ISSN No. 0976-5697

Volume 1, No. 4, Nov-Dec 2010

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

An Empirical Study on Performance Evaluation in Automatic Image Annotation and Retrieval

M. Hemalatha
Head of Department of Software Systems,
Karpagam University
Coimbatore
hema.bioinf@gmail.com

T.Sumathi*
PhDResearch Scholar
Karpagam University
Coimbatore
t_sumathi@yahoo.co.in

Abstract: Advances of information and communication technologies allow the creation of image archives extensively. As a result, the size of images database archives is increasing rapidly. So an efficient image annotation and retrieval system is highly desired. Automatically assigning keywords to images allows one to index, retrieve and understand large collections of data. Many techniques have been proposed for image annotation in the last decade that gives reasonable performance on standard dataset. However most of these works fail to compare their method with other methods that justify the need for more complex models. In this work, we compare the performance of various image annotation methods, and propose that new base line method is that which outperforms the current state of art methods on two standard and one large web data set.

Keywords*:* semantic web, sub space clustering, weighted feature selection, New base line algorithm,Multilabel boosting

I. INTRODUCTION

Given an input image, the goal of automatic image annotation is to assign a few relevant keywords to the image that reflects its visual content. Utilizing this image content to assign a richer and more relevant set of keywords allow us to exploit the fast indexing and retrieval architecture of these search engines for improved image search.

The Semantic Web is a idea of having data on the Web defined and linked in such a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications" [3]. Semantic means adding meaning of data to be discovered by computers. It is a vision of a new architecture for the World Wide Web, characterized by the association of machine-accessible formal semantics with more traditional Web content. The core idea is to create the Meta data expressing the data, which will enable computers to process the meaning of things.

The ultimate goal of the Semantic Web [21, 22] is to transform the Web into a medium through which data can be shared, understood and processed by automated tools.

Semantic Web techniques, which consist of applying knowledge representation techniques in a distributed environment (potentially on a web wide scale), have proven useful in providing richer descriptions of Web resources [4].

Image annotation is a difficult task for two main reasons: first, is the well known pixel- to- predicate which points to the fact that is hard to extract semantically meaning full entities using low level image features like color and texture . Second difficulty arise due to the lack of correspondence between the keywords and image regions, one has to access to the keywords assigned to the entire image and it is not known which regions of the image correspond to these keywords.

Image annotation has been a topic of on-going research for more than a decade and several interesting techniques have been proposed like automatic image annotation and retrieval using Sub space clustering algorithm, weighted feature selection, automatic image annotation and retrieval with multi label boosting, and new base line for image annotation. Most of these techniques define a parametric or non-parametric model to capture the relationship between image features and keywords. The goal of this work is to compare these methods and choose a method that outperforms more complex state-of-art image annotation methods.

II. RELATED WORK

Image annotation surveys have been reviewed by many researchers according to the demanding the needs for annotating images. Bridging the semantic gap for image retrieval is not easy to overcome. In order to overcome the well known problem in semantic gap, automatic image annotation is the solution. However, major difficulty is to make computers understand image content in terms of semantics or high-level concepts. In order to bridge the semantic gap between low-level content and high level concepts could be applied by implementing image analysis and statistical learning approaches or other techniques such as classification etc. Related work has shown the applicability of using machine learning techniques for automatic image annotation in a number of categories. A large number of techniques have been proposed and in that most of the techniques treat this annotation as the problem of translation from image instances to keywords.

Comparison of Related Work

A. Subspace Clustering Algorithm

Image data usually have a large number of dimensions. Traditional clustering algorithms allocate equal weights to these dimensions. But in this we determine relevant features based on histogram analysis and assign greater weight to relevant features based on histogram analysis and assign greater weight to relevant visual tokens with keywords based on the clustering results of clustering algorithm and k-means algorithm.

Content-based image retrieval computes relevance based on visual similarity of low level image features such as color texture, shapes, spatial layout etc. However there is a gap between low-level visual features and semantic meanings. The so-called semantic gap is the major problem that needs to be solved for most CBIR approaches.

In automatic image annotation and retrieval the following steps are employed:

[a] Segment images into visual segments, Extract a quantify features for segments

[b] Cluster segment using clustering algorithm to construct blob tokens

[c] Analyze the correlation between keywords and blob tokens to discover hidden semantics.

Initially, each image will be represented by a set of keywords and visual tokens. A visual token means a segmented region or object and it will be described by a set of low level features like color, texture, and shape. And it is possible that the same visual token can be shared by more than image. If some visual tokens are the same, they will belong to the same cluster. In this approach the images are segmented into a number of visual tokens using normalized cuts. Each visual token will be described by colors, textures, shapes etc., and then we apply clustering algorithms to group similar visual tokens into a blob token. Thus we create a fixed set of blob tokens. This clustering for blob token generation involves first, we cluster visual tokens using K-means assuming equal weights. Next we distribute visual tokens into clusters and update centroids. Third, for each cluster we identify the most important features and discard irrelevant features. Then we update the weights of features in each cluster adaptively using subspace clustering algorithm this step is important because the value of relevant feature in the same cluster should be similar. To determine the link between keywords and blob tokens we construct probability tables. Then the methods like unweighted data matrix, weighted data matrix ,Singular Value decomposition EM algorithm are applied to build the connection between the keywords and blob tokens. Thus the annotation is generated using keywords assigned to all objects in the image.

To evaluate this method they have retrieved images from the test data set using 20 frequent keywords from the vocabulary. The images will be retrieved if the automatically established annotation contains the query keyword. The results are evaluated as follows:

Precision P= **Num_** correct/**Num_** retrieved and Recall r= **Num** _correct/ **Num_** exist and

E measure E (p, r) = 1-(2/ (1/p) + (1/r)

Where Num _ correct means the number of retrieved images which contain the query keyword in its annotation, Num_ retrieval is the number of retrieved images and Num_ exist is the total number of images in the test set containing the query keyword in annotation For evaluation 50 manually labeled image objects we added to test correspondence. If the word predicted by the blob is contained in manually generated keywords of this object then we can say that the blob predicts the word correctly in the right place. From the experimental results we have found out that the precision recall of some query words are zero. So from this we can say that PTS are better than PTK for most query word. The common E measure is calculated based on average precision and recall and E measure of PTS is lower than PTK.

The Problem faced is the similarity of visual tokens will be clustered together and a finite set of tokens are generated. The premise is that if some visual tokens are the same they will belong to same cluster.

B. Weighted feature selection:

This mechanism of annotating image is as follows: first, we calculate the clustered visual tokens using k-means assuming equal weight for all the features. Next we distribute visual tokens into clusters and update centroids. Third for each clusters we identify the most important features and discard the irrelevant feature. At last, the same process will be repeated until the algorithm converges. Thus in step-3 we apply weighted feature selection to determine the relevance of a feature. We represent m features in jth clusters as <fj1, fj2…fjm> and weight of these features as <wj1,wj2…wjm> . For lth feature in jth cluster, fjl, we assume that the denser the distribution of fjl, the more possible that the lth feature is the dominant feature for jth cluster.

$$Density_l = 1 - \frac{Area_{hist(l)}}{Max(Y_{li}) \times \Re_{lX}} .$$

The larger density is, the denser the value distribution for lth feature[22]. Now the new weights are calculated using

$$w_{jl} = \frac{Density_{jl}}{\sum_{l=1}^{m} Density_{jl}}$$

The area of the histogram is :

$$Area_{hist\ (l)} = \sum_{i=1}^{X} Y_{li} \times I_{li}$$

Then we apply clustering algorithm to group similar visual tokens into a blob token. To determine a link between keyword and blob token, we construct a probability table. To annotate the image automatically, we calculate the distance between the given image object and all centeroids of blob tokens and represent this image object with the keyword of the closest blob token. The annotation is generated using keywords assigned to all objects in the image.

In this method the evaluation id done with a total of 42,379 image objects from 4500 training images into 500 blobs using k-means and subspace clustering algorithm, then they have applied 10 different methods to calculate probability tables based on two clustering results. Finally they have 20 different probability tables, which correspond to 10 different methods. In this the methods like correlation method (CRM), Cosine Method (CSM), Singular value Decomposition (SVD), Expected Maximization (EM) are used [22]. The probability tables based on k-means algorithm are denoted as PTK the Probability tables based on subspace weighted feature selection algorithm are denoted as PTS.

While comparing performance of PTS with PTK in terms of average precision p, recall r, and common E measure E.

Table: I

Method	PTK			PTS		
	Avg Prec	Avg Rec	Avg E	Avg Prec	Avg Rec	Avg E
SVD + CRM	0.2141	0.31566	0.744	0.2953	0.4147	0.655
SVD+ CSM	0.12234	0.3877	0.81	0.260	0.525	0.6520

This table shows that the average precisions (column 5) and recall column 6) of PTS are better than PTK column (2 and 3). And E measure of PTS (column 7) is lower than PTK (column 4) so the order of keywords in automatic annotation is the same as the decreasing order of size of corresponding segmented image objects. The Problem faced is frequent keywords are associated with too many different image segments but in frequent keywords have little chance. That is, it never considered the relationship between blob tokens only the relationship between the keywords and blob tokens are considered.

Some subspace algorithms follow top-down strategy [Aggarwal99]. Basically, the top down subspace clustering approaches find initial approximation of the clusters at the beginning. Full feature space is considered and each dimension is weighted equally. Each dimension is assigned weight for each cluster based on clustering results. The updated weights are applied in the following iteration to regenerate the clusters. To the best of our knowledge, this is the first attempt to apply a weighted feature selection algorithm in automatic image annotation which is complementary to subspace clustering at some extent.

C. Multi-label Boosting

The machine learning systems for object recognition is limited by the requirements of single labeled images for training, which are difficult to create or obtain in quantity. So, therefore multi-label data provides a ready means to crate objects recognition systems which are able to deal with large number of classes. The object recognition system named ML- Boost which learns from multi label data through boosting and improves on state of the art multi-label annotation and labeling systems.

ML Boost is able to learn enough from 1500 or so annotated images[3]. It achieves this by learning the correlation between image segments and the accompanying text in a set of training images. Having learnt this, when given any new image it is able to translate it into words, giving both a labeling for the segments and an annotation for the image. The algorithms and techniques which pertain to this kind of learning come largely from the machine translation community, and have been adapted for use in computer vision by Barnard et. al.

In Barnard et al.'s method [3] the image is segmented using normalized cuts, then a feature vector is extracted for each segment containing color, texture, and other cues. To link blobs with words, it is assumed that there are hidden factors which are responsible for generating both the words and blobs associated with that factor. By generating both words and blobs, the concept can then be used to link the two, learning hoe they relate. The Joint probability of a particular blob b and a word w is modeled as P (w, b) =Sum of (P (w/c) P (b/c) P(c))
Where c indexes over the concepts and P© is the concept prior, P (w/c) is a frequency table and P (b/c) is a normal distribution over features [3].

This system provides a hypothesis to the booster which calculates a vocabulary distribution for a document using a novel weighted voting scheme. P(x/b) = Sum of (P (w/c) P(c/b)).

To evaluate this model is evaluated on two tasks annotation and labeling. Each was training on roughly 90% data, or 1667 images and a test set of the remaining was 214 images was used for evaluation of the model's generalization ability[3].

Table: II

	P%	R%	N+	rP%	rP+%
ML Boost	12	20	108	12	24

Annotation provides a straightforward means to determine the effectiveness of the models

D. A New Base Line Method:

Baseline method for image annotation is built on the hypothesis that images similar in appearance are likely to share keywords. In this image annotation is a process of transferring keywords from nearest neighbors. The neighborhood structure is constructed using image features, resulting in a rudimentary baseline models. In this method, given a test image we find its nearest neighbor from the training set and assign all the keywords of the nearest image to the input test image. In this scheme we use K- nearest neighbors to assign the keywords instead of relying on just the nearest one. In multiple neighbor case, we can easily assign the appropriate keywords to the input image using a simple greedy approach. The base line annotation method comprises of a composite image distance measure (JEC or Lasso) for nearest neighbor ranking, combined with label transfer algorithm [1]. Color and texture are recognized as two of the most important low level cues for image representation. Here the most common color descriptors are based on coarse histograms. These color features are frequently utilized within image matching and indexing schemes. Texture is another visual feature and is most frequently captured with wavelet features. In particular Gabor and Haar wavelets have been used and has been quite effective in creating sparse discriminative image features.

The combining distance is done through Joint Equal Contribution (JEC) which combines distances from different descriptors and allow each individual distance to contribute equally to the total combined cost or distance. Image similarity is done by L1-Penalized Logistic Regression [1]. In these images pairs that had at least four common keywords are treated as positive samples for training and those with no common keywords were used as negative samples. Label transfer method is used to transfer n keywords to query image from the query's K nearest neighbors in the training dataset.

III. RESULTS AND DISCUSSION

Performance Evaluation of Automatic Image Annotation and retrieval

The performance and behavior of the baseline and other methods for image annotation on three collections of images have been evaluated focusing on three different settings 1). Performance of individual distance measure 2). Performance of the relevant weighted distance model (Lasso) 3), performance of the Joint Equal contribution (JEC). This performance of all models was evaluated using five measures.

Table: III

Methods	P%	R%	N+	rP%	rP+%
Baseline-RGB	20	23	110	24	49
Baseline-HSV	18	21	110	23	45
Baseline-Haar	6	8	53	12	33
Baseline-Lasso	24	29	127	30	51
Baseline -JEC	27	32	139	33	52
JEC+Lasso	51	61	266	63	103
SVD+CRM	12	26	53	26	38
MBRM	15	20	100	15	20
Multi-label boosting	12	20	108	12	24
Other methods-CRM	16	19	107	-	-

Precision and recall [20]-[21], which are the most popular metrics for comparing CBIR, are also widely used for evaluating the effectiveness of automatic image annotation approaches.

Precision is defined as the ratio of the number of words that correctly retrieved to the total number of words retrieved in every image search. While recall is the ratio of the number of words that retrieved correctly to the number of words. Mean Precision P%, Recall rates obtained by different models (R %), Number of total keywords recalled (N+). In addition we report two retrieval performance measures: rP% denotes the mean retrieval precision for all keywords and rP+% denote the mean retrieval precision for recalled Keywords only.

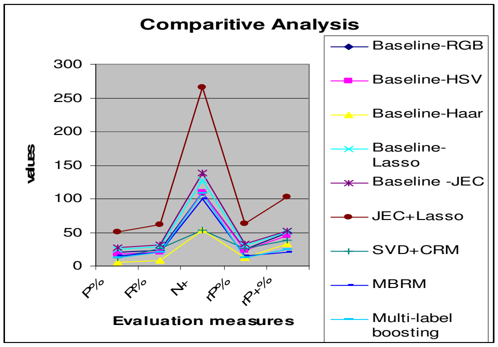

Figure: 1

This result of the experiments clearly states that comparing these base line methods with various image representations like CRM and MBRM, the distance measures induces all individual features. And the results of the baseline methods shows a wide spread in performance scores, ranging from high-scoring HSV and RGB color measures to the potentially less effective quantized Gabor Phase. The combination of individual distances (lasso and JEC) perform significantly better than most other published methods In particular, JEC , which emphasizes equal contribution of all the features distances, shows domination in all five performance measures.

In summary, this baseline annotation methods comprised of a composite image distance measure for nearest ranking, combined with our label transfer algorithm achieves reasonable result for image annotation using such simplistic methods

IV. CONCLUSION

From all the above discussed methods it is clear that new base line approach for image annotation bridges the gap between the pixel representation of images and the semantic meanings. It is clear that a simple combination of basic distance measures defined over commonly used image features can effectively serve as a baseline method to provide a solid test-bed for developing future annotation methods. Thus it is clear new base line method is that which outperforms the current state of art methods. This paper gives a study of image retrieval work towards narrowing down the 'semantic gap'. Recent works are mostly lack of semantic features extraction and user behavior consideration. Therefore, there is a need of image retrieval system that is capable to interpret the user query and automatically extract the semantic feature that can make the retrieval more efficient and accurate.

V. REFERENCES

[1] A New Baseline for Image Annotation Ameesh akadia, Vladimir Pavlovic, and Sanjiv Kumar, Google Research, New York, NY, Rutgers University, Piscata way, NJ makadia@google.com,vladimir@cs.ru tgers. edu,sanjivk@google.com

[2] Review on Statistical Approaches for Automatic Image Annotation Syaifulnizam Abd Manaf#1, Md Jan Nordin*2 #Malaysian Institute of Information Techno logy, University Kuala Lumpur2009 International Conference on Electrical Engineering and Informatics 5-7 August 2009, Selangor, Malaysia.

[3] F. Kang, R. Jin, and R. Sukthankar, "Correlated label propagation with application to multi-label learning," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1719–1726.

[4] M. Wang, X. Zhou, and T. S. Chua, "Automatic image annotation via local multi-label classification," in CIVR '08: Proceedings of the 2008 International conference on Content-based image and video retrieval. ACM, 2008, pp. 17–26.

[5] G. Tsoumakas and I. Katakis, "Multi-label classifica tion: An overview," International Journal of Data Ware housing and Mining, vol. 3, no. 3, pp. 1–13, 2007.

[6] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang, "Correlative multi-label video annotation," in MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia. ACM, 2007, pp. 17–26.

[7] C. G. Snoek, M. Worring, J. C. van Gemert, J. Geusebroek, and A. W. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," in Proceedings of ACM Multimedia, Santa Barbara, USA, October 2006, pp. 421–430.

[8] [Agrawal98] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in Proceedings of the 1998 ACM SIGMOD international conference on Management of data, ACM Press, 1998, pp. 94-105.

[9] [Barnard03] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, M. Jordan, "Matching words and pictures," Journal of Machine Learning Research, Vol.3, pp.1107-1135, 2003.

[10] [Khan02] L. Khan, and L. Wang**,** "Automatic Ontology Derivation Using Clustering for Image Classification," in Proc. of Eighth International Workshop on Multimedia Information Systems, Tempe, Arizona, October 2002, pp. 56- 65.

[11] [Li03] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, pp.1075-1088, 2003.

[12] [MacQueen67] Mac Queen, J., "Some methods for classification and analysis of multivariate observations," in Processing of fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 287-297, 1967.

[13] Yang, C., Dong, M., Hua, J.: Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2006)

[14] Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, and N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007)

[15] Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.A.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: European Conference on Computer Vision, pp. 97–112 (2002)

[16] Blei, D.M., Jordan, M.I.: Modeling annotated data. In: Proc. ACM SIGIR, pp. 127–134(2003)

[17] Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using crossmedia relevance models. In: ACM SIGIR Conf. Research and Development in Informaion Retrieval, New York, NY, USA, pp. 119–126 (2003)

[18] Wang, L., Liu, L., Khan, L.: Automatic image annotation and retrieval using subspace clustering algorithm. In: ACM Int'l Workshop Multimedia Data bases (2004)

[19] Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: Advances in Neural Information Processing Systems, vol. 16 (2004)

[20] H. Muller, W. Muller, D. M. Square, S. M. Maillet and T. Pun, "Performance Evaluation in Content Based Image Retrieval: Overview and Proposals", Pattern Recognition Letters, vol. 22, Apr. 2001.

[21] J. Vogel and B. Schiele, "Performance Evaluation and Optimization for Content Based Image Retrieval", Pattern Recognition Letters, vol. 22, Apr. 2001.

[22] Leiwang leiwang@utdallas.edu Latifurkhan khan@utdallas.edu,Automatic Image Annotation and Retrieval Using Weighted Feature Selection, Department of Computer Science, University of Texas at Dallas, Richardson, Texas 75083.