International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 10, December- 2012

A Survey on Image Parsing

R. Sumathi[1], D. Narmadha [2]
[1] *PG Scholar, Department of Information Technology, Karunya University ,Coimbatore, Tamil Nadu, India*

[2] *Assisstant Professor, Department of Information Technology, Karunya University, Coimbatore, Tamil Nadu, India*

Abstract

Image parsing aims to decompose an image into semantically consistent regions, it is basically a challenging problem. In the earlier days parsing is used for text annotation. Nowadays an effective image parsing facilitates many higher level image processing tasks, such as image editing, region-based image retrieval, and image synthesis. These methods are mainly based on the different learning techniques like supervised learning, unsupervised learning and semi-supervised learning techniques. Image parsing facilitates automatic image annotation, it allow users to access a large image database with textual queries. One common problem shared by most approaches for automatic image annotation is that each annotated word for an image is predicated independently from other annotated words for the same image. So the annotation is defined as label to region for the images.

1. Introduction

Image parsing is a most important part of a huge number of image understanding systems for computer vision tasks. It covers techniques for segmenting images into meaningful objects and labelling them with relative classes. Both problems of segmentation and classification are inconceivable to deal with in most natural domains without implicit or explicit consideration of context provided by other sources (i.e. domain knowledge) or egressing from the image itself (global image features). Natural images consist of a vivid number of visual patterns generated by very different random processes in nature. The main goal of image understanding system is to parsing the input images into its constituent patterns. Image parsing takes effort to find a relative meaningful label for every pixel in an image. Many vision problem uses natural images involve image parsing and a riches of possible applications include: enlarged realities, self-governing navigation, and image database retrieval.

(a) (b) (c)

Figure 1.Image parsing method (a) input image, (b) segmented image, (c) classification and annotation in region level.

The image parsing problem having the great interests in computer vision community, and many methods have been proposed. These methods could be roughly divided into two categories. The first category focuses on unsupervised or weakly supervised learning techniques and the second category is supervised learning techniques. There are different types of image parsing tasks will be defined in the previous works object recognition(One or several pre-specified or learned objects or object classes can be recognized, usually together with their 2D positions in the image or 3D poses in the scene.)Object Identification (An individual illustration of an object is recognized. For example, identification of a specific person's face or fingerprint, or identification of a specific vehicle).Object Detection (The image data are scanned with a specific condition.)Examples include detection of possible abnormal cells or tissues in medical images or detection of a vehicle in an automatic road toll system. Detection based on relatively simple and rapid computations. It is sometimes used for finding smaller regions of interesting image data which can be additionally analysed by more computationally demanding techniques to produce a correct interpretation.

Classification and segmentation is combined in some works.

The LOCUS [1] requires limited supervision, but LOCUS is only reported on a limited number of images. Here in classification the first method POM model [2] is used to learn the structure of the probability model describing the objects as well as the parameters of these distributions. It contains the ObjCut [2] method for limited number of images. And also involves parameter learning and inference for the different structure models. Second, the HGM defines a unified framework to classify an image by recognizing, annotating and segmenting the objects within the image. The third one defines HIM for 2D image parsing which outputs image segmentation and object recognition. This HIM is represented by recursive segmentation and recognition templates in multiple layers. The fourth method LRA [5] used to automatically reassign the labels annotated at the image-level to those contextually derived image regions, i.e., the label to region assignment (LRA) problem. The last method is weakly supervised graph propagation [6] is used to solve the optimization problem in image level.

2. Literature Survey

Computer vision is a research area that has benefitted from machine learning technique like few others: face recognition, object detection and action classification are just a few high-level computer vision tasks in which system that automatically learn from examples are state of the art. The types of learning techniques are supervised learning techniques and unsupervised learning techniques. These techniques are differentiated based on related papers.

2.1 Unsupervised Learning of Probabilistic Object Models for Object Classification, Segmentation and Recognition [2]

An unsupervised method to learn unified probabilistic object models (POMs) for all object-related visual tasks like classification, segmentation, and recognition of an object. This unified object model is suitable, because it provides improvement in segmentation to enable improvement in classification and vice versa. And it combines different image cues to enable accurate segmentation and classification. This method used to learn a POM-IP defined on Interest Points using weak supervision and to use this to train a POM mask, defined on regional features in the images, The POM-IP is assigned by the number of aspects and the number of nodes in each aspects. This provides a combined POM and performs segmentation and localization. This combined model can be used to train POM-edge lets, defined in various sub regions of the object shape, and it gives a full POM with improved performance on classification. The full model pairs the POM-IP, POM-mask, POM-edgelets together and performs inference on the model. The experimental results on large datasets show that the POM is invariant to scale and rotation of the object and performs inference rapidly. This also involves efficient strategies for performing parameter learning and inference for the different structure models. Finally the POM method performs inference, detecting, classifying, and segmenting the object in an image in 5 seconds. The advantage of POM matches object classes between different objects from the same category and enables object recognition from a large number of images. But Scaling and rotation of the object is unknown and less accuracy is based only on image pixels.

2.2 Towards Total Scene Understanding: Classification, Segmentation and Annotation in an Automatic Framework [3]

It is a framework for automatic learning from Internet images and tags (i.e. flickr.com), hence offering a scalable approach with no additional human labour. And able to learn robust scene models from noisy web data such as images and user tags from Flickr.com. The effectiveness of the framework is by automatically classifying, annotating and segmenting images from eight classes depicting sport scenes. A hierarchical generative model is proposed to classify the overall scene, recognizes and segments each object component, and annotates the image with a list of tags. Here the relevant visual objects are presented by regions and patches, while irrelevant visual textual annotations are influenced directly by the overall scene class. First the image is classified as scene. A number of objects can be inferred and segmented by the visual information in the scene, and hierarchically represented by object regions and feature patches, and several tags can be inferred based on the scene class and the object correspondence. In all three tasks, hierarchical generative model outperforms state of the art algorithms. It used to identify the noisy images. Geometry and appearance information of the object is not captured. It provides the complete image classification for the annotation and also identifies the noisy tags. Hierarchical generative model (HGM) doesn't capture the geometry and appearance information of the object.

2.3 Recursive Segmentation and Recognition Templates for Image Parsing[4]

A Hierarchical Image Model (HIM) is defined for 2D image parsing which outputs image segmentation and object recognition. HIM is a discriminative model and contains no model for generating the image. This HIM is represented by recursive segmentation and recognition templates in multiple layers and has advantages for representation, inference, and learning. And it emphasizes the hierarchical S-R pairs. It explicitly represent the segmentation and the labelling of the regions, while more traditional vision approaches. First, The Hierarchical Image Model has a coarse to fine representation. It is capable of capturing long range dependency and exploiting different levels of contextual information. Second, the structure of the HIM allows us to design a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 10, December- 2012

rapid inference algorithm, based on dynamic programming; it enables us to parse the image rapidly in polynomial time. Third, learn the HIM efficiently in a discriminative manner from a labelled dataset. And to demonstrate that HIM is corresponding with the state-of-the-art methods by evaluation on the challenging public MSRC and PASCAL image datasets. Finally, the HIM architecture can be extended to model more complex image phenomena. HIM can able to roughly capture different shaped segmentation boundaries. But it should improve their representational power while maintaining computational efficiency.

2.4 Label to Region by Bi-Layer Sparsity Priors [5]

In general, it is boring to manually annotate the image labels to the corresponding image regions, so the alternative way is to annotate the labels at the image-level. A sparse coding technique is assign automatically the human annotated labels at the image-level to contextually derived semantic regions merged from the over-segmented atomic image patches of the entire image set. For the label-to-region assignment task: the images can first be clustered according to the label information, and the convex l1 norm is applied within each cluster; and priors can be utilized to remove the input images without common labels with the reference image for the sparse reconstruction of the reference image or its candidate regions.

After that roughly select visually similar images from the reference image, and bi-layer sparse coding is applied for these selected images for image annotation. The LRA process has the following characteristics: 1. the bi layer sparse coding aims to implement the usage of merged patches within an image for reconstruct the reference image or semantic region, which guarantees the reliability of label propagation; 2. this process do not require exact image object/concept parsing, however it is far from satisfactory on real world images; and 3. no general model for each label/concept is learnt, so it is scalable to applications with large label set. In addition, the bi-layer sparse coding formulation can be directly applied on new test image to perform multi-label image annotation. It saves run for larger images and advantages neuroimaging segmentation. But it only focuses the consistency relationship is mainly focused.

2.5 Weakly Supervised Graph with Label Propagation for Automatic Image Annotation [6]

To automatically assign the annotated label at image level to those contextually derived semantic regions using weakly supervised graph propagation. A graph is constructed with the over-segmented patches of the image collection as nodes. Image-level labels are imposed graph as weak supervision information, each are corresponds to all patches of one image, and the contextual information across different images at patch level are then mined for assisting the process of label propagation from images to their descendent regions. Steps involved in this method is, the starting process of the algorithm is to over segment each images (i.e divide each image) into multiple homogeneous patches. Here the mean shift segmentation approach is used. Then it divides each image into around 20 coherent patches. Here the solution is general and not bound to any specific image segmentation algorithm. After the image segmentation, text on features, it encodes the both texture and colour information, and it will be extracted. The training images are first wind with a 17-Dimensional filter-bank. Then, the K-means clustering is performed.

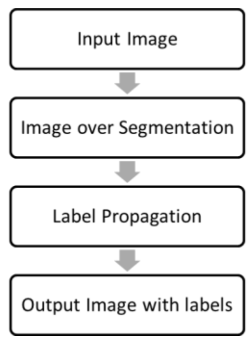

Figure 2.Illustration of the weakly supervised graph propagation.

Finally, each pixel in each image is assigned to the nearest cluster centre. Based on the texton map, normalized texton histogram within a region is computed as the patch feature descriptor. Here the optimization problem is solved by using convex concave programming and it provides higher accuracy in object segmentation and classification. But direct propagation from images to patches is not applicable. But it needs method to impose image labels for the descendent patches.

3. Conclusion

This paper addresses the problem of image parsing, or segmenting all the objects in an image and labels all the categories. The literature survey contains different image parsing methods; supervised learning is the standard for many computer vision tasks such as object recognition or scene categorization. In that the powerful classifiers can obtain impressive results but require sufficient amounts of annotated training data. However, supervised methods have limitations: Annotation is expensive, prone to error, often biased, and does not scale to large datasets. Unsupervised learning studies how systems can learn to

represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. For each input there are no explicit target outputs or environmental evaluations rather the unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output.
Also the methods including ones estimate labels pixel by pixel, ones that aggregate features over segmentation regions. Most of the methods operate with a few pre-defined classes and require a generative or discriminative model and contains optimization problem and less accuracy. The result of HGM encodes a hierarchy of semantic information contained in the scene. Also automatic annotation is not possible for supervised learning and doesn't derive about image retrieval. The weakly supervised image parsing with graph propagation is derived to automatically annotate the label at image level and it facilitate image editing ,image annotation and concept map based image retrieval with more accuracy, but the direct propagation is not possible.

Table 1. The comparison between different survey methods for image parsing

S.No	Methods	Advantages	Disadvantages
[2.1]	Unsupervised Learning of Probabilistic Object Models (POMs) for Object Classification, Segmentation and Recognition	• POM matches object classes between different objects from the same category. • Less time consuming for detecting, classifying, and segmenting the object.	• It uses Markov random field for the classification and thus provides optimization problem. • Scaling and rotation of the object is unknown. • Less accuracy
[2.2]	Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Automatic Framework	• Complete image classification and annotation. • It identifies the noisy tag.	• It doesn't derive the contextual relationships among objects. • Geometry and appearance information of the object is not captured.
[2.3]	Recursive Segmentation and Recognition Templates for Image Parsing	• HIM is able to roughly capture different shaped segmentation boundaries. • It represents complex image structures in a coarse-to-fine manner.	• It should improve their representational power object can be parsed into its consistently aligned constituent parts.
[2.4]	Labels to Region by Bi-Layer Sparsity Priors	• Run time is saved for larger images. • Neuroimaging segmentation is available.	• Consistency relationship is mainly focused. • Unclearness in small patches, so less accuracy.
[2.5]	Weakly Supervised Graph with Label Propagation	• Optimization problem is solved. • Higher accuracy in object segmentation. • Images are segmented based on colour and texture.	• Direct propagation from images to patches is not applicable. • It needs method to impose image labels for the descendent patches.

4. References

[1] J.Winn and N.Jojic, "Locus: Learning object classes with unsupervised segmentation," in Proc. IEEE Int. Conf. Comput. Vis., 2005.
[2] Y.Chen, "Unsupervised learning of probabilistic object models (poms) for object classification, segmentation and recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2008, pp. 1–8.
[3] L.-J.Li, R.Socher, and L.Fei-Fei, "Towards total scene understanding: classification, annotation and segmentation in an automatic framework," in Proc. IEEE Conf. Comput. Vis. Recognit., 2009, pp. 2036–2043.
[4] Long Zhu, Yuanhao Chen, Yuan Lin, Chenxi Lin,Yuille.A, "Recursive Segmentation and Recognition Templates for

Image Parsing" IEEE Transactions Pattern Analysis and Machine Intelligence , Feb. 2012, Volume 34,Issue 2 pp. 359 – 371.

[5] X.Liu, B. Cheng, S.Yan, J. Tang, T.-S.Chua, and H.Jin, "Label to region by bi-layer sparsity priors," in Proc. ACM Multimedia, 2009, pp. 115–124.

[6] Si Liu, Shuicheng Yan, Senior Member, IEEE, Tianzhu Zhang, Changsheng Xu, Senior Member, IEEE, Jing Liu, and Hanqing Lu, Senior Member, IEEE," Weakly Supervised Graph Propagation Towards Collective Image Parsing", IEEE Transactions On Multimedia, vol. 14, no. 2, april 2012.

[7] H.Xu, J. Wang, X.Hua, and S.Li, "Image search by concept map," in Proc. SIGIR, 2010.

[8] B. Russell, A.Efros, J.Sivic, W.Freeman, and A.Zisserman, "Segmenting scenes by matching image composites," in Proc. Adv. Neural Inf. Process. Syst. Conf., 2009, pp. 1580–1588.

[9] R.Rahmani and S.Goldman, "Missl: Multiple-instance semi-supervised learning," in Proc. Int. Conf. Mach. Learning, 2006, pp. 705 -712.

[10] G.Chen, Y.Song, F.Wang, and C.Zhang, "Semi-supervised multilabel learning by solving a sylvester equation," in Proc. SIAM Int. Conf. Data Mining, 2008, pp. 410–419.

[11] D. S. Hochbaum and V. Singh, "An efficient algorithm for co-segmentation," in Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 269–276.

[12] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no. 5, pp. 603–619, May 2002.