ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

AUTOMATIC IMAGE ANNOTATION USING WEAKLY SUPERVISED GRAPH PROPAGATION

[1]Kalaivani.R, [2]Thamaraiselvi.K

[1]ME. Computer and Communication Engineering,
Department of Information Technology, SNS College of Technology, Anna University,
Coimbatore - 35, Tamil Nadu

[2]Assistant Professor, Department of Information Technology, SNS College of Technology,
Anna university,Coimbatore -35, Tamil Nadu

Abstract:- Weakly supervised graph propagation is a method to automatically assign the annotated labels to semantically derived a semantic region. Inputs given are, the training images directory, the labels which are pre-assigned, and the Input Image .In this section, the graph Construction is done with the help of two types of relationships. Consistency Relationship mining, Incongruity Relationship mining. Propagate image labels from patches. The factors needed to be considered are, Patch Label Self-Constraints. Patch–Patch Contextual Relationships, Image-Patch Inclusion Supervision, the supervisions are the supervised and un supervised technique.

Keywords**:** POM,WSG,BSVM,MRF,CCCP

I. INTRODUCTION

Computer vision is a field that includes methods for acquiring, processing, analysing , and understanding images and, a high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.

Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception. computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.

Natural images consist of an intense number of visual patterns generated by very diverse random processes in nature. The objective of image understanding is to parse an input image into its constituent patterns. Image parsing attempts to find a semantically meaningful label for every pixel in an image.

II. IMAGE PARSING TASKS

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 instance of an object is recognized. Examples include identification of a

All Rights Reserved © 2013 IJARCET

specific person's face or fingerprint, or identification of a specific vehicle.

Object Detection

The image data are scanned for 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 fast computations is sometimes used for finding smaller regions of interesting image data which can be further analysed by more computationally demanding techniques to produce a correct interpretation.

III. LEARNING TECHNIQUES

Computer vision is an area of research 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 the state of the art. The types of learning techniques are

- Supervised learning techniques
- Unsupervised learning techniques

- **Supervised learning techniques**

Supervised learning is the standard for many computer vision tasks such as object recognition or scene categorization. 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 techniques**

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. There are no explicit target outputs or environmental evaluations associated with each input; 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.

IV. RELATED WORK

*A.*Colour Image Segmentation

The image can be segmented into different regions. Here the Figure 6.3.2 describes The image is represented as a coarse image and it uses the spatial information from the histogram based window process it segments the image based on the RGB colour values. After the image segmentation it uses the k means to cluster the entire image based on the colour segmentation.

Histograms are constructed by splitting the range of the data into equal-sized bins (called classes). Then for each bin, the number of points from the data set that fall into each bin is counted.

In colour images each pixel is characterized by three RGB values. Here we construct a 3D histogram, and the basic procedure is analogous to the method used for one variable. Histograms plotted for each of the colour values and threshold points are found. The objects can be distinguished by assigning a arbitrary pixel value or average pixel value to the regions separated by thresholds. Each image point is mapped to a point in a colour space, e.g.:

$$\text{Colour}(i, j) = (R\,(i, j), G(i, j), B(i, j)) \qquad (4.1)$$

The points in the colour space are grouped to clusters in equation(4.1). The clusters are then mapped back to regions in the image. K means algorithm for partitioning (or clustering) N data points into K disjoint subsets Sj containing Nj data points so as to minimize the sum-of-squares criterion as in equation

All Rights Reserved © 2013 IJARCET

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

$$J=\sum_{j=1}^{K}\sum_{n\in S_j}|x_n-\mu_j|^2$$

B.WSG Propagation

WSG propagation means to finding the hidden image patches from the unlabelled image.WSG encodes two types of contextual information among image patches, i.e., consistency and incongruity. Finally, the collective image parsing task is formulated as a constrained optimization problem.

C.Graph Construction

In the label propagation algorithm to construct a graph is critical. In this work, the nodes are over segmented image patches, and the ideal edge weights should measure the semantic relationships among the nodes. Here, the semantic relationships include two types of contextual information, one is the consistency relationship, and the other is the incongruity relationship.

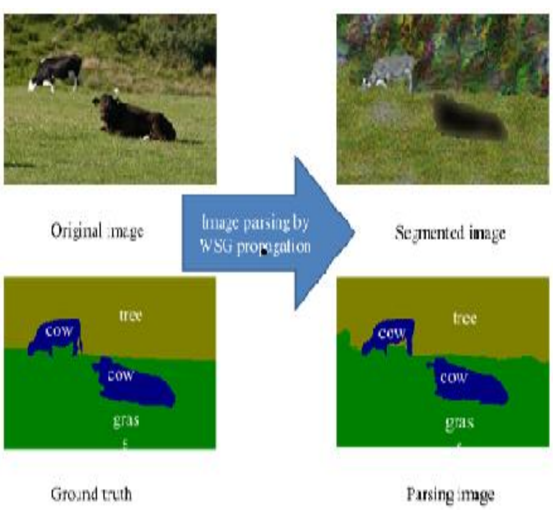

Figure 4.1.image parsing tasks

Sparse coding is used to build the consistency relations among image patches. To reconstruct each image patch as a sparse linear combination of the rest image

patches coming from images with at least one common label. The image patches with nonzero reconstruction coefficients are considered to be similar with the reconstructed patch.

Let h denotes ı (4.2) feature vectors of the image patch, h is column normalized with unitary l_2 norm. φ denotes the coefficient of the derived sparse coding and z_k denotes the parse linear combination .Then φ is derived by solving the optimization problem in equation (4.2)

$$\min\|\varphi\|_{l_1}, \quad \text{s.t.} \quad z_k\varphi = h_k \qquad (4.2)$$

This optimization problem is convexed and can be transformed into general linear programming problem.

D.Incongruity Relationship Mining

To mine the contextual information among the image patches, another graph characterizing the incongruity relationship is introduced. In this graph, the edge weight denotes patch dissimilarity. The higher the edge weight is, the less likely the nodes at the two ends are to be assigned with the same label. To guarantee robustness to noises, sparse the graph by maintaining the farthest neighbours for each image patch and setting other weights to be zero.

For each patch, requires its most dissimilar patches to be labelled differently. Assume that in most cases, the extremely dissimilar patches should be from different labels. Therefore, we put this observation into our formulation.

E.Label propagation

Based on the derived consistency relationship graph and incongruity relationship graph, the task is to propagate labels from images to patches. To obtain the mathematical formulation .

All Rights Reserved © 2013 IJARCET

V. RESULTS AND DISSCUSSIONS

1.Results

Compared with baselines, the proposed WSG algorithm matches much higher accuracies of 0.71, 0.64, and 0.38on the MSRC, COREL-100, and VOC-07 dataset respectively. Since the BSVM classifier is trained at the image level and tested at the patch level, it performs worst. It shows that cross-level label inference is not trivial, and straightforward propagating labels from images to patches is not applicable. A more sophisticated method is required to weakly impose image labels upon their descendent patches.

Contextual image parsing algorithms, including KNN, the bi-layer, and the proposed WSG-based algorithms, all out perform the BSVM-based counterpart. It is because the former three harness the contextual information among the semantic regions in the image collection.

WSG-based algorithm clearly beats the state-of the-art bi-layer sparse coding algorithm to the fact that the weakly supervised information of graph avoids the ambiguities among the smaller patches in the bi-layer sparse coding algorithm and WSG can make use of both consistency and incongruity relationships among patches while the bi-layer method mainly focuses on the consistency relationship.

Detailed comparison results for individual labels are illustration on Figure 5.1 MSRC dataset have 11 out of total 18labels better than the bi-layer method; on COREL-100 dataset have five out of a total of seven labels better and in VOC-07dataset, we have 17 out of 21 labels better than the bi-layer baseline. The results demonstrate the superiority of the proposed algorithm over baseline algorithms.

From these results, we can conclude that during iteration ,information is propagated through the graph effectively. To further show the optimization progress of the proposed algorithm, give some detailed intermediate image parsing results for different iterations can observe that the image parsing results become better and better as the iteration goes.

Therefore, the proposed algorithm is scalable to large-scale applications. Because the pixel-label ground truth label is not provided for the NUS-WIDE-SUB dataset, can not quantitatively report the image parsing results. However,image parsing facilitates image annotation task, which can be quantitatively evaluated.

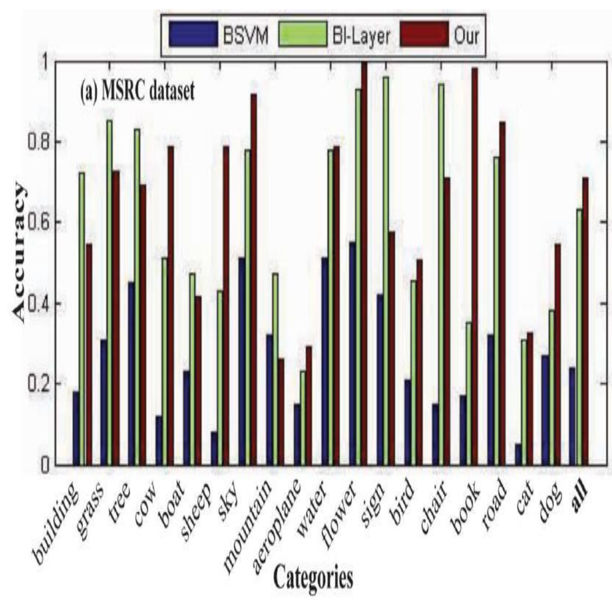

Figure 5.1 MSRC dataset

MSRC dataset Figure 5.1 shows to improve the annotated values using the bi-layer and BSVM algorithms. The horizontal axis shows the name of each label and the vertical axis represents the collective image parsing accuracy.

All Rights Reserved © 2013 IJARCET

2.Discussions

In the process focuses on how to propagate the image labels to their regions and assumes that the image labels have been properly labeled. Unfortunately, in real applications, e.g., for image set flickr.com, image labels are provided by users and inevitably noisy. In this situation, first preprocess the image labels with the state-of-the-art label-refinement algorithm and then run the algorithm based on refined and cleaner labels. The weakly supervised graph propagation is to improve the results of an annotation. The label level spatial contextual relationship for boosting collective image parsing accuracy.

VI. CONCLUSION AND FUTURE ENHANCEMENT

Addresses the problem of image parsing, or segmenting all the objects in an image and labels all the categories. The literature survey contains different proposed image parsing 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 future enhancement is in order to improve an accuracy on concept map based image retrieval. Also automatic annotation is not possible for supervised learning and doesn't not derived about image retrieval. So weakly supervised image parsing with graph propagation is derived to automatically annotate the label at image level and it facilitate image editing ,image annotation. The label level spatial contextual relationship for boosting collective image parsing accuracy.

VII. REFERENCES

[1] Si Liu, Shuicheng Yan, Tianzhu Zhang, Changsheng Xu, Weakly Supervised Graph Propagation Towards Collective Image Parsing,Senior Member, IEEE, Senior Member, IEEE,Jing Liu, and Hanqing Lu, Senior Member, IEEE

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

[3] Chen.Y., (2008), "Unsupervised learning of probabilistic object models (poms) for object classification, segmentation and recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1–8.

[4] Chu.W.-S., Chen.C.-P. and Chen.C.-S., (2010), "Momi-cosegmentation: Simultaneous segmentation of multiple objects among multiple images," in Proc. Asian Conf. Comput. Vis.,pp. 355–368.

[5] Chua.T., Tang.J., Hong.R., Li.H., Luo.Z., and Zheng.Y., (2009), "Nus-wide: A real-world web image database from national university of singapore," in Proc. CIVR, pp. 48–55.

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

[7] He.J., Li.M., Zhang.H., Tong.H., and Zhang.C., (2004), "Manifoldranking based image retrieval," in Proc. ACM Multimedia, pp. 9–16.

[8] Li.L.-J., Socher.R., and Fei-Fei.L., (2009), "Towards total scene understanding: classification, annotation and segmentation in an automatic framework," in Proc. IEEE Conf. Comput. Vis. Recognit., pp. 2036–2043.

All Rights Reserved © 2013 IJARCET

[9] Liu.C., Yuen.J., and Torralba.A., (2009), "Nonparametric scene parsing: Label transfer via dense scene alignment," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,pp. 115–124.

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

[11] Mukherjee.L., Singh.V., and Peng.J., (2011), "Scale invariant cosegmentation for image groups," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2028–2035.

[12] Pan.J., Yang.H., Faloutsos.C. and Duygulu.P., (2004), "Gcap: Graph-based automatic image captioning," in Proc. 4th Int. Workshop Multimedia Data Document Eng.,pp. 146.

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

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

[15] Tu.Z., Chen.X., Yuille.A.L., and Zhu.S.-C., (2005), "Image prsing: Unifying segmentation, detection, and recognition," Int. J. Cmput. Vs., vol. 63, no. 2, pp. 113–140.

[16] Winn.J. and Jojic.N., (2005), "Learning object classes with unsupervised segmentation," in Proc. IEEE Int. Conf. Comput. Vis., 2005.

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

[18] Yuan.J., Li.J., and Zhang.B., (2008), "Scene understanding with discriminative structured prediction," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1–8.

[19] Zhu.G., Yan.S., and Ma.Y., (2010), "Image tag refinement towards low-rank, content-tag prior and error sparsity," in Proc. Int. Conf. Multimedia, pp. 461–470.

[20] [Online]. Available: http://www.flickr.com/

[21] [Online]. Available: http://www.image parsing.com/

[22] [Online]. Available: http://www.acm.caltech.edu/l1magic/

[23] [Online]. Available: http://www.labelme.csail.mit.edu/

All Rights Reserved © 2013 IJARCET