

A Survey on Automatic Image Annotation and Retrieval

Adnan Siddiqui
SOIT, RGPV Bhopal

Nischcol Mishra
SOIT, RGPV Bhopal

Jitendra Singh Verma
SOIT, RGPV Bhopal

ABSTRACT

The Image Annotation process are required to use automated where the strong tagging is required to keep annotate image for making it efficient to provide better results while querying those image annotated database. The process of image annotation should be proceeding while creating the file and to make it strongly labelled. It is a process of machine learning where low level features of images are extracted, clustered and mapped to the semantic. This can be based on training set of data. Automatic image annotation technique can be based on various things either it can observe the images various ways either in texture bases, colour intensity basis or faces included or involved into the images. Next the features are grouped into the cluster and then annotation is done based the category. In this paper we study & describe the how the different automation techniques are working in order to annotate the datasets and then how they are useful in order to query process the data and to release the query work load and to reduce the computation process time. We have compared different annotation techniques on various parameter like segmentation, feature extraction, clustering etc. Furthermore number of models has been observed and reviewed in order to face the challenges being found in those and got rectified in proposed approach.

Keywords-Automatic image annotation, image Tagging, Image Query, image retrieval, annotation models.

1. INTRODUCTION-

Automatic Image Annotation is also known as linguistic indexing or automatic image tagging. Automatic Image Annotation is a process by computer system which automatically assigns meta-data in the form of keywords to a digital image. This application of computer technique is employed in image retrieval systems to organize and locate images of interest from a database.

Annotation Methods regarded to classify as a multi-class image with number of large classes as large as vocabulary size. Generally, image analysis in the form of extorted feature vector and machine learning technique are employed a words that are training annotation to apply automatically annotation to new images. This method is learned the co-relation between image features and training annotation then technique are developed to employed the machine translation to try to translate the textual vocabulary with the 'visual vocabulary' or clustered regions known as blobs. The major steps of annotation are as:

- Segmentation into regions
- Clustering to construct blob-tokens
- Analyze correspondence between key words and blob-tokens
- Auto Annotation

1.1 Segmentation into regions- The regions based image segmentation is basically as a pixel based image segmentation. The pixel based image segmentation involves the initial of seed points. Segmentation approach is examined as a neighboring pixel of seed points and after determining the pixels be should be added to the regions.

1.2 Clustering to construct blob-tokens- The blob are generate from the image features using clustering. When we have a set of images with annotations, we show that probabilistic models allow us to predict the probability of generating a word given the blobs in an image.

1.3 Analyze correspondence between key words and blob- tokens -Here we describe keywords and blob-tokens and identified the relation between blob-tokens and keywords. Here discover between hidden semantics.

1.4 Auto Annotation- In this section, we annotate the image automatically and also calculate the distance between the given image object and all centroids of blob-tokens and represent this image objects with the keywords of a closest blob tokens.

The goal of automatic image annotation is mostly to assist image retrieval by supplying users with a text based interface for search. If successfully images can be retrieved in a way that is similar to search of text documents as many people do on Google. We discuss the reasons for automatic image annotation from two perspectives, manual annotation and CBIR.

The biggest challenge in image annotation is that how we can contact the low-level features and high-level linguistic concept together. In the Automatic image annotation firstly image feature are extracted by colour, texture, shape and then create blobs-token by grouping a similar segmentation and then analyse between keywords and blob-tokens to hidden semantic. A efficient query can be now apply once the proper annotation techniques is applied and thus we can obtain efficient results.

2. AUTOMATIC IMAGE ANNOTATION

Image Annotation or tagging on the image automatically required its indexing and can be done on the various bases– the linguistics is the easiest way to distinguish which was specified in the paper [28] where they have specified that automatic linguistic indexing of pictures is essentially important to content-based image retrieval and computer Object recognition, the statically model specify here about the auto annotation of the images.

Automatic Tagging based on the tag prediction on processing the images and predicting the tagging on them based on their weight, dimensions sort of features but such kind of tagging can lack in accuracy but it is efficient in the case similar kind of image dataset.

Automatic Image Annotation is also known as linguistic indexing or automatic image tagging. Automatic Image Annotation is a process by computer system which automatically assigns meta-data in the form of keywords to a digital image. This application of computer application technique is employed in image retrieval systems to organize and locate images of interest from a database.

Here we are describing the different approaches for automatic image annotation these are:

- Generative model
- Discriminative model
- Graph model

2.1 Generative model: it is based on some sample data to find out hidden fact .It calculate joint probability distribution on observable data or sample data in order to find out text for annotation .Generative models are used in performing machine learning or conditional probability density function .Conditional probability is calculated by bayes' rule.

2.2 Discriminative model: These models do not depend on sample data. This models simply use machine learning for finding annotation word (w) on given unannotated image features (x).Conditional probability in this models can be shown in $P(w/x)$. This model directly assign word to image on unobserved data does not generate sample data like generative model. These models are inherently supervised and have greater performance than generative model. each conditional probability intercepted a classification of labelling .this is the main idea that converted automatic image annotation into classification problem that make easy to annotation under a common semantic label.

2.3 Graph model: Graph model says convert the concept as vertices and relationship among these as edges .thus AIA problem can be converted into graph model. These models have greater potential of improvement.

3. RECENT WORK IN FIELD OF AUTOMATIC IMAGE ANNOTATION

Mori et al. [29], “Image-to-word transformation based on dividing and vector quantizing images with words co-occurrence model”:

This is one of the first attempts at image auto-annotation. The word co-occurrence referred to words often used together. The word co-occurrence statically information is the major question to the natural image processing.They first divide images into rectangular tiles of the same size, and calculate a feature descriptor of colour and texture for each tile. All the descriptors are clustered into a number of groups, each of which is represented by the centroid. Co-occurrence is known as a linguistic term that can be explained as an indicator of an idiomatic expression or semantic proximity. Here we describe machine learning model that used in coocurrence model. History matrix-Records the interactions between users and items as a user-by-item matrix.

Co-occurrence matrix-Transforms the history matrix into an item-by-item matrix, recording which items appeared together in user histories Indicator matrix-Retains only the anomalous (interesting) co-occurrences that will be the clues for recommendation.

Duygulu et al [3], “Object recognition as machine translation: learning a lexicon for a fixed image vocabulary”:

Duygulu et al proposed a machine translation model for image auto-annotation. They argued that region based image annotation is more interesting because global annotation does not give information on which part of the image is related to which label. In their point of view, the process of attaching labels to image regions is analogous to the translation of one form of representation (image regions; French) to another form (labels; English).They first use a segmentation algorithm to segment images into object-shaped regions as shown in fig1.Then, feature quantization is applied to the feature descriptors that are extracted from all the regions, to build a visual vocabulary called ‘blobs’. A ‘blob’ is in fact a representative of a cluster of visually similar image regions. Finally, a machine translation model which was initially proposed for linguistic translation is adopted to build a ‘lexicon’, a translation table containing the probability estimations of the translation between image regions and labels. An unseen image is annotated by choosing the most likely word for each of its regions.

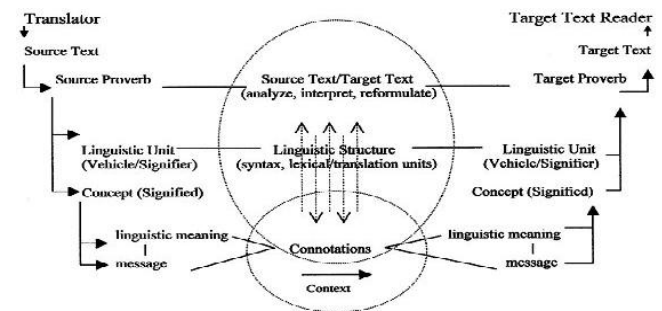


Fig 1: Translation Model

Jeon et al [6], “Automatic image annotation and retrieval using cross-media relevance models”:

Jeon et al proposed Cross Media Relevance Model (CMRM) where the vision information of each image was denoted as blob set which is to manifest the semantic information of image. However, blob set in CMRM was erected based on discrete region clustering which produced a loss of vision features so that the annotation results were too perfect. In order to compensate for this problem, a Continuous Relevance Model (CRM) was proposed in [7].Jeon et al improved the [3] by introducing a generative language model to image annotation. The same process as used was chosen to calculate the blob representation of images. However, as different from the assumption that there exists an underlying one-to-one correspondence between the blobs and words, they only assume that a set of blobs is related to a set of words. This model also shares the keywords of similar image have two types of expansion:-

Document based expansion: Here blob corresponding to each test image are used to find the words and associated probability. This can be done by find the joint probability distribution of blob and words. Each test image can be annotated with all words with associated probability. This type of CMRM is called Probabilistic Annotation-based Cross-Media Relevance

Model (PACMRM). This is useful for rank retrieval of data. Fixed length annotation can be done by restricting PACMRM with top N words. This is called Fixed Annotation-based Cross-Media Relevance Model (FACMRM). FACMRM is popular among the user when no. of annotation is small.

Query based expansion - Here set of blob probabilities are generated, from joint probabilities distribution of words and blobs, using query word.

Putthividhya D. et al [30], "Supervised topic model": This model can uncover the underlying themes of a collection and decompose its documents according to those themes. This analysis can be employed for corpus exploration, document search, and a collection of prediction problems. we describe two advanced topic models: Gaussian-Multinomial PLSA, Gaussian-Multinomial LDA. These three models are based on basic topic models seen above: PLSA and LDA. Given that these basic models are suitable only for modeling one-type of data, advanced topic models are designed to fit multi-type data. Gaussian-Multinomial PLSA (GM-PLSA):

GM-PLSA is a combination of two PLSA models: a standard PLSA to model textual words and a continuous PLSA to model visual features. These two models share a common distribution over latent variable z noted $P(z|d)$.

The whole model, which is represented in figure 3, assumes the following generative process:

1. Select a document d_i with probability (d_i)
2. Choose a latent aspect z_k with probability $(z_k|d_i)$ from a multinomial distribution conditioned on document d_i
3. For each of the words, sample w_m from a multinomial distribution (θ_k) conditioned on the latent aspect z_k .
4. For each of the feature vectors, sample r_n from a multivariate Gaussian distribution $(x|\mu_k, \Sigma_k)$ conditioned on the latent aspect z_k .

Gaussian-Multinomial LDA (GM-LDA):

GM-LDA is a combination of two LDA models: a standard LDA to model textual words and a continuous LDA to model visual features. This model, represented in figure 7, shows that words w_m and regions r_n of an image can come from different topics which means that the whole document can contain multiple topics. Furthermore, we can view θ as high-level representation of the whole document (image features + words).

The joint probability distribution of the set of image features r , the set of associated words w and latent variables θ , z and v is given as follows:

$$p(\mathbf{r}, \mathbf{w}, \theta, \mathbf{z}, \mathbf{v}) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(r_n|z_n, \mu, \sigma) \prod_{m=1}^M p(w_m|\theta) p(v_n|w_m, \beta)$$

GM-LDA model assumes the following generative process:

1. Sample a Dirichlet random variable θ
2. For each of the N image regions:
 - a. Sample $z_n \sim (\theta)$
 - b. Sample a region description r_n conditional on z_n

3. For each of the M words:

- a. Sample $w_m \sim (\theta)$
- b. Sample a word w_m conditional on w_m .

Himali Chaudhari et al [1]: This paper presents the survey of different approaches for automatic annotation and annotation based image retrieval. This paper aims to cover the latent space and generative approaches for automatic image annotation.

Shalini K. Kharkate, Prof. Nitin J. Janwe et al [2]: This paper proposed an algorithm to annotate image by comparing test image feature vector with feature matrix of training data sets and similar and dissimilar image pairs. Ranking technique is used for transferring the key word from similar image pair to the test image by counting local frequency of keywords [2].

Recent techniques for AIA based image retrieval generally divided into two types of approaches, the probabilistic modelling methods and the classification methods.

The probabilistic modelling methods: The probabilistic modelling methods aim to develop a relevance model to represent the correlation or joint probability distribution between images and keywords [4]. [3] Propose to treat image annotation as a process of machine translation. They introduced a Translation Model (TM) based on statistics. They used the method to translate a visual vocabulary into keywords. The other typical method is the Latent Dirichlet Allocation model [5].

Recently growing trend for retrieving the images statistical classification is used to group the images into rough semantic classes, such as graph-photograph, textured-non textured. thus image is categorized and by permitting adaptive semantic searching and by narrowing down the search space in database image retrieval get enhanced.

Feng S et al [8], "Multiple bernoulli relevance models for image and video annotation": Feng et al proposed to improve CMRM and CRM. Compared with other discrete models, these methods can evidently improve annotation accuracy. it is statistical generative model uses set of annotated training set for annotation. Images are divided into rectangular grid and real valued feature vector are computed using training set. These methods employ a nonparametric method to estimate a Gaussian distribution. Word probability is estimated using Bernoulli model and image feature probability is estimated using non-parametric kernel density function.

Tianxia Gong, Shimiao Li, Chew Lim Tan et al [9]: Tianxia et al proposed a framework of using language models to represent the word-to-word relation utilizing probabilistic models. On the other hand, the discriminative model trains a separate classifier from visual features for each tag. These classifiers are used to predict particular tags for test image samples [10,11]. Similarly, we can also train a regression model (regression coefficients) to predict tags for test images, taking features as predictors (input variables) and tags as responses (output labels). In image annotation and retrieval, SVM is a widely used machine learning method. SVM can generate a hyper plane to separate two data sets of features and provide good generalization. Annotation based image retrieval is based on the theory of text retrieval systems. Many document retrieval and indexing techniques were incorporated into IR Systems. In this section

we will discuss some of the important document retrieval techniques.

S. Deerwester et al [12], "Indexing by latent semantic analysis": Deerwester et al proposed Latent Semantic Indexing (LSI) as a document retrieval technique to address the some of the shortcomings inherent in traditional lexical matching techniques. LSI deals with the problems of synonymy (Many words refer to same object) and polysemy (Many words have multiple meanings). LSI tries to search for something that is closer to representing the underlying semantics of a document, rather than just by matching specific keywords. Latent semantic indexing method starts with the creation of terms by document matrix. Then this high dimensional matrix is further decomposed into a reduced dimension matrix called Singular Value Decomposition (SVD). This filters out the noise found in a document, such that two documents that have same semantics will be located close to one another in a multi-dimensional space. However LSI has some drawbacks such as reduced dimensions are difficult to interpret, SVD is computationally expensive, performance and speed level degrades when applied to large scale collection. As LSI has number of deficits due to its unsatisfactory and incomplete theoretical foundation.

Hofmann et al [14], "Probabilistic Latent Semantic Indexing": Hofmann et al proposed the probabilistic LSI (PLSI) model, as an alternative to LSI. The roots of PLSI go back to the LSI. Like LSI, PLSI also deals with synonymous as well as polysemous words. PLSI is an automated document indexing technique, in which each document is represented by its word frequency. PLSI is also known as the aspect model which is the Latent variable model in which latent variables are associated with observed variables. Consequently, it has a more robust statistical foundation, and is able to provide a proper generative data model. PLSI is based on the Expectation Maximization (EM) algorithm. While PLSI is one of the good text analysis technique it has some drawbacks such as it is incomplete since provide no probabilistic model at the level of documents, leads to over fitting problems if there are too many parameters in the model and it's not clear how to assign how to assign probability to a document outside of the training data. To address the limitations here proposed a unsupervised, generative model called Latent Dirichlet Allocation (LDA). It is closely related to PLSI. It is a powerful generative probabilistic model developed for modelling words in a document. In LDA each document is a mixture of a small number of latent topics, here each topic is characterized by a distribution over words.

Lavrenko et al. [7] argued that the process of quantization from continuous image features into discrete blobs, as the approach used by the machine translation model and the CMRM model, will cause the loss of useful information in image regions. Lavrenko et al. [7] also proposed a model in this paper named Continuous Relevance Space Model (CRM). Lavrenko et al. [7] proposed that each part of divided image can be best described by continuous-valued feature vector. Here continuous features are directly associated with words and does not require an intermediate clustering stage. This model also allows rank retrieval in response to a text query. CRM makes no assumption about topological structure. It is a kind of doubly non-parametric approach where expectation is computed on every individual point in training set. The main part of that model is a special mapping function G which maps image region r to real-valued feature vector g. Real valued feature vector specify the object location, color, shape in the original image.

R.Jin et al. [22] proposed a new framework named Coherent Language Model (CLM) and Active Learning. It considers the word to word correlation. Word to word correlation is some relation among predicated annotated word for an image. The word to word correlation is important particularly when image features are insufficient in determining an appropriate word annotation. The EM algorithm, for word to word correlation, is used to find optimal language model for given image. Active learning is used to reduce uncertainty by selectively sample examples for labeling. Active learning is used generally in iterative manner. This model automatically determines the annotation length. A variation of CLM named Coherent Language Model Flexible Length (CLMFL) provides more accurate result than CLM.

S.Zhang et al [21] proposed a technique of automatic image annotation using group sparsity. It regularizes the problem features related selection and solve annotation as a retrieval problem. It uses concept of regressor for representing similarity of image pair if regressor value is positive means pair are similar if negative pair are dissimilar. To incorporating the the keyword information similarity unction are used. The quality of pair similarity highly influence the overall performance. Image has many features so finding out similarity important features are used and other can be pruned by assigning weighted distance zero. Expectation maximization algorithm are used to feedback similar and dissimilar pair and iteratively improve the performance by finding better pair.

4. COMPARISON OF DIFFERENT IMAGE ANNOTATION TECHNIQUE

Table1 Comparison of different technique

S. No	Model Name	Author	Segmentation type	Object Recognition	Feature extraction	Annotation type
1	Co-occurrence Model	Mori et al	Each image in training set is divided into non-overlapped parts	Statical learning of image to word relationship.	Clusters are made by vector quantization technique	Probability for each word in each centroid is estimated statistically and word is assigned to that part of image
2	Translation Model	Duygulu	Region based segmentation.	K-means vector quantization.	Blobs extraction and then map to words	Word to blob is used to annotate.

3	CMRM Model	Jeon	Blob segmentation is performed.	Natural way probabilities and image annotation using shared keywords.	Based on blobs feature “blobs” extracted.	Use the keywords shared by similar images to annotate new images
4	CRM Model	Lavrenko	NA	Blob recognition.	Document based expansion and Query based expansion.	The process of quantization from continuous image features into discrete blobs.
5	Coherent Language Model	R.Jin, J.Liu	NA	Word to word is used and word is related or mapped to image.	Word to word relation.	Active learning is used generally in iterative manner. automatically determines the annotation length.
6	MBRM Model	Feng	The images are divided into rectangular grid and real value.	image feature probability is estimated using non-parametric kernel density estimate	The word probability is estimated using Bernoulli model	Image annotation are hierarchical and having greatly varying length.
7	Maximum Entropy	J.Jeon	NA	Rectangular regions of image and these regions is cluster, using k-means algorithm.	uniform distribution when no information is available. it automatically assign a weight to features	Annotation is based on maximum atrophy.
8	Group Sparsity Model	S. Zhang	No Segmentation, sparse feature	Word to word matching is done.	sparse feature extraction technique.	Annotation using group sparsely.
9	Tag Propagation	Matthieu Guillaumin	NA	weighted combination of the tag absence/ presence Among neighbors.	Based on distance weight technique.	Prediction based tagging and annotation.

5. CHALLENGES BEING FOUND WHILE REVIEWING THE LATEST PROPOSED MODEL-

Weak Labelling Results: Weak label techniques not give appropriate result or work while querying the image database.

Repeated Tagging: Training time is getting high while getting the same kind of tagging or annotation with the image in first model described.

Threshold Scheme: Time gape monitoring or observing the time gaping between the snap captured was not monitored with the current thresholding scheme to classify feature vector.

Unclear Images: Challenges being found to working with low dimension or low intensity image while working with annotating them.

Fixed Annotation: Challenges found in some technique for statically annotation and in some technique also we found that it takes long time to compute and finding proper annotation.

Multiple Mapping: We also find in some model that they perform multiple mapping for same image to perform annotation.

6. CONCLUSION

In this paper, we attempted to provide a comprehensive survey on automatic image annotation techniques. As a survey paper, we might not include each and every aspect of individual works; however we have focused on the latent space approaches and statistical models of automatic image annotation. Here we are describing how its work and what its advantages of these model, we have discussed recently developed model which have been used by the authors for image annotations and retrieval. In the literature review we have found number of challenges such as long computation time, weak signatures, lack of redundancy, static annotations and mapping issues been observed in previous references and thus we came to the point of requirement of such a technique which can improve the overcomes the challenges and give a best throughput solution to the challenges in less computation time, the techniques which are described are efficient but having lacking in some category which can be overcome by design based on the challenges or disadvantages.

7. REFERENCES

- [1] Himali Chaudhari et al, “A Survey on Automatic Annotation and Annotation Based Image Retrieval”, International Journal of Computer Science and Information Technologies, Vol. 5 (2) ,1368- 1371 , 2014.

- [2] Shalini K. Kharkate, Prof. NitinJ. Janwe, “A Novel Approach For Automatic Image Annotation Using Color Saliency”, *International Journal of Innovative Research in Computer and Communication Engineering* Vol.1, Issue 5, pp 1101-1108, July 2013.
- [3] Duygulu, P., Barnard, K., Freitas, J., Forsyth, D.A.,” Object recognition as machine translation: learning a lexicon for a fixed image vocabulary”, *Proceedings of 7th European Conference on Computer Vision*, Copenhagen, Denmark, pp. 97–112,2002.
- [4] Zhu, S., Liu, Y.,” Semi-supervised learning model based efficient Image annotation. “*IEEE Signal Process. Lett.* 16 (11), 989–992, 2009.
- [5] Blei, D.M., Jordan, M.I ,”Modeling annotated data.”,*Proceedings of 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’03)*, Toronto, Canada, pp. 127–134,2003.
- [6] Jeon J, Lavrenko V, Manmatha R.,” Automatic image annotation and retrieval using cross-media relevance models”. *Proc. of Int. ACM SIGIR Conf. On Research and Development in Information Retrieval*, Toronto, Canada,119-126,Jul.2003.
- [7] Lavrenko V, Manmatha R, Jeon J.” A model for learning the semantics of pictures”. *Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems*,pp119-126, 2003.
- [8] Feng S, ManmathaR ,Lavrenko V. “Multiple bernoulli relevance models for image and video annotation.” *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR’04)*, Washington DC, USA,1002-1009, Jun. 2004.
- [9] Tianxia Gong, Shimiao Li, Chew Lim Tan,”A Semantic Similarity Language Model to Improve Automatic image annotation”,*22nd International Conference on Tools with Artificial Intelligence*, p 197-203, 2010.
- [10] D. Grangier and S. Bengio, “A discriminative kernel-based approach to rank images from text queries”, *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 30, no. 8, pp. 1371–1384,2008.
- [11] Y. Chen, J. Wang, and D. Geman, “Image categorization by learning and reasoning with regions”, *The journal of Mach. Learn. Res.*, vol. 5, pp. 913–939,2004.
- [12] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, “Indexing by latent semantic analysis”, *J. Soc. Inform. Sci.* 41, 391_407, june 1990.
- [13] M.W. Berry, S.T.Dumais, and G.W. O’Brien, “Using Linear Algebra for Intelligent Information Retrieval,” *SIAM Rev.*, vol. 37, no. 4, pp.573-595, 1995.
- [14] T. Hofmann, “Probabilistic Latent Semantic Indexing,” *Proc.22ndInt’l Conf. Research and Development in Information Retrieval SIGIR’99*, 1999.
- [15] D.M. Blei and A.Y. Ng, and M.I. Jordan, “Latent Dirichlet Allocation”, *Advances in neural information processing systems conference*, vol. 3, pp. 601-608, 2003.
- [16] Carneiro, Gustavo, Chan, Antoni B, Moreno, Pedro J, Vasconcelos, Nuno, “Supervised learning of semantic classes for image annotation and retrieval”, *IEEE Computer Society*, pp 392-410, 2007.
- [17] Claudio Cusano, GianluigiCiocca, RaimondoSchettini,” Image annotation using SVM”,*Proc. SPIE 5304, Internet Imaging V*, 330, December 22, 2003.
- [18] Changbo Yang, Ming Dong,FarshadFotouhi, “Region Based Image Annotation Through Multiple Instance Learning”, *ACM*, 435-438, 2005.
- [19] Andrews, I. Tsochantaridis,T.Hofmann,“Support Vector Machine for Multi-Instance Learning”,*NIPS*,2002.
- [20] XiaojunQi,Yutao Han,” Incorporating multiple SVMs for automatic image annotation, *Pattern Recognition*”, Computer Science Department, Utah State University, pp 728 – 741, 2007.
- [21] Shaoting Zhang, Junzhou Huang, Hongsheng Li, and Dimitris N. Metaxas, *Automatic Image Annotation and Retrieval Using Group Sparsity*, *IEEE June 2012 (838-849)*
- [22] Rong Jin, Joyce Y. Chai, Luo Si, *Effective Automatic Image Annotation Via A Coherent Language Model and Active Learning*, *ACM*, pp 892-899, 2004
- [23] O. Maron and A. L. Ratan. “Multiple-Instance Learning for Natural Scene Classification”,*Fifteenth International Conference on Machine Learning*, Pages 341-349,1998.
- [24] P. Brown, S. D. Pietra, V. D. Pietra and R. Mercer. “The mathematics of statistical machine translation: Parameter estimation”, *Ininternational journal of Computational Linguistics*,19(2),263-311, 1993.
- [25] A. Berger, S. Pietra and V. Pietra. “A Maximum Entropy Approach to Natural Language”, *Processing. In Computational Linguistics*, pages 39-71, 1996.
- [26] FlorentMonay, Daniel Gatica-Perez,“On Image AutoAnnotation with Latent Space Models”, *Proceedings of the eleventh ACM international conference on Multimedia*,Pages 275-278,NY, USA,2003.
- [27] FlorentMonay, Daniel Gatica-Perez,“PLSAbased Image AutoAnnotation: Constraining the Latent Space”, *Proceedings of the 12th annual ACM international conference on Multimedia*, Pages 348-351, NY, USA,2004.
- [28] JiaLi,“Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach”, *IEEE Transactions on Pattern Analysis and Machine Intelligence archive*, Volume 25 Issue 9, Page 1075-1088 , September 2003.
- [29] Yasuhide MORI, Hironobu TAKAHASHI and RyuichiOKA, “Image-to-word transformation based on dividing and vector quantizing images with words”, *Neural Networks in Boltzmann machine*,1999.
- [30] Putthividhya,D., Attias,H.T., Nagarajan,S.S.”, *Supervised topic model for automatic image annotation*,”*IEEE International Conference on Acoustics Speech and Signal Processing* , Pages 1894 – 1897, 2010.