IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.- 6, Issue-12, Dec.2018 E-ISSN: 2347-2693

Knowledge Based Unsupervised Object Discovery Using Probabilistic Randomized Hough Transform (PRHT) With Deep Learning Classification (DLC)

Mereena Johny[1*], L. Haldurai[2]

[1]Department of Computer Science, Kongunadu Arts and Science College, Coimbatore, India
[2]Department of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore, India.

[*]*Corresponding Author: mereenajohny123@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 27/Dec/2018, Published: 31/Dec/2018

Abstract: Latent topics models have become a popular paradigm in many computer vision applications due to their capability to discover semantics in visual content. Various knowledge based object discovery algorithms for the classification problem in dependent images are appearing in the literature. However, these algorithms mostly suffer from the following two problems: image metadata and time measures. To overcome this kind of problem, this paper presents a Probabilistic Randomized Hough Transform (PRHT) with Deep Learning Classification Algorithm (DLC) algorithm performs the object discovery and localization used by deep learning classification algorithm. The proposed method of object regions are efficiently matched across images using a Probabilistic Randomized Hough Transform with Deep Learning Classification that evaluates the confidence for each candidate correspondence considering both appearance and spatial consistency. The achieved PRHT-DLC has high accuracy and performance increases compared to the previous method of Pipeline method and Latent Dirichlet allocation (LDA) algorithms.

Keywords: Image Mining, Image Retrieval, Probabilistic Randomized Hough Transform, Deep learning, Unsupervised object discovery.

I. INTRODUCTION

Image mining is a development potential technology for data mining which involves in multiple disciplines; it is also a challenging field which extends traditional data mining from structured data to unstructured data such as image data.

Knowledge-Based Object localization and detection is highly challenging because of intra-class variations, background clutter, and occlusions present in real-world images. While significant progress has been made in this area over the last decade, as shown by recent benchmark results, most state-of-the-art methods still rely on strong supervision in the form of manually-annotated bounding boxes on target instances. Since those detailed annotations are expensive to acquire and also prone to unwanted biases and errors, recent work has explored the problem of weakly-supervised object discovery where instances of an object class are found in a collection of images without any box-level annotations.

Application of World Wide Web and the internet is increasing exponentially, the need for finding an image in internet is also increasing rapidly. A huge amount of image databases are added every minute and so is the need for effective and efficient image retrieval system. Retrieving an image having some characteristics in a big database is a crucial task. Searching for an image among a collection of images can be done by different approaches.

Currently is a growing interest in Unsupervised Object Discovery and Localization technique because of the limitations inherent in text based systems, as well as the large range of possible uses for efficient image retrieval. The present technology is adequate to search images using textual information. But it requires humans to personally describe image information in the database. This is very impractical for large amount of image databases. It is possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling issues.

Current object localization approaches heavily rely on strong supervision, in the form of training images that have been manually collected, labeled and annotated. These approaches learn to detect an action using bounding box annotations and recognize using action class labels from training data. Since supervised methods have the annotated ground truth at their disposal, they can take the advantage of learning detectors and classifiers by fine-tuning over the training data.

Most commercial image retrieval systems associate keywords or text with each image in the corpus and require the user to enter a keyword or textual description of the desired image. Standard text retrieval techniques are then used to identify the relevant images. Unfortunately this text-based approach to image retrieval has numerous drawbacks. Associating keywords or text with each image is a tedious and time-consuming task since it must be done manually or at best semi-automatically; image processing technology is not advanced enough to allow the automatic construction of textual image descriptions except in well-defined and tightly focused domains.

This paper presents a Probabilistic Randomized Hough Transform (PRHT) with Deep Learning Classification Algorithm (DLC) for Caltech4, LabelMe, and PSACAL07-6 × 2 data sets

II. RELATED WORK

A. Joulin, F. Bach, and J. Ponce (2010) discussed to combine the existing tools for bottom up image segmentation such as normalized cuts, with kernel methods commonly used in object recognition. These two sets of techniques are used within a discriminative clustering framework: the goal is to assign foreground/background labels jointly to all images, so that a supervised classifier trained with these labels leads to maximal separation of the two classes [1].

D. Mimno, et.al., (2011) addressed the dimensionality reduction methods for text, such as latent Dirichlet allocation, often produce low-dimensional subspaces (topics) that are obviously flawed to human domain experts. The contributions of this techniques are threefold: (1) An analysis of the ways in which topics can be flawed; (2) an automated evaluation metric for identifying such topics that does not rely on human annotators or reference collections outside the training data; (3) a novel statistical topic model based on this metric that significantly improves topic quality in a large-scale document collection from the National Institutes of Health (NIH)[2].

T. Deselaers, B. Alexe, and V. Ferrari (2012) presented a novel approach that can cope with extensive clutter as well as large scale and appearance variations between object instances. To make this possible to exploited generic knowledge learned beforehand from images of other classes for which location annotation is available. Generic knowledge facilitates learning any new class from weakly supervised images, because it reduces the uncertainty in the location of its object instances. Meanwhile, they proposed a conditional random field that starts from generic knowledge and then progressively adapts to the new class[3].

Z. Chen, et.al., (2013) addressed to go one step further to study how the prior knowledge from other domains can be exploited to help topic modeling in the new domain. This problem setting is important from both the application and the learning perspectives because knowledge is inherently accumulative. The human beings gain knowledge gradually and use the old knowledge to help solve new problems[4].

M. Rubinstein, J. Kopf, C. Liu, and A. Joul (2013) presented a new unsupervised algorithm to discover and segment out common objects from large and diverse image collections. In contrast to previous co-segmentation method algorithm performs well even in the presence of significant amounts of noise images (images not containing a common object), as typical for datasets collected from Internet search. The key insight to the algorithm is that common object patterns should be salient within each image, while being sparse with respect to smooth transformations across images. They proposed to use dense correspondences between images to capture the sparsity and visual variability of the common object over the entire database, which enables us to ignore noise objects that may be salient within their own images but do not commonly occur in others. To performed extensive numerical evaluation on established co-segmentation datasets, as well as several new datasets generated using Internet search[5].

A. Faktor and M. Irani (2014) discussed to define a “good image cluster” as one in which images can be easily composed (like a puzzle) using pieces from each other, while are difficult to compose from images outside the cluster. The larger and more statistically significant the pieces are, the stronger the affinity between the images. This gives rise to unsupervised discovery of very challenging image categories. To further show how multiple images can be composed from each other simultaneously and efficiently using a collaborative randomized search algorithm. This collaborative process exploits the “wisdom of crowds of images”, to obtain a sparse yet meaningful set of image affinities, and in time which is almost linear in the size of the image collection[6].

A. Joulin, K. Tang, and L. Fei-Fei (2014) tackled the problem of performing efficient co-localization in images and videos. Co-localization is the problem of simultaneously localizing (with bounding boxes) objects of the same class across a set of distinct images or videos. Building upon recent state-of-the-art methods, to show how they are able to naturally incorporate temporal terms and constraints for video co-localization into a quadratic programming framework. Furthermore, by leveraging the Frank-Wolfe algorithm (or conditional gradient), To showed how their optimization formulations for both images and videos can be reduced to solving a succession of simple integer programs, leading to increased efficiency in both memory and speed[7].

Z. Niu, G. Hua, X. Gao, and Q. Tian (2014) addressed the problem of recognizing images with weakly annotated text tags. Most previous work either cannot be applied to the scenarios where the tags are loosely related to the images; or simply take a pre-fusion at the feature level or a post-fusion at the decision level to combine the visual and textual content. Instead, to first encode the text tags as the relations among the images, and then propose a semi-supervised relational topic model (ss-RTM) to explicitly model the image content and their relations. In such way, it can efficiently leverage the loosely related tags, and build an intermediate level representation for a collection of weakly annotated images[8].

L.Haldurai et.al. (2016) proposed a novel approach of parallel indexing the color and feature extraction of images and genetic algorithm has been implemented. Its main functionality is image-to-image matching and its intended use for still-image retrieval. The evaluation criteria are provided by the genetic algorithm and have been successfully employed as a measure to evaluate the efficacy of content-based image retrieval process [9].

C. Wang, K. Huang, W. Ren, J. Zhang, and S. Maybank (2015) proposed the latent category learning (LCL) in large-scale cluttered conditions. LCL is an unsupervised learning method which requires only image-level class labels. Firstly, they use the latent semantic analysis with semantic object representation to learn the latent categories, which represent objects, object parts or backgrounds. Secondly, to determine which category contains the target object, they proposed a category selection strategy by evaluating each category’s discrimination. Finally, they proposed the online LCL for use in large-scale conditions. Evaluation on the challenging PASCAL VOC 2007 and the large-scale ILSVRC 2013 detection datasets shows that the method can improve the annotation precision by 10% over previous methods[10].

M. Cho, S. Kwak, C. Schmid, and J. Ponce (2015) addressed an unsupervised discovery and localization of dominant objects from a noisy image collection with multiple object classes. The setting of this problem is fully unsupervised, without even image-level annotations or any assumption of a single dominant class. This is far more general than typical co-localization, co-segmentation, or weakly-supervised localization tasks. To tackled the discovery and localization problem using a part-based region matching approach: To use off-the-shelf region proposals to form a set of candidate bounding boxes for objects and object parts. These regions are efficiently matched across images using a probabilistic Hough transform that evaluates the confidence for each candidate correspondence considering both appearance and spatial consistency. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them[11].

*L.Haldurai et.al. (2016) e*lucidates various methods and techniques found to be very effective when combined with Genetic algorithm to derive optimal solution and to increase the computation time of retrieval systems [12].

Zhenxing Niu, et.al. (2018) addressed to tackle common object discovery in a fully unsupervised way. Generally, object co-localization aims at simultaneously localizing objects of the same class across a group of images. Traditional object localization/detection usually trains specific object detectors which require bounding box annotations of object instances, or at least image-level labels to indicate the presence/absence of objects in an image. Given a collection of images without any annotations, they proposed fully unsupervised method is to simultaneously discover images that contain common objects and also localize common objects in corresponding images. Without requiring knowing the total number of common objects, to formulate this unsupervised object discovery as a sub-graph mining problem from a weighted graph of object proposals, where nodes correspond to object proposals and edges represent the similarities between neighboring proposals [13].

L.Haldurai et.al. (2018) [14] elucidates the precise picture of studies related to unsupervised object discovery and localization in knowledge based topic models and also to find relevant techniques to consider all possible datasets to localization measures which proves to be the most important criteria for image classification.

Zhenxing Niu, et.al. (2018) addressed an unsupervised discovery and localization of dominant objects from a noisy image collection with multiple object classes. The setting of this problem is fully unsupervised, without even image-level annotations or any assumption of a single dominant class. This is far more general than typical co-localization, co-segmentation, or weakly-supervised localization tasks. To tackled the discovery and localization problem using a part-based region matching approach: To use off-the-shelf region proposals to form a set of candidate bounding boxes for objects and object parts. These regions are efficiently matched across images using a probabilistic Hough transform that evaluates the confidence for each candidate correspondence considering both appearance and spatial consistency. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them [15].

III. Research Methodology

In this paper, proposed a Probabilistic Randomized Hough Transform (PRHT) with Deep Learning Classification Algorithm (DLC) algorithm that can be easily "instantiated" both in main memory and on top of different underlying different datasets. The proposed method accepts the simulation parameters as input which contains the MATLAB R2013a simulation where the Probabilistic Randomized Hough Transform (PRHT) with Deep Learning Classification Algorithm (DLC) for Caltech4, LabelMe, and PSACAL07-6 × 2 data sets. These overall proposed flow diagrams in figure 2 follows a cluster procedure form begin to end state.

A. Image Data Preprocessing

Image data preprocessing is a technique that involves transforming raw image data into an understandable format. In this pre-processing stage is an enhancement of the image information that contains unwanted distortions or enhances some image features important for further processing. In this stage, original RGB Seed image database of pixels size (1109 x 1069 x 3 uint8) is resized into 256 x 256 dimensions without pixel information loss using 'bicubic' interpolation method. After that, images must be of the identical size and are supposed to be associated with indexed images on a common color map.

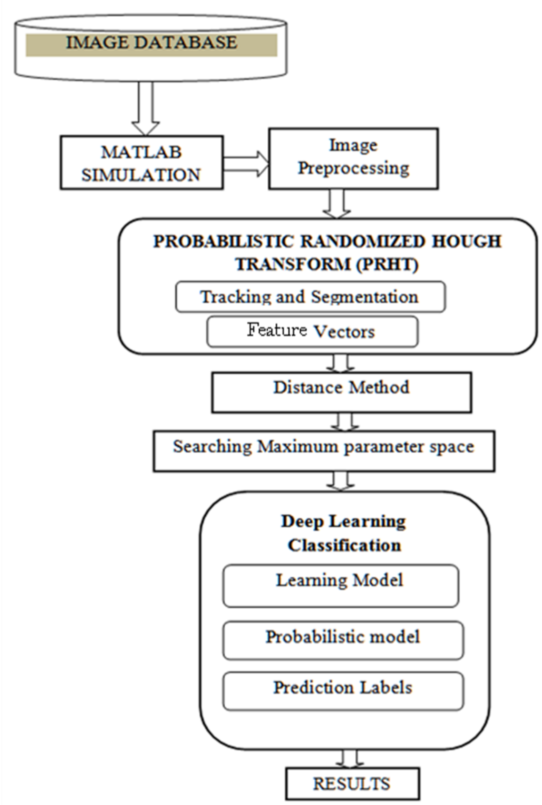


Fig 2: Flow Diagram of Proposed System

B. Probabilistic Randomized Hough Transform (PRHT)

The Probabilistic Randomized Hough Transform (PRHT) has been recognized as a very powerful tool for the detection of parametric curves in images. It is implemented by a voting process that maps image edge points into manifolds in a properly defined parameter space. The Circular Hough Transform (CHT) is one of the modified versions of the HT. The purpose of using CHT is to find the circular patterns within an image scene. The CHT is used to transform a set of feature points in the image space into a set of accumulated votes in a parameter space. Then, for each feature point, votes are accumulated in an accumulator array for all parameter combinations. The array elements that contain the highest number of votes indicate the presence of the shape. A circle pattern is described by (1) where x0 and y0 are the coordinates of the center and r is the radius of the circle.

PRHT on an image, the image is segmented into smaller blocks. For each block, PRHT is applied and the specific number of feature points is extracted. The block size reduces when the number of blocks is large, due to which the image content present in the block is also reduced, leading to better extraction of features. Thus, for each block, the number of features present is less and extraction of these features leads to precise choosing of the prominent peaks. The block size is decided by the number of blocks the image is segmented into. For N number of blocks, the image consists of √ N number of blocks along rows and √ N number of blocks along columns. The block image dimension is given by eqn. 1.

$$Block\ Width = \frac{Image\ width}{\sqrt{N}}$$

$$Block\ Height = \frac{Image\ height}{\sqrt{N}}$$

Algorithm 1: PROBABILISTIC RANDOMIZED HOUGH TRANSFORM (PRHT)
Input: Images *I*; candidate curve *Cs*; Features *f*
Output: *PRHT features*
Process
Step 1: Randomly sample a number of pixels and implement a converging mapping into a feature point $f \in Cs$
Step 2: Compute the spatial location function,

$$h(x) = [(I(x) - I_{min}) / (I_{max} - I_{min})] * h_{max} + h_{min},$$

h_{max} is the target maximum value; h_{min} is the target minimum value;
Step 3: Compute weight function

$$PRHT(x) = \sum_{x \in I(x)} w(h(x), I(x))$$

where w(h(x), $I(x)$) represents the weighting function that measures the similarity between the local neighbourhoods of the pixel at the spatial locations.

Step 4: Compute the normalized weighting image function

Step 5: Store PRHT training features of the normalized training vectors image

C. Distance Measures

a. Bhattacharyya Distance

The Bhattacharyya distance measures the similarity of two discrete or continuous probability distributions. It is closely related to the Bhattacharyya coefficient which is a measure of the amount of overlap between two statistical samples or populations. Both measures are named after A. The coefficient can be used to determine the relative closeness of the two samples being considered. It is used to measure the separability of classes in classification and it is considered to be more reliable than the Mahalanobis distance, as the Mahalanobis distance is a particular case of the Bhattacharyya distance when the standard deviations of the two classes are the same.

$$\text{Bhattacharyya} = \sum_{i=1}^{n} \sqrt{\Sigma a_i . \Sigma b_i} \qquad eqn.(2)$$

b. Manhattan Distance

The Manhattan distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance between two items is the sum of the differences of their corresponding components. The formula for this distance between a point X=(X_1, X_2, etc.) and a point Y= (Y_1, Y_2, etc.) is:

$$\text{Manhattan} = \sum_{i=1}^{n} |X_i - Y_i| \qquad eqn.\ (3)$$

Where n is the number of variables, and X_i and Y_i are the values of the i^{th} variable, at points X and Y respectively.

D. Deep Learning Classification (DLC)

DLC are specialized type of Neural network NNs that was originally used in image processing applications. They are arguably most successful models in AI inspired in biology. The input image is partitioned into a group of non-overlapping rectangles and a maximum value is given for each such sub-region. The research work use max-pooling in vision for the following reasons-The computation of upper layers is reduced by the removal of non-maximal values. Suppose a max-pooling layer is cascaded with a convolutional layer. The input image can be translated by a single pixel in 8 directions. 3 out of 8 possible configurations produce exactly the same output at the convolutional layer if max-pooling is done over a 2x2 region.

This jumps to 5/8 for max-pooling over a 3x3 region. A form of translation invariance is provided by this. The dimensionality of intermediate representations is reduced by max-pooling because it provides additional robustness to position.

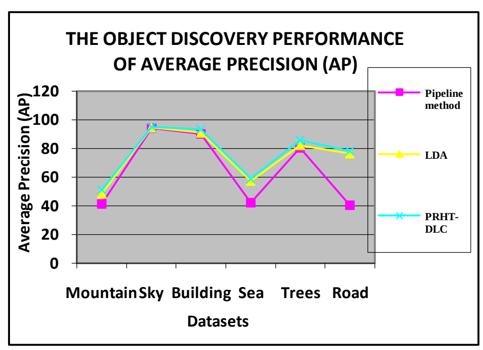
IV. PERFOMANCE EVALUATION

The research work taken as LabelMe-12-50k dataset consists of 50,000 JPEG images (40,000 for training and 10,000 for testing), which were extracted from LabelMe. Each image is 256x256 pixels in size. 50% of the images in the training and testing set show a centered object, each belonging to one of the 12 object classes shown in Table 6.1. The remaining 50% show a randomly selected region of a randomly selected image (“clutter”).

The dataset is a quite difficult challenge for object recognition systems because the instances of each object class vary greatly in appearance, lighting conditions, and angles of view. Furthermore, centered objects may be partly occluded or other objects (or parts of them) may be present in the image.

Table 1: THE OBJECT DISCOVERY PERFORMANCE OF AVERAGE PRECISION (AP).

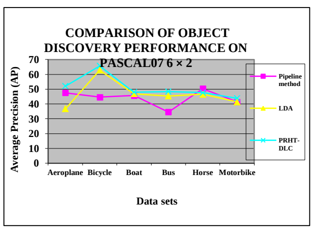
Methods	Mountain	Sky	Building	Sea	Trees	Road
Pipeline method	41.17	94.17	90.12	42.05	80.37	40.39
LDA	48.37	94.67	91.01	57.32	82.31	76.45
PRHT-DLC	**51.24**	**95.25**	**93.47**	**59.04**	**85.73**	**78.38**



The PASCAL07- 6x2 subset consists of images from 6 classes (aeroplane, bicycle, boat, bus, horse and motorbike) for Left and Right aspects of each class, resulting in a total of 12 class/aspect combinations. The PASCAL07-all subset consists of 42 class/aspect combinations covering 14 classes and 5 aspects (Left, Right, Frontal, Rear, Unspecified). From these results, the proposed can see that PRHT-DLC produced the highest log results on six categories.

Table 2: THE OBJECT DISCOVERY PERFORMANCE ON PASCAL07 6 × 2

Methods	Aeroplane	Bicycle	Boat	Bus	Horse	Motorbike
Pipeline method	47.51	44.62	45.73	34.51	50.31	41.40
LDA	36.71	62.93	46.84	45.39	46.61	41.40
PRHT-DLC	**52.12**	**65.71**	**47.96**	**48.40**	**47.68**	**43.88**



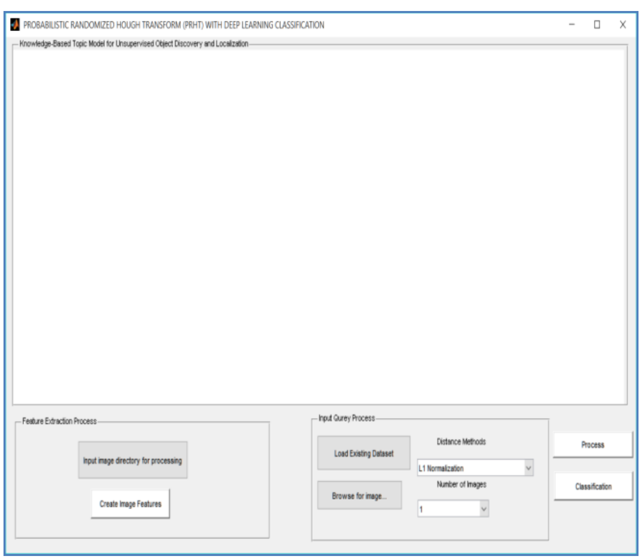


Fig.A1: Main Graphical User Interface (GUI) Window

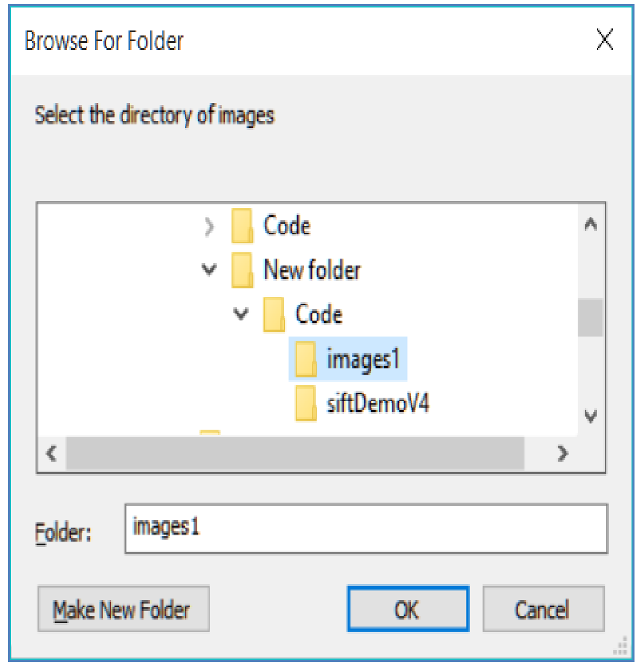


Fig.A2: Open Images for Feature Extraction Process

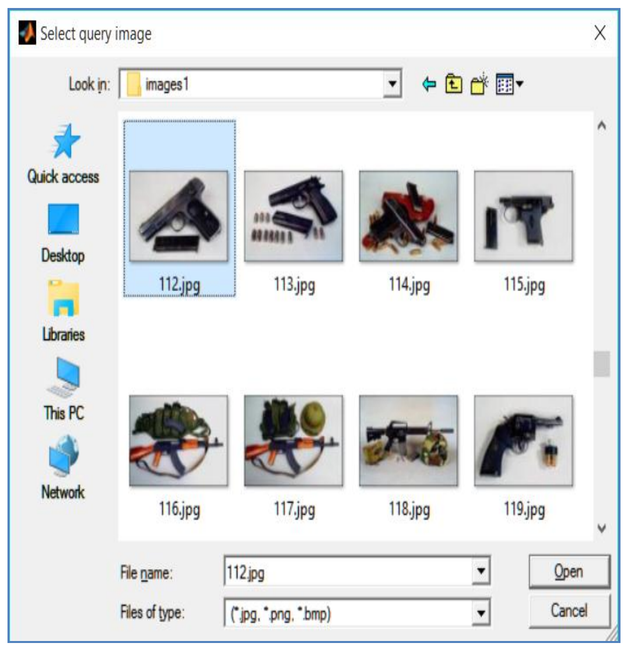


Fig.A3: Selecting Query image for discovery process

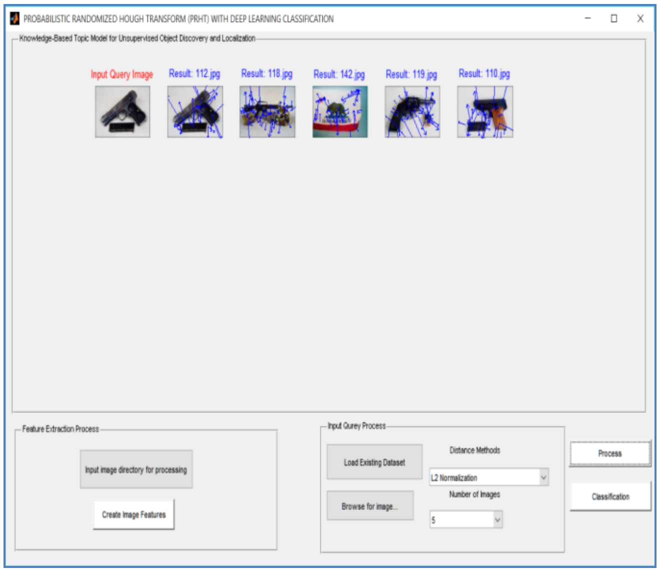


Fig.A4: Discovery Classification Result

V. CONCLUSION

This paper presents a Probabilistic Randomized Hough Transform (PRHT) with Deep Learning Classification (DLC) algorithm Framework. A unique method for PRHT feature extraction using probabilistic Hough Transform Peaks, has been proposed in this research. As a result, the visual knowledge is efficiently exploited and incorporated into topic modeling, and hence topic coherence is significantly improved to facilitate object discovery and localization. The extensive experiments on the Caltech, LabelMe and PASCAL datasets demonstrated the advantages of the proposed model for improving topic coherence. It is shown that our method significantly outperforms the unsupervised methods for object discovery and localization. The experimental results are analyzed with several constrains such as number of dimensions versus objective and average precision. Based on the PRHT-DLC results generated on this research, it concludes that accuracy and performance increases compared to the previous Pipeline method and Latent Dirichlet Allocation (LDA) algorithms.

REFERENCES

[1] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In CVPR, 2010.

[2] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in Proc. EMNLP, 2011, pp. 262–272.

[3] T. Deselaers, B. Alexe, and V. Ferrari, "Weakly supervised localization and learning with generic knowledge," Int. J. Comput. Vis., vol. 100, no. 3, pp. 275–293, 2012.

[4] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Leveraging multi-domain prior knowledge in topic models," in Proc. IJCAI, 2013, pp. 2071–2077.

[5] M. Rubinstein, J. Kopf, C. Liu, and A. Joulin, "Unsupervised joint object discovery and segmentation in Internet images," in Proc. CVPR, Jun. 2013, pp. 1939–1946.

[6] A. Faktor and M. Irani, "Clustering by composition'—Unsupervised discovery of image categories," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 6, pp. 1092–1106, 2014.

[7] A. Joulin, K. Tang, and L. Fei-Fei, "Efficient image and video co-localization with Frank–Wolfe algorithm," in Proc. ECCV, 2014, pp. 253–268.

[8] Z. Niu, G. Hua, X. Gao, and Q. Tian, "Semi-supervised relational topic model for weakly annotated image recognition in social media," in Proc. CVPR, 2014, pp. 4233–4240.

[9] L. Haldurai and V. Vinothini, "Parallel Indexing on Color and Texture Feature Extraction using R-Tree for Content Based Image Retrieval", International Journal of Computer Sciences Engineering, Volume-03, Issue-11, pp.11-15, 2015.

[10] C. Wang, K. Huang, W. Ren, J. Zhang, and S. Maybank, "Largescale weakly supervised object localization via latent category learning," IEEE Trans. Image Process., vol. 24, no. 4, pp. 1371–1385, 2015.

[11] M. Cho, S. Kwak, C. Schmid, and J. Ponce, "Unsupervised object discovery and localization in the wild: Part-based matching with bottomup region proposals," in Proc. CVPR, pp. 1201–1210, 2015.

[12] L. Haldurai, T. Madhubala and R. Rajalakshmi, "A Study on Genetic Algorithm and its Applications", International Journal of Computer Sciences Engineering, Volume-04, Issue-10, pp.139-143, 2016.

[13] Zhenzhen Wang ; Junsong Yuan, "Simultaneously Discovering and Localizing Common Objects in Wild Images", IEEE Transactions on Image Processing, Vol: 27, Issue: 9, 2018.

[14] Mereena Johny and L. Haldurai, "A Brief Survey on Dynamic Topic Model for Unsupervised Object Discovery and Localization", International Journal of Computer Sciences Engineering, Volume-06, Issue-09, pp.567-571, 2018.

[15] Zhenxing Niu, Gang Hua, Le Wang, Member, and Xinbo Gao, "Knowledge-Based Topic Model for Unsupervised Object Discovery and Localization", IEEE transactions on image processing, vol. 27, no. 1, 2018

Authors Profile

Mereena Johny M.sc, Pursuing Mphil Research Degree in Kongunadu Arts and Science College at Coimbatore.She did her UG and PG degree in Kongunadu Arts and Science College at Coimbatore.

Mr. L. Haldurai has done his Master of Computer Applications and Master of Philosophy at Bharathiar University. He currently works as Assistant Professor in Department of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore. His research work focuses on Data Mining, Image Mining and Data Communications. He has 10 years of Teaching experience and 5 years of Research Experience.