ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Literature Survey of Systems for Object Detection in Unconstrained Video Sequences

Shyamal D. Patil[1], Sonali Rangdale[2]

ME Student, Dept. of I.T., Siddhant College of Engineering, Savitribai Phule Pune University, Pune, India[1]

Professor, Dept. of I.T., Siddhant College of Engineering, Savitribai Phule Pune University, Pune, India[2]

ABSTRACT: Automatic detection objects from a video feed for multiple purposes had been addressed by researchers in varied technological ways. In the early age of this research some of them focused on detecting the object from video by assuming it as a unrelated sequence of static image frames. With further development many researchers considered using additional information available from the video frame. However not many considered the spatio temporal linkage of frames of the same video. With advances in pattern recognition algorithms and further evolution in the field multicore product detection framework is proposed. Present paper provides a summary of sequential development in the field of video based object detection since earlier reported articles in the literature till some of the state-of-the-art techniques.

KEYWORDS: object detection, video sequence, smart TV, multicore product detection framework, spatio-temporal information

I. Introduction

It's been a while since online shopping has begun replacing the retail shopping market. The ease of online shopping attracted the customers due to the fact that they can now easily sit in their homes surfing for products on internet browser and apply for shopping at their will while sitting on their couches. If one thinks with a little more detail and inspects the online shopper mentality, there is a point that most of the times, the products which shoppers see on television soaps, films or documentaries will remain in their minds as a hunch to buy. The shoppers then go online searching for these items, precisely exactly the same item. However, most often it's a disappointing scenario that builds when the shopper don't find the exact thing that they are looking for, which perhaps they had seen in some of the films or videos while watching the television. There goes the idea that if the shoppers could get the complete buying information of any such article when it flashes on their television screens. And that's one of the reasons why many of researchers and engineers chasing the goal of developing a system that can detect an object in an unconstrained video, such as television footage.

For example, consider a case of a potential buyer watching a video such as a fashion show on television. He/she likes a particular hat as seen wore by a model. Now how he/she can find out the buying information of that hat. Is it possible for the television to software to provide it? Well, probably a few years ago the answer to the question could have been no. However, with the launch of smart television sets in market this can be possible. A smart TV, sometimes referred to as connected TV or hybrid TV, is a television set or set-top box with integrated Internet and Web 2.0 features, and is an example of technological convergence between computers and television sets and set-top boxes. Besides the traditional functions of television sets and set-top boxes provided through traditional broadcasting media, these devices can also provide Internet TV, online interactive media, over-the-top content, as well as on-demand streaming media, and home networking access. In smart TVs, the operating system is preloaded or is available through set-top box. The software applications or apps can be preloaded into the device, or updated or installed on demand via an app store or app marketplace, in a similar manner to how the apps are integrated in modern smartphones. The technology that enables smart TVs is also incorporated in external devices such as set-top boxes, Blu-ray players, game consoles, digital media players, hotel television systems and phones and other network connected interactive devices that utilize television type display outputs. These devices allow viewers to search, find and play videos, movies, photos and other content from the Web, on a cable TV channel, on a satellite TV channel, or on a local storage drive. In summary the smart television opens the access and possibilities of interactive TV watching.

DOI: 10.15680/IJIRCCE.2015. 0312052

However, the simply internet surfing on the television or accessing and using many other open market shared software applications is not the end of smart TV. The viewers are still not having the full interactive TV viewing experience. Going back to our base idea of television enabled shopping, it needs the objects from the smart television video footage to be detected and then later, say after the end of programme or with a notifier the viewer can have a choice of going through list of detected items or articles with their full possible buying information available at the perusal of the viewer or now the potential shopper. And therefore, in the modern age, many of the engineers in the field of computer science, information technology and consumer electronics are looking forward to the solution of the problem of object detection from an unconstrained video.

Present paper is a literature survey paper, going through a list of advances in this problem. The paper intends to put forward a valuable summary of many noteworthy efforts as published in literature in this field for the use of industries and the developers. Section II presents the forwarded literature survey.

II. Related work

One can easily comprehend the idea of object detection from the video, assuming the video as a sequence of slower moving quasi-steady sequence of still frames. And exactly this had been the initial effort by researchers. Later on, researchers moved ahead with the ideas of considering a spatio-temporal movement of frames or images from the video sequence. And recently there had been few successful efforts showing object detection in a continuous video using multi-cue algorithms. In present section, reviewing the literature one shall see a section wise classified approach to understand better the evolution of advances.

A. *Object Detection from video as sequence of unrelated static frames*

The object detection task can be addressed by considering the video as an unrelated sequence of frames and perform static object detection In 2009, Felzenszwalb et al. [1] described an object detection system based on mixtures of multiscale deformable part models. Their system was able to represent highly variable object classes and achieves state-of-the-art results in the PASCAL object detection challenges. They combined a margin-sensitive approach for data-mining hard negative examples with a formalism we call latent SVM. This led to an iterative training algorithm that alternates between fixing latent values for positive examples and optimizing the latent SVM objective function. Their system relied heavily on new methods for discriminative training of classifiers that make use of latent information. It also relied heavily on efficient methods for matching deformable models to images. The described framework allows for exploration of additional latent structure. For example, one can consider deeper part hierarchies (parts with parts) or mixture models with many components.

Leibe et al. [2] in 2007, presented a novel method for detecting and localizing objects of a visual category in cluttered real-world scenes. Their approach considered object categorization and figure-ground segmentation as two interleaved processes that closely collaborate towards a common goal. The tight coupling between those two processes allows them to benefit from each other and improve the combined performance. The core part of their approach was a highly flexible learned representation for object shape that could combine the information observed on different training examples in a probabilistic extension of the Generalized Hough Transform. As they showed, the resulting approach can detect categorical objects in novel images and automatically infer a probabilistic segmentation from the recognition result. This segmentation was then in turn used to again improve recognition by allowing the system to focus its efforts on object pixels and to discard misleading influences from the background. Their extensive evaluation on several large data sets showed that the proposed system was applicable to a range of different object categories, including both rigid and articulated objects. In addition, its flexible representation allowed it to achieve competitive object detection performance already from training sets that were between one and two orders of magnitude smaller than those used in comparable systems.

Recently in last decade, methods based on local image features have shown promise for texture and object recognition tasks. Zhang et al. [3] in 2006, presented a large-scale evaluation of an approach that represented images as distributions (signatures or histograms) of features extracted from a sparse set of key-point locations and learnt a Support Vector Machine classifier with kernels based on two effective measures for comparing distributions. They first evaluated the performance of the proposed approach with different key-point detectors and descriptors, as well as different kernels and classifiers. Then, they conducted a comparative evaluation with several modern recognition methods on 4 texture and 5 object databases. On most of those databases, their implementation exceeded the best reported results and achieved comparable performance on the rest. Additionally, we also investigated the influence of background correlations on recognition performance.

In 2001, Viola and Jones [4] in a conference on pattern recognition described a machine learning approach for visual object detection which was capable of processing images extremely rapidly and achieving high detection rates. Their work was distinguished by three key contributions. The first was the introduction of a new image representation called the "integral image" which allowed the features used by their detector to be computed very quickly. The second was a learning algorithm, based on AdaBoost, which used to select a small number of critical visual features from a larger set and yield extremely efficient classifiers. The third contribution was a method for combining increasingly more complex classifiers in a "cascade" which allowed background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The cascade could be viewed as an object specific focus-of-attention mechanism which unlike some of the previous approaches provided statistical guarantees that discarded regions were unlikely to contain the object of interest. They had done some testing over face detection where the system yielded detection rates comparable to the best of previous systems. Used in real-time applications, the detector runs at 15 frames per second without resorting to image differencing or skin color detection.

In 2000, Weber et al. [5] proposed a method to learn heterogeneous models of object classes for visual recognition. The training images, that they used, contained a preponderance of clutter and the learning was unsupervised. Their models represented objects as probabilistic constellations of rigid parts (features). The variability within a class was represented by a join probability density function on the shape of the constellation and the appearance of the parts. Their method automatically identified distinctive features in the training set. The set of model parameters was then learned using expectation maximization. When trained on different, unlabeled and non-segmented views of a class of objects, each component of the mixture model could adapt to represent a subset of the views. Similarly, different component models could also specialize on sub-classes of an object class. Experiments on images of human heads, leaves from different species of trees, and motor-cars demonstrated that the method works well over a wide variety of objects.

B. *Object Detection using additional information from video progression sequence*

On the other hand, the problem of object detection can be tackled by utilizing the additional information offered by the progression of the video sequence. Ayvaci and Soatto [6] in 2012, presented a method for detecting objects in a scene. While functionally important properties such as graspability cannot be ascertained from passive imaging data, they had defined properties that a moving image of an object must have in order to correlate to topological properties of the scene, such as being partially surrounded by the medium. As they showed, occlusions play a key role in the detection of detachable objects. Leveraging some of the prior work, they also showed that once (binary) occlusion regions are available, integrating local ordering information into a coherent depth ordering map could be achieved by simple linear programming. The key to their approach was to convert a supervised segmentation problem into an unsupervised one using occlusions as the supervision mechanism. As a result, they had a fully unsupervised method for detecting and segmenting an unknown number of objects, and estimating their number in the meantime, all by solving a linear program. Despite their efforts to manage errors in the occlusion detection stage, they still suffered from complete failures of the occlusion detection mechanism. In many cases, it was due to insufficient motion in the scene, and the results improve under extended temporal observation. However, results as they demonstrated could still be useful as initialization of a more involved optimization over an extended temporal observation. Their approach had a few tuning parameters, but fewer than most competing schemes since they performed model selection. Also, the prescribed approach shared the limitation of all schemes that break down the original problem (detached object detection) into a number of sequential steps, whereby failure of the early stages of processing caused failure of the entire pipeline.

It is a great challenge to detect an object that is overlapped or occluded by other objects in images. For moving objects in a video sequence, their movements can bring extra spatio-temporal information of successive frames, which helps object detection, especially for occluded objects. In 2011, at a multimedia conference, Liu et al. [7] proposed a moving object detection approach for occluded objects in a video sequence with the assist of the SPCPE (Simultaneous Partition and Class Parameter Estimation) unsupervised video segmentation method. Based on the preliminary foreground estimation result and object detection information from the previous frame, an n-steps search (NSS) method was utilized to identify the location of the moving objects, followed by a size-adjustment method that adjusted the bounding boxes of the objects. Their conducted several experimental results showed that the proposed approach achieved good detection performance under object occlusion situations in serial frames of a video sequence.

As one can apprehend, the detection of moving objects in complex environments with various types of motion is a difficult problem because the camera motion and the object motion are mixed. Kim et al. [8] in 2010 proposed a moving object detection algorithm that used motion clustering and classification from only two consecutive image frames captured by a free-moving camera. The proposed moving object detection had no assumption about the camera motion and the environmental conditions. Their experimental results showed that the proposed moving object detection was accurate within an accuracy of 7 pixels on the average.

Most global motion estimation (GME) methods are oriented to video coding while video object segmentation methods either assume no global motion (GM) or directly adopt a coding-oriented method to compensate for GM. In 2008, Qi et al. [9] proposed a robust fast GME method for video object segmentation with 1) a fast initial estimate using a combination of 3-step search and MV prediction, 2) a robust estimate using object information, and 3) a robust estimate considering neighborhood to eliminate outliers. Both subjective and objective results demonstrated that the proposed method was more robust, faster, and more suitable for object segmentation than some of the referred methods.

C. *Object Detection using additional information from co-occurrence and/or spatial relationships between object labels*

Most of the approaches as mentioned so far fail to detect complex objects and perform well only on ideal conditions. In the case of video based approaches, these mostly concentrate on using motion information to detect moving objects, which may not work well for difficult objects; in the case of hat, as talked about in introduction, the motion of the hat would be masked by the motion of the person wearing the hat. The significant challenge posed by complex objects gives rise to using additional information to improve the detection results. For still images, most approaches employ additional information from co-occurrence and/or spatial relationships between object labels.

Kumar and Hebert [10] from the Carnegie Mellon University in 2005 presented a two-layer hierarchical formulation to exploit different levels of contextual information in images for robust classification. Each layer was modeled as a conditional field that allowed one to capture arbitrary observation dependent label interactions. The proposed framework had two main advantages. First, it use to encode both the short-range interactions (e.g., pixel wise label smoothing) as well as the long-range interactions (e.g., relative configurations of objects or regions) in a tractable manner. Second, the formulation was general enough to be applied to different domains ranging from pixel wise image labeling to contextual object detection. The parameters of the model were learned using a sequential maximum-likelihood approximation. The benefits of their proposed framework were demonstrated on four different datasets and comparison results were presented.

The sliding window approach of detecting rigid objects (such as cars) is predicated on the belief that the object can be identified from the appearance in a small region around the object. Other types of objects of amorphous spatial extent, for example trees, sky, etc. however, are more naturally classified based on texture or color. In 2008, Heitz and Koller [11] from Stanford university, in one of their project presented at conference, combined recognition of these two types of objects into a system that leverages ``context" toward improving detection. In particular, they clustered image regions based on their ability to serve as context for the detection of objects. Rather than providing an explicit training set with region labels, their method automatically grouped regions based on both their appearance and their relationships to the detections in the image. They showed that their things and stuff (TAS) context model produces meaningful clusters that are readily interpretable, and helps improve their detection ability over state-of-the-art detectors. They also presented a method for learning the active set of relationships for a particular dataset. They presented results on object detection in images from the PASCAL VOC 2005/2006 datasets and on the task of overhead car detection in satellite images, demonstrating significant improvements over state-of-the-art detectors.

Torralba et al. [12] of MIT in 2010, discussed one approach for combining local and global features in visual object detection and localization. However, their system was also not that perfect. For example, sometimes objects used to appear out of context and might be accidently eliminated if the local evidence was ambiguous. The only way to prevent this was if the local detector gives a sufficiently strong bottom-up signal. Conversely, if the detector makes a false positive error in a contextually plausible location, it would not be ruled out by their system. However, their argument was that even people can also suffer from such “hallucinations”. In more general terms, they projected their system as a good example of probabilistic information fusion, an approach which is widely used in other areas such as speech recognition, which combines local acoustic models which longer-range language models.

Wang et al. [13] of China in their paper of 2011 presented a novel approach for multiclass object detection by combining local appearances and contextual constraints. They first constructed a multiclass Hough forest of local

DOI: 10.15680/IJIRCCE.2015. 0312052

patches, which could well deal with multiclass object deformations and local appearance variations, due to randomization and discrimination of the forest. Then, in the object hypothesis space, a new multiclass context model was proposed by them to capture relative location constraints, disambiguating appearance inputs in multiclass object detection. Finally, multiclass objects were detected with a greedy search algorithm efficiently. Their experimental evaluations on two image data sets showed that the combination of local appearances and context achieved the performance level as good as other modern systems in multiclass object detection.

Context is critical for reducing the uncertainty in object detection. However, context modeling is challenging because there are often many different types of contextual information coexisting with different degrees of relevance to the detection of target object(s) in different images. It is therefore crucial to devise a context model to automatically quantify and select the most effective contextual information for assisting in detecting the target object. Nevertheless, the diversity of contextual information means that learning a robust context model requires a larger training set than learning the target object appearance model, which may not be available in practice. Zheng et al. [14] in their work, proposed a novel context modeling framework without the need for any prior scene segmentation or context annotation. They formulated a polar geometric context descriptor for representing multiple types of contextual information. In order to quantify context, they proposed a new maximum margin context (MMC) model to evaluate and measure the usefulness of contextual information directly and explicitly through a discriminant context inference method. Furthermore, to address the problem of context learning with limited data, they exploited the idea of transfer learning based on the observation. Two novel context transfer learning models were proposed by them which utilized training samples from source object classes to improve the learning of the context model for a target object class based on a joint maximum margin learning framework. They conducted experiments on PASCAL VOC2005 and VOC2007 data sets, a luggage detection data set extracted from the i-LIDS data set, and a vehicle detection data set extracted from outdoor surveillance footage. The results validated the effectiveness of the proposed models for quantifying and transferring contextual information, and demonstrated that they outperformed related alternative context models.

D. *Object Detection using spatio-temporal information as an cue*

The utilization of additional information has been approached in recent years to try to overcome the challenges posed by object detection. Plenty of approaches tackle this task in static images utilizing co-occurrence and/or spatial relationships. However, very few approaches address this problem for videos. These additionally include temporal relationships to exploit the inherent spatio-temporal information.

Accurate detection of moving objects is an important precursor to stable tracking or recognition. Sheikh and Shah [15] in 2005 presented an object detection scheme that had three innovations over other existing approaches. From an intuitive point of view, using the joint representation of image pixels allowed local spatial structure of a sequence to be represented explicitly in the modeling process. The entire background was represented by a single distribution and a kernel density estimator was used to find membership probabilities. The joint feature space provided the ability to incorporate the spatial distribution of intensities into the decision process, and such feature spaces have been previously used for image segmentation, smoothing and tracking. A second novel proposition in their work was temporal persistence as a criterion for detection without feedback from higher-level modules. The idea of using both background and foreground color models to compete for ownership of a pixel using the log likelihood ratio had been used before for improving tracking in. However, in the context of object detection, making coherent models of both the background and the foreground changed the paradigm of object detection from identifying outliers with respect to a background model to explicitly classifying between the foreground and background models. The likelihoods obtained are utilized in a MAP-MRF framework that allowed an optimal global inference of the solution based on local information. The resulting algorithm performed suitably in several challenging settings.

In video object classification, insufficient labeled data may at times be easily augmented with pairwise constraints on sample points, i.e, whether they are in the same class or not. In their paper of 2004, Yan et al. [16] proposed a discriminative learning approach which incorporated pairwise constraints into a conventional margin-based learning framework. Their proposed approach offered several advantages over existing approaches dealing with pairwise constraints. First, as opposed to learning distance metrics, the new approach drove its classification power by directly modeling the decision boundary. Second, most previous work handled labeled data by converting them to pairwise constraints and thus led to much more computation. The proposed approach could handle pairwise constraints together with labeled data so that the computation was greatly reduced. Their proposed approach was evaluated on a people classification task with two surveillance video datasets.

Enormous uncertainties in unconstrained environments lead to a fundamental dilemma that many tracking algorithms have to face in practice: Tracking has to be computationally efficient, but verifying whether or not the tracker is following the true target tends to be demanding, especially when the background is cluttered and/or when occlusion occurs. Due to the lack of a good solution to this problem, many existing methods tend to be either effective but computationally intensive by using sophisticated image observation models or efficient but vulnerable to false alarms. This greatly challenges long-duration robust tracking. In 2008 Yang et al. [17] presented a novel solution to this dilemma by considering the context of the tracking scene. Specifically, they integrated into the tracking process a set of auxiliary objects that were automatically discovered in the video on the fly by data mining. Auxiliary objects had three properties, at least in a short time interval: 1) persistent co-occurrence with the target, 2) consistent motion correlation to the target, and 3) easy to track. Regarding these auxiliary objects as the context of the target, the collaborative tracking of these auxiliary objects led to efficient computation as well as strong verification. Their extensive experiments had exhibited exciting performance in very challenging real-world testing cases.

The lack of content understanding does not allow smart TVs to provide consumers with a seamless TV shopping experience. To purchase interesting items displayed in the current TV show, consumers must inconveniently resort to a store or the Web. Object detection is one of the tasks that is required for realizing the TV shopping use case, but the detection of complex objects poses a significant challenge. Thereby, Fleitis et al. [18] proposed a multi-cue product detection framework for TV shopping. Three main characteristics define their proposed approach. Firstly, it was generic in the sense that it was not tied to a specific object detection approach. Secondly, it did not make any assumption about motion in the video. Thirdly, it utilized three cues as additional information to improve the detection results of a target product class. The appearance cue was related to the probability of a product occurrence of corresponding to the target class. The other two consisted of topological and spatio-temporal relationships between the target product class and a related, easier-to detect object class. These enforced spatial relationships within a video frame and across consecutive frames, respectively. The proposed approach jointly considered the three cues as a path-optimization problem that aims at selecting the correct product occurrences and weed out false positive detections. The empirical results demonstrated the advantages of their proposed framework in improving the detection results.

III. CONCLUSION

The problem of object detection from an unconstrained video, e.g. television, has been elaborated in detail. Such a system with high accuracy defines the next phase of the smart television sets and is a research area for many of researchers from machine learning fraternity. The paper has presented a survey of literature work pertaining to state-of-the-art published systems for object detection from videos. The pragmatic literature survey explains that, initially researchers started with developing an object detection system employing little information such as object appearances in sequential frames or with the help of some co-occurring object information. Thereafter the focus of research shifted to detecting objects with spatio-temporal information as gained from the video frames. However, the recent advances in the field had proved that, using multi-cue optimization of object detection yields much higher accuracy rates of detecting objects to specified classes, minimizing the false detections.

REFERENCES

1. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627 1645, Sep. 2010.
2. B. Leibe, A. Leonardis, and B. Schiele, “Robust object detection with interleaved categorization and segmentation,” Int. J. Comput. Vis., vol. 77, nos. 1-3, pp. 259-289, May 2008.
3. J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local features and kernels for classification of texture and object categories: A comprehensive study,” in Proc. Conf. Comput. Vis. Pattern Recognit.Workshop (CVPRW), Jun. 2006, p. 13.
4. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1. Dec. 2001, pp. 511-518.
5. M. Weber, M. Welling, and P. Perona, “Towards automatic discovery of object categories,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2. Jun. 2000, pp. 101-108.
6. A. Ayvaci and S. Soatto, “Detachable object detection: Segmentation and depth ordering from short-baseline video,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pp. 1942 1951, Oct. 2012.
7. D. Liu, M.-L. Shyu, Q. Zhu, and S.-C. Chen, “Moving object detection under object occlusion situations in video sequences," in Proc. IEEE Int. Symp. Multimedia (ISM), Dec. 2011, pp. 271-278.
8. J. Kim, G. Ye, and D. Kim, ``Moving object detection under free-moving camera," in Proc. 17th IEEE Int. Conf. Image Process. (ICIP), Sep. 2010, pp. 4669-4672.

9. B. Qi, M. Ghazal, and A. Amer, ``Robust global motion estimation oriented to video object segmentation," IEEE Trans. Image Process., vol. 17, no. 6, pp. 958 967, Jun. 2008.
10. S. Kumar and M. Hebert, ``A hierarchical field framework for unified context-based classification," in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 2. Oct. 2005, pp. 1284-1291.
11. G. Heitz and D. Koller, ``Learning spatial context: Using stuff to find things," in Proc. 10th Eur. Conf. Comput. Vis. (ECCV), 2008, pp. 30-43.
12. A. Torralba, K. P. Murphy, and W. T. Freeman, ``Using the forest to see the trees: exploiting context for visual object detection and localization," Commun. ACM, vol. 53, no. 3, pp. 107-114, Mar. 2010.
13. L. Wang, Y. Wu, T. Lu, and K. Chen, ``Multiclass object detection by combining local appearances and context," in Proc. 19th ACM Int. Conf. Multimedia (MM), 2011, pp. 1161-1164.
14. W.-S. Zheng, S. Gong, and T. Xiang, ``Quantifying and transferring contextual information in object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, pp. 762-777, Apr. 2012.
15. Y. Sheikh and M. Shah, ``Bayesian modeling of dynamic scenes for object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 11, pp. 1778-1792, Nov. 2005.
16. R. Yan, J. Zhang, J. Yang, and A. G. Hauptmann, ``A discriminative learning framework with pairwise constraints for video object classification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 4, pp. 578-593, Apr. 2006.
17. M. Yang, Y. Wu, and G. Hua, ``Context-aware visual tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 7, pp. 1195 1209, Jul. 2009.
18. F. Fleitis, H. Wang, and S. Chen, “Enhancing Product Detection With Multicue Optimization for TV Shopping Applications,” IEEE Transaction on Emerging Topics in Computing, vol. 3, no. 2, pp. 161-171. June 2015.