iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

A Review on Image Annotation Generation

Ms. A.S. Aher [1], Prof. Dr. S. M. Kamalapur[2]
[1]*M.E. Studen* , [2]*Associate Professor) Department of computer Engineering, K. K. Wagh Institute of Engineering Education & Research, Nashik Savitribai Phule Pune University, Maharashtra, India.*

Abstract*: *Image annotation, tagging, semantic descriptors are used to search an image from large image dataset. To search similar image, low level feature matching is used. In text based image search process image tags, annotations or image semantics are used. Annotation can be generated using manual, semi-automatic or automatic process. The manual and semi-automatic annotations are time consuming. In automatic image annotation, tag is automatically assigned to an image. There are various techniques for annotation generation. This work aims to study various image annotations generation, image matching and image searching techniques based on execution strategies and efficiency evaluation.

Keywords: *Community detection, co-occurrence model, feature extraction, image annotation, semantic descriptors.*

I. INTRODUCTION

In machine learning process, image matching and image retrieval are important techniques. These techniques are adopted by various search engines such as google, bing, yahoo, etc. Large number of search engines applies search over the surrounding text present in image, image names and provided manual tags.

In recent work, content based image retrieval is emerging technique. For such retrieval the meaningful descriptors of images are required. These descriptors can be generated manually, semi-automatic or automatic technique. The manual and semi-automatic techniques are time consuming.

Automatic tag annotation process is supervised learning process. It requires training phase. Using training dataset, testing images are annotated.

Apart from the content based image retrieval, semantic based image annotation and retrieval process is challenging task. In machine learning technique statistical inference can be evaluated from low level image visual features and image concepts. Using the evaluated inference semantic descriptor for an image can be generated. Such semantic description is also called as image annotation or image signature.

The statistical inference can be evaluated using various techniques such as network co-occurrence model, Semantic Concept Co-Occurrence Models, etc.

Paper is organized as follows: section I introduces the image annotation techniques and methods. Section II gives the literature review. Section III concludes the paper.

II. LITERATURE WORK

The approaches for various visual feature extraction, semantic concept co-occurrence model, image semantic descriptor generation and network analysis for community detection and image retrieval are discussed here.

A. Visual Feature Extraction

For identification of objects in an image, visual feature extraction technique is used. In visual feature extraction process various non redundant informative values are extracted from the given image.

Image may contain one or more objects and the background scene. For identification of objects in a scene minimum boundary rectangles [10] principle is used. In this technique, an envelope is created around the object in terms of polygon, rectangle or circle. These envelopes can be cropped and saved as an individual image representing a single object from set of object scene.

For feature extraction, colour GIST [2] is a method that summarizes the image gradient information in terms of scale and orientation. This GIST technique convolves the image with 32 graber filters and generates the 32 feature map. This feature map is then divided in equal 16 sections i.e. 4X4 grid and generates the average value of each grid. Assembling all the 16 section values for each feature map 16X32 feature values are extracted.

Pyramid of Histogram of Oriented Gradients (PHOG) [6] is a technique for object detection. This technique identifies the shape and spatial layout of the shape. For this, it counts the gradient orientation of object in an image. It initially applies the canny

edge detection technique to identify edges in an image. After edge detection it quantizes the gradient information of edges form 0 to 180^0rotation. PHOG with oriented edges [16] technique identifies the gradient information of edges from 0 to 360^0 rotation. These techniques use 4 level of pyramid.

Scale-Invariant Feature Transform (SIFT) [3] describe local features in images. To create feature description important points in a image are extracted. These points are referred as visual words. Visual words vocabulary is also called as bag of visual words. Image key point extraction, Speeded Up Robust Features (SURF)[24] is one more technique. This technique is quite faster than SIFT technique. SURF is based on Hessian matrix technique. SURF evaluates the approximation of image descriptors using integral images.

B. Semantic Concept Co-Occurrence Models

Image annotation is nothing but labelling to the image. These annotations are used to identify or search an image or set of similar images. These annotations can be provided manually or can be derived by analysing visual features of objects. Generally images are annotated using the objects containing in an image, location, landmark, etc.

Two or more objects may co-occur in multiple images. The idea to predict the co-occurrence factor of multiple images and order sequence annotation analysis [17]. The pair wise concept occurrence [7] [11] technique generates co-occurrence matrix of label co-occurrence count.

Multiple objects co-occurrence [8] [12] i.e. multiple concept learning is the technique uses multi-correlation probabilistic matrix factorization model. This technique generates image-with-concept, concept-with-concept and image-with-image correlation information. To refine the generated annotations, correlation concept is used. This technique is similar to the collaborative filtering technique [1]. The collaborative model generated based on information tapestry on various online sources. This technique is also useful in recommendation [20] and image retrieval system [15]. In some cases co-relation can be studied in terms of graph theory. Tree structure of spatial information [21] includes the information related to the object, its location, positive model information and negative model information.

Hierarchical image annotation [13] uses concept of ontology. The ontology is generated using semantic and visual concepts and finds inter-concept correlation and generates the concept hierarchy.

The basic limitation of these techniques is the lack of dataset information for training to generate co-relation information. But data provider like Google, wordnet do not consider their visual co- occurrence. Google uses web co-occurrence while wordnet provide semantic meaning similarity information.

The concept co-occurrence model is applicable in object detection and labelling technique. It provides some unspoken clues that help to generate appropriate image annotations. In visual feature extraction process there may have ambiguity in object detection. This model provides pair wise constraints that help to resolve ambiguity.

C. Image Semantic Descriptors

Image semantic descriptors define the natural language processing sentence to define an image. This sentence is based on the visual features. Semantic description process helps in bridging gap between low level image features and high level image semantics. Semantic description helps in image grouping, categorization and also in searching process.

Semantic descriptors [22] [21] are generated based on image/scene analysis. The generated semantic descriptors are used for image annotation and image retrieval process. Ali et al. [18] proposed a technique to generate image semantic descriptors. After analysing the visual features, the meaningful triplet is generated as an intermediate step. This triplet is called as meaning space. It includes object, action and scene. Using the triplet objects it defines the scene semantics. This Triplet is only used for semantic description generation and not used for finding similar objects or for defining the scene similarity.

For image classification, attribute generation [14] is one of the techniques. This technique is applicable for classification when there is insufficient availability of training dataset. Image interpretation based on visual feature recognition and cross category similarity identification [20] technique mainly focuses on vehicle and animal dataset. It do not simply detects the object but annotate the localize parts and compare the sharable localized objects.

D. Network Structure and Community Detection:

In an online social networking sites user can add multiple images and can tag those images with one or more descriptive keywords. These tags can be used to analyse the relationship among multiple objects.

Liu et al. [15] proposed network structure in terms of graph where each objects represents the graph node and edge weight represents the number of co-occurrence count. Using this graph structure multi-model similarity is generated. Video based similarity search technique [9] proposed using random walk technique. Flickr distance [23] used to measure correlation between concepts. A latent topic language model (LTVLM) is constructed in which Jensen-Shannon (J-S) divergence distance is used as a flickr distance. Using LTVLM visual conceptual network (VCNet) is generated to store information regarding conceptual relationship among multiple objects.

Based on the network structure, community can be detected using various techniques such as graph partitioning, hierarchical clustering, etc.

Community analysis is again a technique that provides more detailed relationship analysis of underlying communities. Modularity optimization [4] is a technique for community analysis based on hierarchical clustering. This technique identifies the co-occurrence patterns from generated communities based on the closeness level.

Linan Feng, Bir Bhanu [25] proposed semantic concept and co-occurrence pattern framework is applied to generate image annotations. These annotations are used for image storage and retrieval. Initially each object feature is extracted using visual feature extraction technique and then mapping of visual features to the concept semantics is applied.

The related work of visual feature extraction that uses minimum bounding rectangles for identification of objects, colour GIST for image gradient information, PHOG for object detection, SIFT for image features is described here. The Semantic Concept Co-Occurrence Models predict the co-occurrence factor. The image semantic descriptor generates sentence description for an image. Network structure discovers the co-occurrence pattern in network.

III. CONCLUSION

There are various techniques that help in image annotation and refinement of image annotations. Various techniques are studied independently like to extract visual feature for object detection, annotation generation or recommendation based on co-occurrence network pattern, etc. There is a need to develop a system that provides an efficient and faster solution to bridge the gap between visual features and concept annotation.

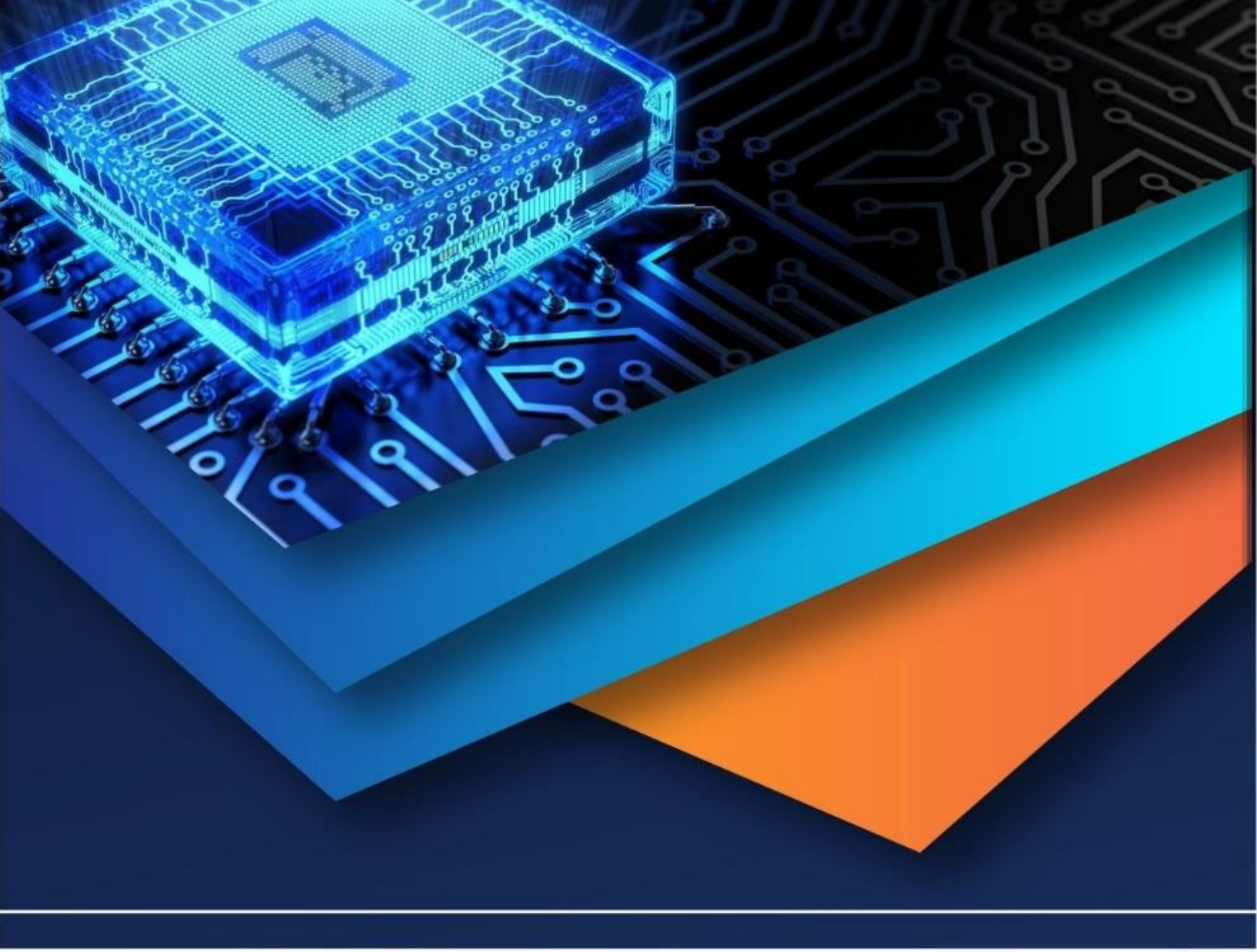
IV. ACKNOWLEDGMENT

Authors would like to thanks Prof. Dr. K. N. Nandurkar, Principal and Prof. Dr. S. S. Sane, Head of Department of Computer Engineering, K.K.W.I.E.E.R., Nashik for their kind support and suggestions. We would also like to extend our sincere thanks to all the faculty members of the department of computer engineering and colleagues for their help.

REFERENCES

[1] Goldberg, D. Nichols, B. M. Oki, and D. B. Terry, 1992, Using collaborative filtering to weave an information tapestry.
[2] A. Oliva and A. Torralba, 2001, Modeling the shape of the scene: A holistic representation of the spatial envelope.
[3] D. Lowe, 2004, Distinctive image features from scale-invariant keypoints.
[4] M. E. J. Newman, 2004 ,Fast algorithm for detecting community structure in networks
[5] S. Uchihashi and T. Kanade, 2005, Content-free image retrieval by combinations of keywords and user feedbacks
[6] A. Bosch, A. Zisserman, and X. Munoz, 2007, Representing shape with a spatial pyramid kernel.
[7] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie,2007 ,Objects in context.
[8] J. Fan, Y. Gao, and H. Luo, 2007,Hierarchical classification for automatic image annotation.
[9] W. H. Hsu, L. S. Kennedy, and S-F. Chang, 2007, Video search re-ranking through random walk over document-level context graph,
[10] Jordan Wood ",Minimum Bounding Rectangle pp 660-661," in Springer US, 2008,pp 660-661.
[11] C. Galleguillos, A. Rabinovich, and S. Belongie, 2008, Object categorization using co-occurrence, location and appearance
[12] J. Fan, Y. Gao, and H. Luo, 2008, Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel Image annotation.
[13] C. H. Lampert, H. Nickisch, and S. Harmeling, 2009, Learning to detect unseen object classes by between-class attribute transfer.D. Liu, X. Hua, L. Yang, M. Wang, and H. Zhang, 2009, Tag ranking
[14] L. Torresani, M. Summer, and A. Fitzgibbon, 2010, Efficient object category recognition using classemes
[15] S. Hwang and K. Grauman, 2010, Reading between the lines: Object localization using implicit cues from image tags.
[16] F. Ali, M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth, 2010,Every picture tells a story: Generating sentencesFrom images.
[17] A. Farhadi, I. Endres, and D. Hoiem, 2010, Attribute-centric recognition for cross-category generalization.
[18] F. Cacheda, V. Carneiro, D. Fernndez, and V. Formoso, 2011, Comparison of collaborative filtering algorithms: Limitations of current techniquesproposals for scalable, high-performance recommender systems.
[19] B. Siddiquie, R. S. Feris, and L. S. Davis, 2011, Image ranking and retrieval based on multi-attribute queries.
[20] F. X. Yu, R. Ji, M. H. Tsai, G. Ye, and S.-F. Chang, 2012, Weak attributes for large-scale image retrieval.

[21] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li, 2012, Flickr distance: A relationship measure for visual concepts
[22] Farhadi, I. Endres, and D. Hoiem, 2010, Attribute-centric recognition for cross-category generalization
[23] P M Panchal1, S R Panchal, S K Shah3,2013, A Comparison of SIFT and SURF.
[24] Linan Feng, Bir Bhanu, 2016, Semantic Concept Co-Occurrence Patterns for Image Annotation and Retrieval.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY