

# An Extended Survey on Various Techniques used in Automatic Image Annotation

K. Kishore Anthuvan

Assistant Professor, Dept of CSE,  
Christ College of Engg. & Tech,  
Pondicherry, India.

C. Jayanavithraa

Final Year M.Tech, Dept of CSE,  
Christ College of Engg. & Tech,  
Pondicherry, India.

T. Kamaleshwar

Final Year M.Tech, Dept of CSE,  
Christ College of Engg. & Tech,  
Pondicherry, India.

**Abstract** - Automatic Image Annotation (AIA) is an automatic methodology that maps low-level visual options for the high-level semantic options of the given image. Image annotation approaches would like an annotated dataset to be told a model for the knowledge between images and words. Unfortunately, preparing a labeled dataset is extremely time consuming and expensive. During this paper we tend to plan the comparison on varied techniques and algorithms to give optimized and efficient automatic image annotation.

**Keywords** - Automatic image annotation; semantic gap; semi supervised learning; artificial neural network; genetic rule.

## I. INTRODUCTION

A picture may be a resemblance of past recollections that is cherished by each individual all their life. Over the years the numbers of images being captured and shared have fully grown exponentially. There are many factors to blame for this growth. Firstly, in recent days the digital cameras permit individuals to capture, edit, store and share high quality pictures with great ease compared to the previous film cameras. Secondly, the provision of low value of memory and hard disk drives. Thirdly, the recognition of social networking sites like Facebook, MySpace have given the user a further interest to share photos on-line with their friends across the world.

The typical methodology of bridging the semantic gap is thru the automated image annotation (AIA) that extracts semantic options using machine learning techniques. Automatic image annotation (also referred to as automatic image tagging or linguistic indexing) is that the method by that a ADP system mechanically assigns data within the type of captioning or keywords to a digital image. This application of computer vision techniques is employed in image retrieval systems to prepare and find pictures of interest from information. With the exception of Automatic Image Annotation, annotation may be done manually. But the latter technique was being time overwhelming and involving sizable overhead motives the employment of the previous technique i.e. Automatic Image Tagging.

## II. LITERATURE SURVEY

### A. Semi-supervised approach

According to [1] Automatic image annotation using semi-supervised generative modeling. An annotation scheme within the semi-supervised framework to include untagged pictures into the coaching section. A generative modeling approach in 2 steps. In the initial, an initial mixture model is made for every thought using its labeled pictures. Within the second step, the parameters of mixture models are updated through incorporating the untagged pictures.

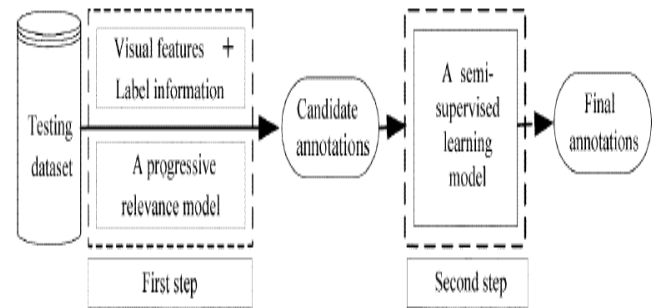


Fig. 1. Semi-supervised learning model

Image annotation plays a serious role for categorization the images in massive datasets and photo-sharing societies. An annotation scheme within the semi-supervised frame exertion to include untagged pictures in to the coaching section. In this system a generative modeling approaches in 2 steps. In the first step, an initial mixture model is made for every conception using its labeled pictures. This step is resembling the utmost chance (ML) estimation of parameters in mixture models supported the labeled pictures. Within the second step, the parameters of mixture models are updated through incorporating the untagged pictures. Since the untagged pictures generate incomplete observations for clusters, they tend to personalized EM rule steps to use the untagged signatures for change parameters of Gamma distributions fitted in to the clusters. In spite of the many semi-supervised learning algorithms, the planned approach is inductive and will annotate each take a look at image not enclosed within the training section.

Comparing to the supervised annotation systems, it reaches higher accuracy when incorporating the untagged samples. However, in our rule, the coaching section desires longer than the supervised annotation, as a result of tend to used repetitive EM rule with an repetitive nature. Additionally reveal that SSL rule can improve the annotation metrics for the datasets that aren't organized supported PSU protocol. One of the greatest challenges is related to the extraction of correct options from pictures. Using multiple descriptors for texture or new options like form may improve the annotation results. Our system doesn't take into account the relation between completely different words. An appropriate model for words correlations or their co-occurrences may facilitate to refine unrelated words doled out to a picture.

*B. Statistical modeling approach*

In this paper[2], incontestable the statistical modeling approach to the matter of automatic linguistic indexing of images for the aim of image retrieval. They tend to used classified images to train a lexicon of many ideas automatically. Wavelet-based options to describe native color and texture within the pictures. When analyzing all coaching pictures for a plan, a 2 dimensional multi resolution hidden markov model (2DMHMM) is made and hold on in a very thought lexicon. Image in one class are thought to be instances of a stochastic process that characterizes the class. To live the extent of association between a picture and therefore the matter description of an image class, tend to calculate the chance of the incidence of the image supported the stochastic process derived from the class. The incontestable that the planned strategies is used to train models for 600 completely different linguistics ideas and these models is wont to index images linguistically.

- The major blessings of this approach was
- Models for various ideas is severally trained and retrained
- A relatively large number of ideas is trained and hold on, and
- Image pixels among and across resolutions is taken into thought with probabilistic chance as a universal live

The system implementation and therefore the analysis methodology have many limitations. They train the concept lexicon using only 2d images while not a way of object size. It's believed that the thing recognizer of persons is sometimes trained using 3D stereo with motion and a way of object sizes. Training with 2nd still pictures probably limits the power of accurately learning ideas. As seen by one among the anonymous reviewers, the COREL image information isn't ideal for training the system as a result of its biases. For example, images in some classes, e.g.,

“tigers,” are rather more alike than a general sampling of images depicting the thought. On the opposite hand, pictures in some classes, e.g., “Asia,” are widely distributed visually, creating it not possible to coach such a plan using solely a small collection of such images. Until this limitation is completely investigated, the analysis results according should be understood cautiously.

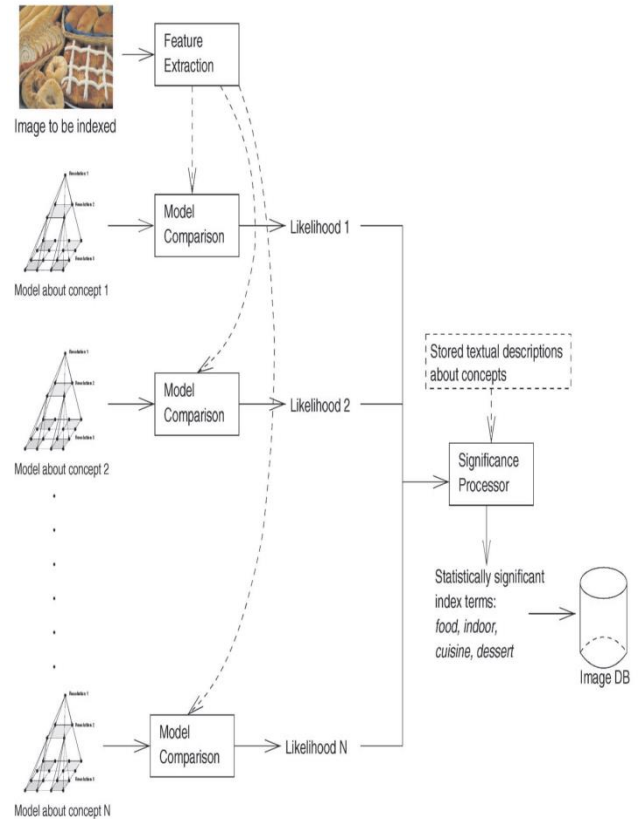


Fig. 2. Statistical modeling architecture

The hybrid strategies [3], that mix each world and native features for image illustration. A feature vector, may be a measurement of various features of image, would be the input of classification network. To create a feature vector using native options, 1st image shall section into completely different region before extracting options. Varied image segmentation techniques planned within the past however still not found correct methodology. Process value of native methodology is higher compared to world methodology because in global methodology no need to section a picture. For image annotation [4], low-level vector's options are calculated iteratively for every region within the image, either by using Hu moments, by Legendre moments or by Zernike moments. These vector's options are fed into the input layer of the neural network that's already trained, where every of the input neurons or nodes corresponds to every part of those options. And therefore the output neurons of the neural network represent the category labels of images to be classified and

annotated. Then each region is annotated by the corresponding label found by neural network classifier.

The input layer of the neural network incorporates a variable variety of input nodes. Regarding feature's extraction methodology, it's seven input nodes once adopting Hu moments, 9 input nodes once adopting Zernike moments, and 10 input nodes once using Legendre moment. However, the quantity of input nodes of the neural networks is modified or accrued once using Zernike and Legendre moments, as a feature's extraction methodology, so as to extend the accuracy of the annotation system. an image annotation system using region growing as image segmentation rule, moments and Multilayer Neural Network. For input feature vector for neural network they use Hu moments, Legendre moment and Zernike moment for feature extraction methods.

C. Baseline strategy

Image Annotation Refinement using Dynamic Weighted voting based on Mutual information [5] when result obtained by classification network might not be satisfactory because it may contain noise i.e. keywords that aren't associated with the image. It's necessary to refine this annotation result because several of annotation keywords are inappropriate for image content. Several researchers have strived to create novel algorithms of automatic annotation to enhance the standard of annotation. This paper technique utilizes global low-level image features and a simple combination of basic distance measures to search out nearest neighbors of a given image. The keywords are then assigned employing a greedy label transfer mechanism. The planned baseline methodology outperforms these progressive strategies on two standards and one large internet dataset.

A baseline live can give a robust platform to match and higher perceive future annotation techniques. Baseline strategies mix basic distance measures over terribly simple world color and texture options. K-Nearest Neighbors computed using these combined distances type the idea of our simple greedy label transfer rule. Our thorough experimental analysis reveals that nearest neighbors, even once using the individual basic distances, will outstrip variety of existing annotation strategies. What is more, a straightforward combination of the essential distances (JEC), or a mixture trained on noisy labeled information (Lasso), outperforms the most effective progressive strategies on 3 completely different datasets. These somewhat surprising results build a case for revisiting the progressive strategies and punctiliously analyzing their completely different modeling and coaching steps to know why they fail to attain performance at the amount of those oversimplified baseline strategies.

D. Neural network approach

Artificial neural network [6] may be a learning network which will learn from examples. Once network is trained by learning strategies with training sample information, it will build call for new sample. Typically ANN may be a multi-layer network build by assortment of interconnected nodes referred to as somatic cell. It consists of three layer neural network. 1st layer is input layer that has neurons adequate to dimension of input sample. Second layer is hidden layer, in hidden layer the selection of variety of hidden layer and therefore the variety of neuron at every layer is open problems in ANN approaches and therefore the output layer contain neuron adequate to variety of categories. Every connecting edge between neuron of various layers is related to weight. An activation function that generates output supported weight of connecting edge and output of previous layers neurons. Learning methodology is has to train neural network like back-propagation method.

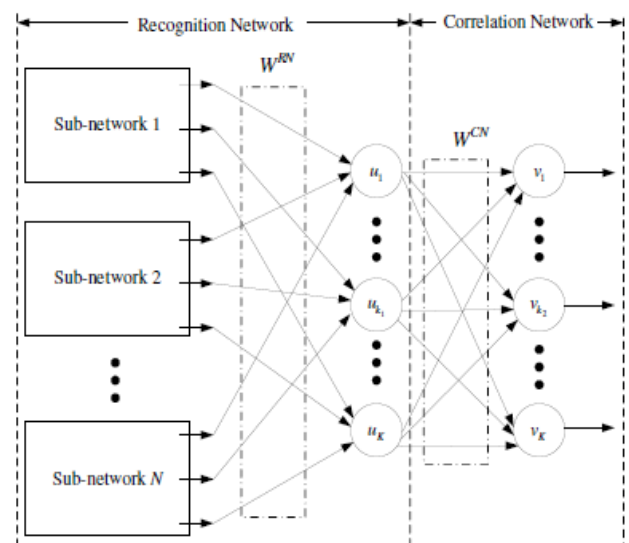


Fig. 3. Neural Network model

Neural Network models have certain common characteristics. They're given a group of inputs  $X=(X_1, X_2, \dots, X_m)$  and their corresponding set of outputs  $Y=(Y_1, Y_2, \dots, Y_n)$  for a definite method. Here the input  $X$  may be a visual feature vector of the image and output  $Y$  may be a keyword vector that is use to label image. The output turn out by a neuron for given input  $X$  is calculated by following formula:

$$Y=f(WX+B) \tag{1}$$

Here (1),  $W$  is that the weight matrix, that is that the weight matrix of connection link weight (synaptic weight) between neurons.  $B$  may be a bias vector of the layer of network. The ultimate output depends on the transfer operate  $f(\cdot)$ , that method the incoming data from other neurons. With the right choice of parameter  $W$  and  $B$  will get desired target output. This alternative is completed in training section of network. Training a neural network

means that adapting its connections so the model offers the required process behavior for all inputs.

This model is divided into 2 networks, Recognition Network and Correlation Network. Recognition network has set of sub network and therefore the variety of sub network is set by the quantity of divided region of a picture. Recognition network produces a keyword vector for the image and Correlation network enhances annotation performance by using the keyword correlation information. Each genetic rule and back propagation rule are wont to train recognition network whereas solely back propagation rule is employed to train correlation network and it's experimented on an artificial image dataset. Within the method of image annotation, image is 1st divided into many regions that produce a visible feature vector for every region. The input layer of recognition network receives the feature vector of a region and output layer would generate a keyword vector. This keyword vector indicates that keyword should be chosen to label the input region. Finally correlation network receives keyword vector from recognition network and refine annotation using keyword correlation data.

### III. CONCLUSION

The idea with automatic image tagging is that tags are automatically captioned and assigned to the digital image. These tags ought to describe each vital part or side of the image and its context. Automatic image tagging is done supported the visual content of the image, contextual data, or employing a mixture of those two approaches.

In this survey, I have conclude that neural Network offers higher performance once image is classifies by quite one class i.e. keywords. We can simply train neural network than alternative classification network and once a neural network is trained we will simply annotate sample images. Overall AIA system may be a very challenging task and research has been getting to develop efficient AIA system.

### REFERENCE

- [1] S. Hamid Amiri, Mansour Jamzad, Automatic image annotation using semi-supervised generative modeling 0031-3203/& 2014Elsevier
- [2] D. Zhang, Md. M. Islam, G. Lu, —A review on automatic image annotation techniquesl ELSEVIER Pattern Recognition 45 (2012) 346-362
- [3] H. Bouyerbou, S. Oukid, N. Benblidia, K. Bechkoum, —Hybrid Image Representation Methods for Automatic Image Annotation: A Surveyl ICSES 2012- International Conference on Signals and Electronic Systems WROCLAW, POLAND, sep-2012
- [4] Z.chen, H. Fu, Z. Chi, D. Feng, —An Aduptive Recognition Model for Image Annotationl, IEEE Transaction on Systems, Man, and Cybernetics—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 6, Nov. 2012
- [5] M. Oujaoura, B. Minaoui, M. Fakir, —Image Annotation using Moments and Multilayer Neural Networks, Special Issue of Internationa Journal of Computer Applications (0975-8887) on Software Engineering, Database and Expert systems-SEDEXS, sep-2012
- [6] Haiyu Song, Xiongfei Li, Pengjie Wang, —Image Annotation Refinement Using Dynamic Weighted Voting Based on Mutual Informationl JOURNAL OF SOFTWARE, VOL. 6, NO. 11, NOVEMBER 2011