VOL.2, ISSUE 3, MARCH-2013 ISSN NO: 2319-7471

PSO based Automatic Image Annotation using Weakly Supervised Graph Propagation

R. Sumathi[1], D. Narmadha[2]
[1] PG Scholar, [2] Assistant Professor, Department of Information Technology,
Karunya University, Coimbatore, Tamil Nadu, India
[1]sumathirajendran2226@gmail.com,[2]narmadha@karunya.edu

Abstract: In this work, a weakly supervised graph propagation method is proposed to automatically assign the annotated labels at image level to those contextually derived semantic regions. The images can be segmented into different regions using color image segmentation. The graph is built with the over-segmented patches of the image pool as nodes. Image-level labels are carried out on the graph as weak supervision information over sub graphs, each of which relates to all patches of one image, and the contextual information from end to end different images at patch level are then mined to assist the process of label propagation from images of their descendent regions. The Weakly Supervised Graph propagation encodes two types of contextual information among image patches, i.e., consistency and incongruity. The decisive optimization problem is proficiently solved Particle swarm optimization (PSO) method. It iteratively tries to improve a candidate solution with regard to a given measure of quality for optimization. Experiments using the data sets show the effectiveness of the proposed method for the task of collective image parsing.

Keywords: Weakly supervised graph propagation, Particle swarm optimization(PSO), Image annotation, Image Classification, Image segmentation.

1. INTRODUCTION

Image parsing is an essential part of many of image understanding systems for computer vision tasks. It incorporates techniques for segmenting images into significant objects and labelling them with semantic classes. This task is mainly used in online photo sharing sites like 'flicker', 'behold', which has a large number of images with a user given labels, and these labels are refined by exploiting correlations across images. The segmentation and the classification problems are impossible to deal with in most natural domains without implicit or explicit consideration of context provided by other sources (i.e. Sensors or a domain knowledge) or emerging from the image itself (neighbourhood or global image features).
Image annotation is the process to automatically assign metadata in the state of captioning to a digital image in a computer system. This method can be viewed as a classification of multi-class images with a huge number of classes as huge as the vocabulary size. Generally, image analysis in the form of extracting structural vectors and the trained annotation words is used by machine learning techniques in effort to automatically put on annotations to new images. The basic methods well-read the connections between image structures and training annotations, then the methods are settled by machine translation to go to translate the textual language with the 'visual terms', or grouped regions known as blobs.

The advantages of the automatic image annotation contrasted with content-based image retrieval are the queries can be more logically stated by the user. CBIR usually needs users to search images by the way of concepts like color, texture, finding instance queries. Certain image types in example images may take priority over the model that the user is truly concentrating on. The old methods of image retrieval have depended on manually annotated images, which is costly and time-consuming. 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 chance model describing the objects as well as the limits of these distributions. It has the ObjCut [2] method for a limited number of images. And also involves parameter learning and inference for the different structure models. Second, the HGM defines a unified framework to categorize an image by recognizing, segmenting and annotating the objects within that particular class. The third one defines HIM for 2D image parsing which gives image segmentation and object recognition. Hierarchical Image Model is presented by recognizing and segmenting several layers. The fourth method LRA [5] used to automatically re-assign the labels annotated on the image-level to those derived contextually from image regions, i.e., the label to region assignment (LRA) problem.

This paper presents a framework for parsing the input signal into a categorised structure (i.e.) images into semantically consistent regions. A graph based approach is proposed to solve the weakly supervised image parsing task. Its inputs are collection of images with annotation and the outputs are semantic consistent regions with required labels.

2. RELATED WORKS

Computer vision is a research area that has benefitted from machine learning technique like few others: face recognition, object detection and 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.

An unsupervised method [2] is to learn unified probabilistic object models (POMs) for all objects-related visual tasks like classification, segmentation, and recognition of an object. It provides improvement in segmentation to enable improvement in classification and vice versa. The advantage of POM matches object classes between different objects of the same class and enables object recognition from many of the images. But Scaling and rotation of the object is unknown and less accuracy is based only on image pixels.

A Hierarchical Image Model (HIM) [4] is defined for 2D image parsing which outputs image segmentation and object recognition. HIM is a discriminative model and has no model for generating the image. It explicitly represents the segmentation and the labelling of the image regions. HIM can able to roughly capture different shaped segmentation boundaries. But it should improve their representational power while maintaining computational efficiency.

A sparse coding technique [5] is to assign automatically the human annotated labels at the image-level to contextually derive semantic regions merged from the over-segmented atomic image patches of the entire image set. In this method, the bi-layer sparse coding formulation can be directly applied to new test image to do multi-label image annotation. It saves run for larger images and advantages neuroimaging segmentation. But it only focuses the consistency relationship is mainly focused.

This is a method to automatically assign the annotated label at image level to those contextually derived semantic regions using weakly supervised graph propagation [6]. 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 corresponding to all patches of one image, besides the contextual information across different images at patch level are then mined for assisting the process of label propagation from images of their descendent regions. It tries to cut the optimization problem via CCCP. But the accuracy level is low. Perhaps, image parsing may increase extra from contexts that the sky often looks above water in images.

This method [16] is a novel approach that provides effective and robust color image segmentation images. To form segmented regions this method pre-processes an image by using the MS algorithm that reserves the required disjointedness characteristics of the image. By using the graph structures the segmented regions are represented, and the Ncut method is applied to do generally improved clustering. On the other hand, the application of the region adjacent graph and Ncut methods of the resulting segments, instead of directly to the image pixels, produces superior image segmentation performance. This method requires suggestively lower computational complexity and, then only it is possible to real-time image processing.

3. PROPOSED SYSTEM

The previous methods don't improve the accuracy due to optimization in image segmentation and classification so the proposed system is mainly focused to improve the accuracy of the image segmentation and classification using the PSO algorithm. The overall method has three steps

- Color image segmentation
- WSG propagation
- PSO algorithm

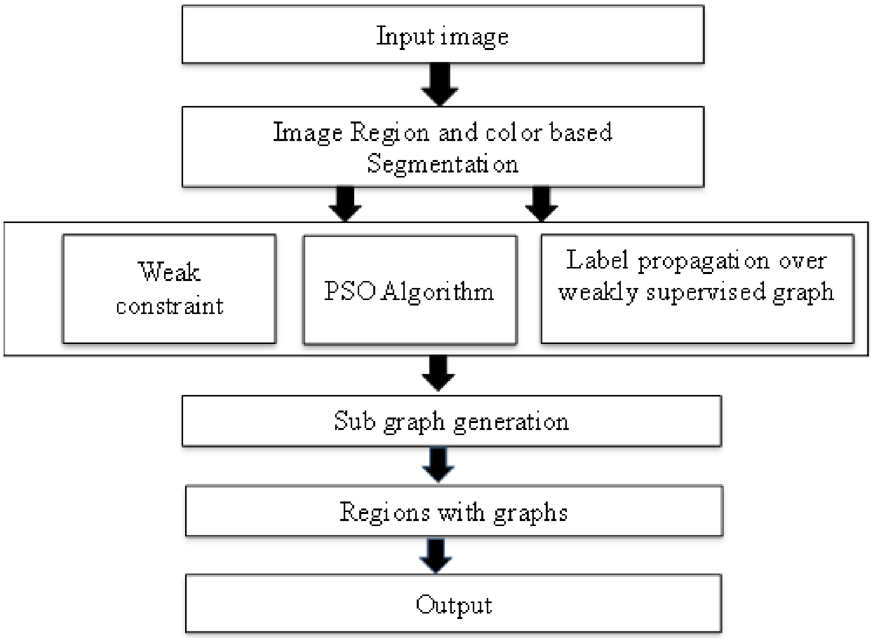


Figure 1. Overview of the proposed framework

VOL.2, ISSUE 3, MARCH-2013 ISSN NO: 2319-7471

A. Color Image Segmentation

The image can be split into different regions. In colour image segmentation 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 color values. After the image segmentation, the k means applied to cluster the entire image based on the color segmentation.

For that, the histograms are constructed by parting the range of the data into equal-sized bins (called classes). Then for all bin the number of points from the data set that fall under each bin is calculated. In color images every pixel is classified by three RGB values. Then to build a 3D histogram, the basic procedure is corresponding to the method used for one variable. Histograms plotted for each of the color values and threshold points are found in the pixels. The objects can be individualized by allocating an arbitrary pixel value or average pixel value to the regions separated by thresholds. Each image point is plotted to a point in the color area, such as,

$$\text{Color}(i, j) = (R(i, j), G(i, j), B(i, j)) \tag{1}$$

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

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

B. WSG Propagation

WSG codes consistency and incongruity contextual information among images. Then, 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 best edge weights should size the semantic relationships among the nodes. Here, the semantic relationships consist of two types of contextual information, first is the consistency relationship, second is the incongruity relationship. In the consistency relationship sparse coding is used to build the 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 to the reconstructed patch. Let h denotes the feature vectors of the image patch," h" is calumny normalized to unitary l2 norm. Denotes the coefficient of the derived sparse coding. Then is derived by solving the optimization problem in the equation

$$\min \|\varphi\|_{l_1}, \quad \text{s.t.} \quad z_k \varphi = h_k \tag{3}$$

To mine the contextual information among the image patches, 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 to the same label.

D. Label Propagation

Based on the derived consistency relationship graph and incongruity relationship graph, the task is to transmit labels from images to patches. To find the mathematical construction for this task, the following factors need be taken into consideration. First the Patch Label Self-Constraints use the patch value is to be range from [0,1]. Second the Patch–Patch Contextual Relationships is to integrate the consistent relationship between patches into the formulation. Finally, Image-Patch Inclusion Supervision defines the weakly supervised label information imposed by image labels.

E. PSO Algorithm

Particle swarm optimization (PSO) is a computation method that optimizes a problem by iteratively trying to increase a candidate solution with regard to a given measure of quality. This uses a number of particles that set up a swarm moving everywhere in an N-dimensional search space looking for the best solution. Every particle takes track of its coordinates in the solution space, which are related with the best solution that is achieved to this point by that particle is called as personal best position (pbest) and the other best value achieved until now by any particle in the neighbourhood of that particle is called as global best position (gbest). Each particle tries to alter its position using the following information.

- Recent positions.
- Recent velocities.
- Space between the recent position and pbest.
- Space between the recent position and gbest.

This algorithm mainly focused a solution to the multilevel image thresholding problems. The number of threshold levels is the dimension of the problem. Such as, if there are "m" threshold levels, the ith particle is represented as follows:

$$Xi = (Xi1, Xi2, Xi3, \ldots\ldots\ldots\ldots\ldots, Xim) \tag{4}$$

- **Swarm initialization:** For a population size p, the particles are randomly produced between the minimum and the maximum bounds of the threshold values.
- **Objective function evaluation**: The particle's objective function values are evaluated using the objective functions.

VOL.2, ISSUE 3, MARCH-2013 ISSN NO: 2319-7471

- **Initialization of the pbest and gbest:** The objective values gotten above for the initial particles of the swarm are fixed at the initial pbest values of the particles. The best value between all the pbest values is known as gbest.
- **Velocity evaluation:** The new velocity of each particle is calculated.

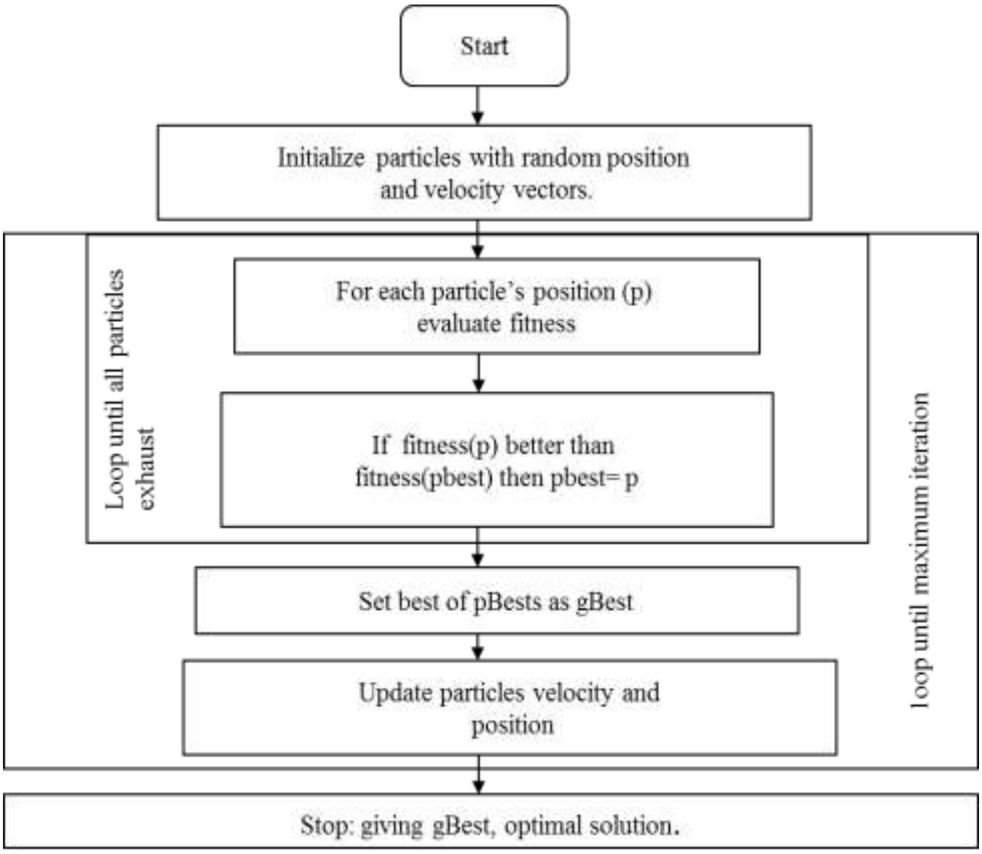


Figure 2. The PSO Algorithm Flow Chart

- **Swarm Updating:** The position of the particle is updated. The objective function values are calculated for the updated positions of the particles. If the new value is developed than the previous pbest, the new value is recognized to pbest. Likewise, gbest value is also updated as the best pbest.
- **Stopping criteria**: If the stopping criteria are met, the positions of particles signified by gbest are the optimal threshold values. Otherwise, the process is repeated.

4. RESULT ANALYSIS

There are 200 images taken from the dataset to calculate the accuracy of the proposed system. It has different labels like "water", "cow", "sheep", "horse", "sky", "car", "human", "bicycle", "airplane", "building", "tree", "grass". The PSNR and RMS error value and the computation time is calculated to define the accuracy of the proposed system.
The two images were taken to find the RMS error value. The ground truth image is taken as a reference image and the existing method image and the proposed images compared, the RMS of the pairwise differences of the two images can serve as a measure how far on average the error is from 0.The quality of the threshold images can also be evaluated through Peak Signal to Noise Ratio (PSNR) measure.

$$\mathrm{PSNR} = 20\log_{10}\frac{255}{\mathrm{RMSE}} \quad (5)$$

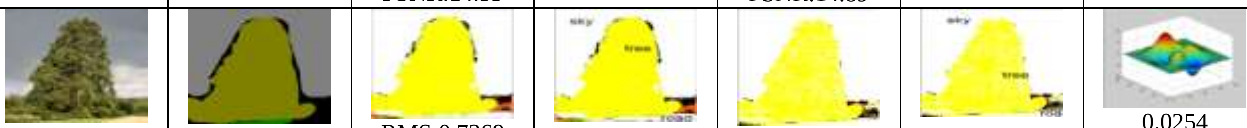
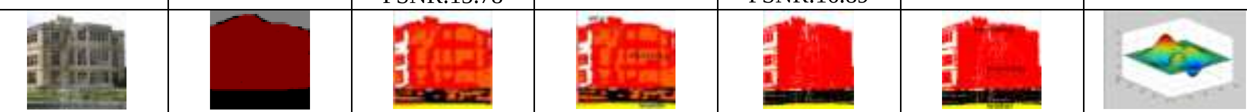
Where,

$$\mathrm{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[\mathrm{I}(i,j) - \mathrm{I}(i,j)]^2} \quad (6)$$

In the observation of the computation time, the proposed method is faster than the existing methods. It is shown in the table the CPU time increases with the number of thresholds.

VOL.2, ISSUE 3, MARCH-2013 ISSN NO: 2319-7471

Table 1. Result analysis for RMS, PSNR, and Computation time

Images	Ground Truth	Existing system		Proposed system		Computation Time (secs)
		Accuracy calculation	Annotation (label)	Accuracy calculation	Annotation (label)	
		RMS:0.0468 PSNR:15.43		RMS:0.0275 PSNR:19.84		0.0312
		RMS:0.8218 PSNR:12.45		RMS:0.6590 PSNR:17.57		0 .0156
		RMS:0.6590 PSNR:13.48		RMS:0.5326 PSNR:17.16		0.0156
		RMS:0.5112 PSNR:14.55		RMS:0.2775 PSNR:14.89		0.0312
		RMS:0.7369 PSNR:12.37		RMS:0.6303 PSNR:21.78		0.0254
		RMS:0.9225 PSNR:15.78		RMS:0.7492 PSNR:16.89		0.0127
		RMS:0.4503 PSNR:25.34		RMS:0.3863 PSNR:27.56		0.0254
		RMS:0.5483 PSNR:28.16		RMS:0.4490 PSNR:28.79		0.0542
		RMS:0.8298 PSNR:17.71		RMS:0.6935 PSNR:18.53		0.0459
		RMS:0.7998 PSNR:20.12		RMS:0.4956 PSNR:22.45		0.0156

5. CONCLUSION

This system carries out the problem of collective image parsing under the weakly supervised setting with a novel WSG-based label propagation method and improves the accuracy of the image parsing task. Most of the methods use a few pre-defined classes which are generative or discriminative model and has optimization problem and an automatic annotation is not possible for supervised learning techniques and image retrieval is not possible. This proposed method is different from traditional image parsing methods; it requires only image-level label annotations as input. WSG can absorb weak label information from images and propagate them among patches concurrently. So weakly supervised image parsing with graph propagation is derived to automatically annotate the label at image level and PSO algorithm is used for the optimization to improve the accuracy.

REFERENCES

[1]. F 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.
[13]. P.D. Sathya, R. Kayalvizhi,” PSO-Based Tsallis Thresholding Selection Procedure for Image Segmentation”, In: International Journal of Computer Applications (0975 – 8887)Volume 5– No.4, August 2010.
[14]. Chun-Chieh Tseng, Jer-Guang Hsieh and Jyh-Horng Jeng, “Fractal image compression using visual-based particle swarm optimization,” Image and Vision Computing 26 (8) (2008),1154-1162.
[15]. Yi-Tung Kao, Erwie Zahara and I-Wei Kao, “ A hybridized approach to data clustering,” Expert Systems with Applications 34 (2008), 1754-1762.
[16]. Wenbing Tao, Hai Jin, Senior Member, IEEE, and Yimin Zhang, Senior Member, IEEE,”Color Image Segmentation Based on Mean Shift and Normalized Cuts”, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 37, No. 5, October 2007.