# Geo-Location And Information Retrieval For On-Premise Signs

**Charulata**

Asst. Professor, Dept. of CSE, SJBIT, charulata@sjbit.edu.in

**Abstract:** Image recognition has become an integral and important part of today's technical world. The various application scenarios give rise to a key technique of daily life visual object recognition. On-premise signs (OPSs), a popular form of commercial advertising, are widely used in our living life. The OPSs often exhibit great visual diversity (e.g., appearing in arbitrary size), accompanied with complex environmental conditions (e.g., foreground and background clutter). Observing that such real-world characteristics are lacking in most of the existing image data sets, in this paper, we first proposed an OPS data set, in which comprises of OPS images of different businesses which are basically collected from Google's Street View. Further, for addressing the problem of real-world OPS learning and recognition, we developed a probabilistic framework based on the distributional clustering, in which we proposed to exploit the distributional information of each visual feature (the distribution of its associated OPS labels) as a reliable selection criterion for building discriminative OPS models. This approach is simple, linear, and can be executed in a parallel fashion, making it practical and scalable for large-scale multimedia applications. This project provides very simple and modified features, which are very easy to view and operate. This project is designed and organized in a very simplified manner to hold the details of images and return the geo-location of the image.

**Index Terms:** Real-world objects, street view scenes, learning and recognition, object image data set.

## 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 rely on mobile devices to maintain an always-on relation to information and personal networks and thereby can access in-situ information related to nearby 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. To build a recognizable image for a business to attract customers, each store has its own on-premise sign (defined as *OPS* hereafter), which is a visually consistent image for a brand and contains a mixture of text (e.g. the business's name) and graphics (e.g. corporate trademarks/logos). For example, shows the OPSs attached to building or properties operated by Starbucks. Recognizing these OPSs as a means to identify different stores can help to retrieve nearby store information for enabling a new class of multimedia applications, such as social shopping, intelligent navigation, mobile visual search, and mobile augmented reality. Therefore, this study proposes a probabilistic framework for learning and recognizing OPSs in real scene images. In the literature, previous researchers have developed recognition techniques to identify business brands *OPS* (defined as *categories* hereafter) by recognizing logo or trademark images or texts, contained in OPSs. However, the visual contents of a business's OPSs are not required to be limited with logo/trademark images or texts. In addition, even though businesses tend to keeps the visual design of their OPSs consistent while being distinguishable from those of other brands, they might flexibly adapt the OPS design to fit the practical conditions when deploying the OPSs. Thus, even for those OPSs of the same business, they might still exhibit great diversity of visual appearance, such as the variations in color, font style, and OPS size. Further, the spatial geometry of the visual components also varies, e.g. some of the OPS texts are kept in one line but others are separated into multiple lines. Real-world characteristics of erected OPSs, such as arbitrary size, position, viewing angles, perspective distortion, occlusions, varying lighting conditions, foreground and background clutter, etc., make logos, trademarks, or texts in OPSs occupy a relatively smaller area or be occluded by other objects in real scene images. These characteristics make the existing solutions fail to identify logos, trademarks, or texts in OPSs. On the contrary, our approach exploits a probabilistic framework to extract discriminative visual words of each OPS category, and therefore is able to recognize and localize each OPS within images by using the learnt OPS model. In particular, 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, that is, 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. Instead of generating strong labels for real-scene images, we resort to an alternative learning technique, which is weakly supervised by a dataset with each image labeled with the OPS category it contains, that is, a weakly labeled image. For this purpose, we first proposed an OPS dataset, namely OPS-62 in which totally 4,649 OPS images of 62 different businesses (categories) are collected from Google's Street View and weakly labeled, that is, all training images of an OPS category are known to contain the corresponding OPS but no precise labeling of location in the images is needed. As illustrated before, the OPS-62 dataset demonstrates challenging real-world characteristics as mentioned earlier, and it allows the co-existence of multiple OPSs in a single image. Sample images and more details of the OPS-62 dataset are referred. 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. As a summary, the main contributions of this work include:

1) We develop a probabilistic framework for real-world object recognition from a communication system perspective. The proposed approach requires no pixel wise object labeling in the learning phase and the

recognition scheme is simple, linear, and can be executed in a parallel fashion.

2) We suggest a visual saliency based procedure to reduce noisy visual words while generating the codebook of object categories.
3) We propose to learn reliable object models by employing the distributional clustering to measure the discriminative of code words with respect to each object category.
4) This study creates a new benchmark for object recognition, OPS data set, which has more real-world characteristics lacking in most existing image datasets.

The authors have proposed the system to address the problem based on the communication theory. From a communication system perspective, an image can be regarded as a visual signal which is nothing but a set of code words to transfer information from its source which is encoder to a destination which is the decoder. It is generated by an OPS encoder that is a random code word selector and modulated by its corresponding OPS category. Understanding the term Visual Word Set, authors consider a set which denote the universal set of code words that is the signal alphabet. Then, the visual word set of an OPS imagecan be extracted using bag-of-words (BoW) models, which is represented as a set which includes visual word at a super pixel region, and *p* is the number of super pixels. The problem is formulated by recognizing and localizing OPSs as a super pixel classification problem. A super pixel 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. Further understanding, OPS Codebook Generation, authors consider OPS category, the category alphabet that is nothing but a codebookused by the associated imagescan be constituted by the subset. However, as it is evident, we can observe that not every code word is discriminative of the category. 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 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, that is graph-based visual saliency (GBVS), are employed to help possibly filter out the background (non-OPS) regions for reducing the number of noisy visual words extracted. Coming to Discriminative OPS Model, considering a set of visual words extracted fromweakly labeled street view images which is a set of discriminative code words for each OPS category will be computed using the distributional clustering. Understanding Superpixel Labeling, considering an input image over-segmented intosuperpixels, a testing set can be arranged, is the corresponding OPS category of the superpixel. This system maps a superpixel intoOPS categoriesthrough a learnt decision making function.

## II. RELATED WORK

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews use secondary sources, and do not report new or original experimental work. The goal of literature survey is to understand the positions of other academics who have studied the problem/issue that is under consideration and include that in paper or project. This can done by comparing and contrasting, simple summarization, or anynumber of ways that show that some research concerning the considered problem/project is done. The task of recognizing and localizing OPSs in real-world scenes can be viewed as a problem of real-world visual object recognition. The visual template based matching techniques exploit pre-defined patterns to discover the correspondences in given images. The task of recognizing and localizing OPSs in real-worldscenes can be viewed as a problem of real-world visual object recognition [8], [17], [18]. 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], [13] 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 such as the variations in color, font style, and OPS size, which makes defining basic templates for an OPS unpractical as the number ofOPSs scales up. To the best of our knowledge, due to all thereal-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 [14], [20], [21]. Supervised learning is the mainstream paradigm in modeling visual objects for recognition and localization [20]. Yeh*et al.* [18] addressed the problem of concurrent object recognition and localization according to the data-dependent region hypothesis. Hoi *et al.* [20] 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 [21]. A compromise is to learn the object models from weakly labeled images [14], [15]. For example, Vijayanarasimhan*et al.* [15] presented an SVM based active learning approach for training online object detectors but user intervention is required. A leading approach in this field is using probabilistic Latent Semantic Analysis (pLSA) [14], [22]. Russell *et al.* [23] searched for objects by applying multiple segmentations to each image and employed pLSA and Latent Dirichlet Allocation (LDA) to discover the latent topics. Further, there are many studies gathering training and testing data from the web resources such as Google and Flickr search engines [24]. Fergus *et al.* [14] adopted images from the Google search engine and applied pLSA models to recognize the corresponding object categories. In addition, Schroff*et al.* [22] exploited the text, metadata, and visual cues to train a robust classifier and build a database for image recognition. Moreover, Li *et al.* [24] proposed a framework which collects web images and learns the class models iteratively such that the collected image dataset can grow larger and the learnt models can become more robust simultaneously. However, there are some open and common issues with the related schemes. For example, due to the unsupervised nature of latent variable models [14], [22], it is not a trivial task to choose the number of latent topics to use and how to pick the representative topics of the recognizable objects. Also, the objects are required to occupy a minimum proportion of the associated image, or most latent topics will be largely polluted by the noisy visual features and no reliable object models can be obtained [14]. Based on the above discussions, we know that learning object models from weakly labeled images is a practical way to deal with large scale object recognition. However, most of the existing approaches would fail to address street view images containing real-world objects in great visual diversity, such as an often but challenging case that objects are visible at small size with a very cluttered background. Motivated by the communication theory of transmitting signals over a noisy channel, we proposed an alternative solution based on the distributional clustering to overcome the drawbacks in the previous proposed systems.

## III. PROBLEM STATEMENT

**Visual Word Set**: Let $\Omega = \{w1, w2, \ldots, wj, \ldots, wN\}$ denote the universal set of code words $wj$, i.e. the signal alphabet. Then, the visual word set of an OPS image $It$ can be extracted using bag-of-words (BoW) models, which is represented as a set $Sit = \{s1, s2, \ldots, sr, sr+1, \ldots, sp\}$, where $sr \in \Omega$ a visual word at a superpixel region $r$ , and $p$ is the number of superpixels in $It$ . 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.

**OPS Codebook Generation**: For each OPS category $Ci$, the category alphabet (i.e. a codebook) $\Omega Ci$ used by the associated images of $Ci$ can be constituted by the subset of $\Omega$. 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 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.

**Discriminative OPS Model**: Given a set of $t$ visual words $\{v1, v2, \ldots, vt\}$ extracted from $n$ weakly-labeled street view images $\{I1, I2, \ldots, In\}$, a set of discriminative code words $\Omega+Ci$ for each OPS category $Ci$ can be computed using the distributional clustering [30]. After obtaining $\Omega+Ci$ for each OPS category $Ci$ the discriminative OPS model can be created.

**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 the corresponding OPS category of the superpixel$Pi$ . This system maps a superpixel$Pi$ into $m$ OPS categories $COPS$ through a learnt decision making function $D(Pi\ ) = l \in COPS$. 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 $M$ Once the system labels all superpixels in an input image,an OPS location can be determined by finding a boundary enclosing a set of clustered superpixels which are classified as being the same OPS category.

## IV. WORKING

The system is proposed to have the following modules:

**On-Premise Signs:** In this Module, These are signs that are located on the same premises on which the activity is conducted. Any property, on which a sign is placed, that is not integral to the activity, or is separated from the activity by a roadway, highway, common driveway, or other obstruction, or is at such distance that the sign is closer to the highway than the activity is not considered on-premises. Also, if the sign is located on a narrow strip of land whose only real purpose is to accommodate the sign, and is not used for the advertised activity, the sign cannot be considered on-premises. These rules apply regardless of whether the properties are under the same ownership. On-premises signs in the controlled area may be subject to registration.

**Image Data Set**: Instead of generating strong labels for real-scene images, we resort to an alternative learning technique, which is weakly supervised by a dataset with each image labeled with the OPS category it contains, i.e., we, learning involves a significant amount of human labor, and thereby is usually not feasible for training a real-scene OPS model. Instead of generating strong labels for real-scene images, we resort to an alternative learning technique, which is weakly supervised by a dataset with each image labeled with the OPS category it contains, i.e., a weakly labeled image , learning involves a significant amount of human labor, and thereby is usually not feasible for training a real-scene OPS model. Instead of generating strong labels for real-scene images, we resort to an alternative learning technique, which is weakly supervised by a dataset with each image labeled with the OPS category it contains, i.e., a weekly labeled image akly labeled image.

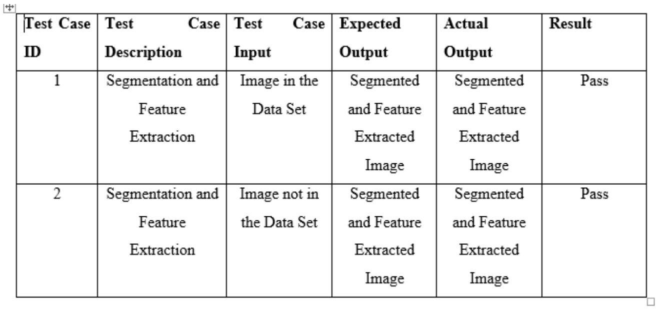
**Recognition:** The task of recognizing and localizing OPSs in real-world scenes can be viewed as a problem of real-world visual object recognition which is a visually consistent image for a brand and contains a mixture of text (e.g. the business's name) and graphics (e.g. corporate trademarks/logos).Nature of digital information has become increasing visual, and so has the need for companies to locate and identify in the digital ocean. Explore what the industry leader in image recognition technology has to say about making sense of visual content in this digital world.

**Learning**: Learning is the act of acquiring new, or modifying and reinforcing, existing knowledge, behaviors, skills, values, or preferences and may involve synthesizing different types of information. The ability to learn is possessed by humans, animals and some machines. Progress over time tends to follow learning curves. Learning is not compulsory; it is contextual. It does not happen all at once, but builds upon and is shaped by what we already know. To that end, learning may be viewed as a process, rather than a collection of factual and procedural knowledge. Learning produces changes in the organism and the changes produced are relatively permanent.

## V. EXPERIMENTAL RESULTS

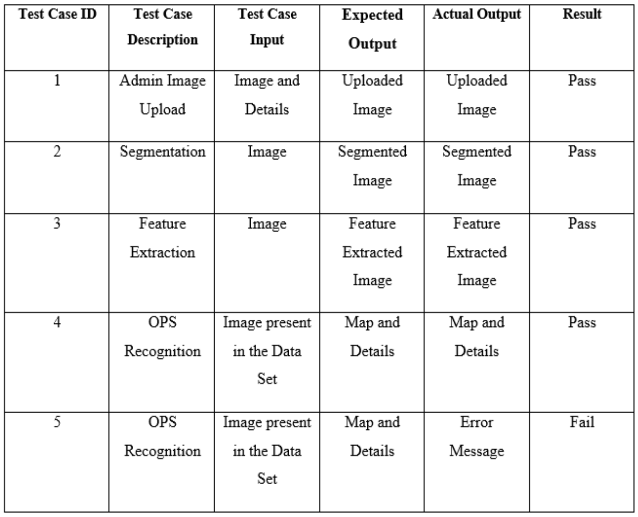
Objects in real-world images are often in arbitrary size, position, and viewing angles, accompanied with perspective distortion, foreground and background clutter, varying lighting conditions, etc [3], [17]. Observing that such real-world characteristics are lacking in most of the existing imagedatasets [2], [3], we focus on the street view scenes containing commercial advertising OPSs

.



| Test Case ID | Test Case Description | Test Case Input | Expected Output | Actual Output | Result |
|---|---|---|---|---|---|
| 1 | Segmentation and Feature Extraction | Image in the Data Set | Segmented and Feature Extracted Image | Segmented and Feature Extracted Image | Pass |
| 2 | Segmentation and Feature Extraction | Image not in the Data Set | Segmented and Feature Extracted Image | Segmented and Feature Extracted Image | Pass |

As it is evident from the table above, images given as the input for the proposed system which are present in the dataset give positive results.



| Test Case ID | Test Case Description | Test Case Input | Expected Output | Actual Output | Result |
|---|---|---|---|---|---|
| 1 | Admin Image Upload | Image and Details | Uploaded Image | Uploaded Image | Pass |
| 2 | Segmentation | Image | Segmented Image | Segmented Image | Pass |
| 3 | Feature Extraction | Image | Feature Extracted Image | Feature Extracted Image | Pass |
| 4 | OPS Recognition | Image present in the Data Set | Map and Details | Map and Details | Pass |
| 5 | OPS Recognition | Image present in the Data Set | Map and Details | Error Message | Fail |

The above table gives detailed results pertaining to each module. Input is given in the form of images to segmentation and feature extraction modules and it has proved to be a success.

## VI. CONCLUSION

In this work, we proposed a probabilistic framework for learning and recognizing real-world OPSs from weakly labeled street view images, in which the technique of distributional clustering is exploited to benefit the selection of discriminative visual words and the construction of effective OPS models, as motivated by the communication theory. Meanwhile, we proposed the OPS image dataset which contains more real world characteristics as a new benchmark for visual object recognition. However, in view of the low average recall values relatively, the OPS recognition in real-world scenes is still a challenging problem.

## 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: StateSmall 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. UbiquitousComput., 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] 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 burstinessfor 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. Let, vol. 31, no. 12, pp. 1728–1739, Sep. 2010.

[13] Zamir, A. Darino, and M. Shah, "Street view challenge: Identification of commercial entities in street view imagery," in Proc. 10th ICMLAWorkshops, 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?"PLoSComput. 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.