kleban, X. Xie, and W.-Y. Ma, “Spatial pyramid mining for logo detection in natural scenes,” in Proc. IEEE ICME, Apr. 2008, pp. 1077–1080.

[9] A. Joly and O. Buisson, “Logo retrieval with a contrario visual query expansion,” in Proc. 17th ACM Int. Conf. Multimedia, 2009, pp. 581–584.

[10] S. Romberg, L. G. Pueyo, R. Lienhart, and R. van Zwol, “Scalable logo recognition in real-world images,” in Proc. 1st ACM ICMR, 2011, pp. 1–25.

[11] J. Revaud, M. Douze, and C. Schmid, “Correlation-based burstiness for logo retrieval,” in Proc. 20th ACM Int. Conf. Multimedia, 2012, pp. 965–968.

[12] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, et al., “Automatic detection and recognition of Korean text in outdoor signboard images,” Pattern Recognit. Lett., vol. 31, no. 12, pp. 1728–1739, Sep. 2010.

[13] A. Zamir, A. Darino, and M. Shah, “Street view challenge: Identification of commercial entities in street view imagery,” in Proc. 10th ICMLA Workshops, vol. 2. 2011, pp. 380–383.

[14] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, “Learning object categories from internet image searches,” Proc. IEEE, vol. 98, no. 8, pp. 1453–1466, Aug. 2010.

[15] S. Vijayanarasimhan and K. Grauman, “Large-scale live active learning: Training object detectors with crawled data and crowds,” in Proc. IEEE Conf. CVPR, Aug. 2011, pp. 1449–1456

[16] P. Siva and T. Xiang, “Weakly supervised object detector learning with model drift detection,” in Proc. IEEE ICCV, Nov. 2011, pp. 343–350.

[17] N. Pinto, D. D. Cox, and J. J. DiCarlo, “Why is real-world visual object recognition hard?” PLoS Comput. Biol., vol. 4, no. 1, pp. 1–6, Jan. 2008.

[18] T. Yeh, J. Lee, and T. Darrell, “Fast concurrent object localization and recognition,” in Proc. IEEE Conf. CVPR, Jun. 2009, pp. 280–287.

[19] W.-L. Zhao and C.-W. Ngo, “Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection,” IEEE Trans. Image Process., vol. 18, no. 2, pp. 412–423, Feb. 2009.

[20] L. Wu, S. Hoi, and N. Yu, “Semantics-preserving bag-of-words models and applications,” IEEE Trans. Image Process., vol. 19, no. 7, pp. 1908–1920, Jul. 2010.

[21] G. Kim, C. Faloutsos, and M. Hebert, “Unsupervised modeling of object categories using link analysis techniques,” in Proc. IEEE Conf. CVPR, Jun. 2008, pp. 1–8.

[22] F. Schroff, A. Criminisi, and A. Zisserman, “Harvesting image databases from the web,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 4, pp. 754–766, Apr. 2010.

[23] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, “Using multiple segmentations to discover objects and their extent in image collections,” in Proc. IEEE Conf. CVPR, Jun. 2006, pp. 1605–1614. [24] L.-J. Li and L. Fei-Fei, “OPTIMOL: Automatic online picture collection via incremental model learning,” Int. J. Comput. Vis., vol. 88, no. 2, pp. 147–168, 2010.

[25] (2011). ICMLA StreetView Recognition Challenge [Online]. Available: http://www.icmla-conference.org/icmla11/challenge.htm

[26] M. Capdevila and O. W. M. Florez, “A communication perspective on automatic text categorization,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 7, pp. 1027–1041, Jul. 2009.

[27] S. Haykin, Communication Systems. New York, NY, USA: Wiley, 2000.

[28] Y.-H. Kuo, H.-T. Lin, W.-H. Cheng, Y.-H. Yang, and W. H. Hsu, “Unsupervised auxiliary visual words discovery for large-scale image object retrieval,” in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 905–912.

[29] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in Proc. ANIPS, 2007, pp. 545–552.

[30] L. D. Baker and A. K. McCallum, “Distributional clustering of words for text classification,” in Proc. 21st ACM ACM SIGIR Conf. Res. Develop. Inf. Retr., 1998, pp. 96–103.

[31] E. Nowak, F. Jurie, and B. Triggs, “Sampling strategies for bag-offeatures image classification,” in Proc. ECCV, 2006, pp. 490–503.

[32] K. van de Sande, T. Gevers, and C. Snoek, “Evaluating color descriptors for object and scene recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

[33] W.-H. Cheng, C.-W. Wang, and J.-L. Wu, “Video adaptation for small display based on content recomposition,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 1, pp. 43–58, Jan. 2007.

[34] B. Alexe, T. Deselaers, and V. Ferrari, “What is an object?” in Proc. IEEE CVPR, Jun. 2010, pp. 73–80. [35] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer-Verlag, 2006.