02 Days International Conference on Recent Innovations in Engineering, Science, Humanities and Management
Gayatri Vidya Parishad College for Degree & P.G Courses, Rushikonda, Visakhapatnam (India) (ICRIESHM)
11th-12th August 2016, www.conferenceworld.in ISBN: 978-93-86171-02-3

IMAGE TAG RANKING USING NEAREST NEIGHBOUR SEARCH

Angaru Aruna[1], Sreerama Murthy Velaga[2],

[1]M. Tech Student, [2]Asso. Professor, Dept. of CSE, GMR Institute of Technology, Rajam, India

ABSTRACT

Now-a-days Social Media is becoming very popular with digital images. There are several CBIR techniques available and specifically used to retrieve the images based on the visual content. CBIR became ineffective for increasing image data .In such cases, tag based image retrieval came into existence which is more effective than CBIR in identifying a relevant image based on image tag. Since it consuming more time to tag images manually Automatic Image Annotation(AIA) is developed. Most of the image annotation techniques are based on multi-label classification problem which requires clear annotations and also take lot of time for training. This limitation can be shown by developing a novel approach that merges the strength of tag ranking with the power of matrix recovery. Tag ranking is done in descending order of their relevance to the given image, which significantly solves the problem to some extent. Experimental results show the effectiveness of our proposed method for tag ranking. This tag ranking also reduces the training time significantly for image search.

Keywords: Image annotation, tag ranking, nearest neighbour, matrix recovery, trace norm.

I. INTRODUCTION

Now-a-days social networking sites are very popular with digital images. The important research topic is to retrieve images accurately from enormous collections of digital photos.

Content Based Image retrieval (CBIR) is the process of identifying and extracting images from the huge set of image database automatically extract features based on the visual content such as color, shape, edge and texture on the basis of user's query [1]. The main objective of CBIR is high retrieval accuracy with the less retrieval time. The difference between the retrieved image and query image is the semantic gap in CBIR system. CBIR technique is said to be effective and efficient only if the semantic gap is minimum [2]. To overcome this limitation of CBIR, many algorithms are developed for tag based image retrieval (TBIR) which shows images by manually attached keywords or tags. Since it consuming more time to label images manually Automatic ImageAnnotation (AIA) came into existence.

The main objective of image annotation is to automatically annotate an image with relevant keywords/tags which represent its visual content. Most of the studies registers image annotation as a multi-label classification problem. In our proposed framework, we focus on the tag ranking approach for automatic image annotation [3]–[5]. Instead of deciding each tag, if should be attached to a given image, tag ranking is done in descending order relevant to the given image, significantly solves the problem to some extent. In order to train a relevant model for tag prediction, it requires a large number of training images with clear and correct annotations is the main

 ISBN: 978-93-86171-02-3

objective of this approach. Instead of making binary decision for each tag, the tag ranking approach significantly reduces the complexity of a problem, and also leads to a better performance than the traditional classification based approach for image annotation[6].

Baby, white, ears, eyes, nose, **child**, boy, bed, **kid,** girl, face, pink.

Fig 1: sample image of corel5k dataset with tags. Highlighted ones are relevant tags of a image.

In real world applications they are different algorithms developed for tag ranking, which tend to perform very poor when compared to the number of tags with limited number of training images is limited [7]. We address this limitation by casting tag ranking into a matrix recovery problem. Major idea is to aggregate the prediction models for different tags into a matrix. Instead of learning each prediction model alone, we propose to learn all the prediction models simultaneously by considering the theory of matrix recovery, where a trace norm regularization is introduced to capture the dependency among different tags to control the model complexity.

Recently, Makadia et al. (2008); Guillaumin et al. (2009) proposed different new algorithms for automatic image annotation based on nearest neighbour methods[8]. Guillaumin et al. (2009) carefully learn embeddings into metric spaces which join a different image set descriptors and assign specific tag weights to overcome label sparsity [9]. As a result the algorithm improves both precision and recall over state-of-art methods.

In our work we show, both empirically and theoretically, that with the introduction of trace norm regularization, a reliable prediction model can be learned for tag ranking even when the tag space is more with limited number of training images. Although the trace norm regularization has been studied widely for classification [10], [11] purpose, this will be the first study that utilizes trace norm regularization for tag ranking.

II. RELATED WORK

2.1 Automatic Image Annotation

Automatic image annotation aims to assign a set of relevant keywords or tags to an image, which can show its visual content of a digital image. Content based ranking of images is hard for text documents because they do not have any text in their content part. It plays a major role in combining the semantic gap between low-level features and high-level semantic content of images. Image search can be done by using annotations and semantic tags that are present in image of a image dataset. However, these tags which are entered by the users

manually consumes more time for large image dataset. Thus, AIA has been a most important research topic from past decades. AIA methods require a set of training images, from which annotations for the images are determined earlier. The process of AIA is as follows

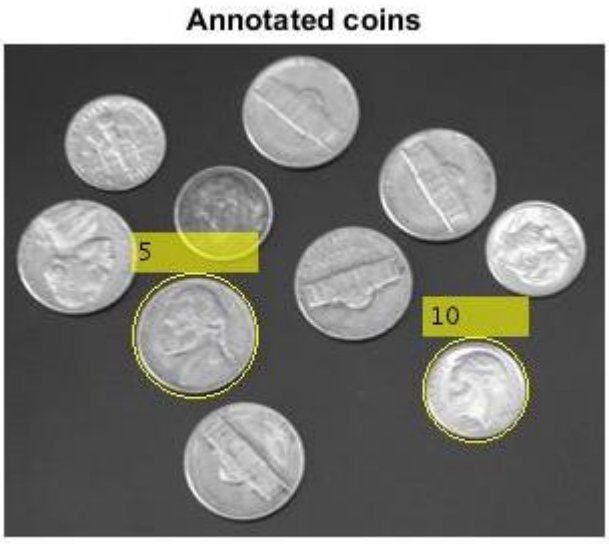


Fig 2: Annotated Image

The training image dataset has been previously loaded into the computer system in order to process the image which is given as input. This technique uses both local and global features for calculating the existence of the training dataset of a given image. First splits the given image into various combination of image features based on scaling by GIST, SIFT and different colour histograms like RGB, HSV. Then the images are compared with required image dataset, then different features in images remain unnoticed where the similar tags retrieved back and annotated automatically.

Image Annotation According to [9], (AIA)automatic image annotation method categorized into three groups: (i) generative models [12, 14], which are designed to model the joint distribution between tags and visual features, (ii) discriminative models [13, 15] that view image annotation as a classification problems where each keyword is treated as an independent class, and (iii) search based approaches[8,16]. Search based approaches are more popular when compared to generative and discriminative models. Here we discuss briefly about search-based techniques which are developed for image annotation.

The search based approaches are based on the single idea that visually similar images are share similar keywords/tags [10].Let us assume a example with a test image say I, which primary finds out a set of training images that are visually common to nI, and then it assigns the tags that are most popular among the similar images. A divide-and-conquer framework is proposed in [45] which identifies the salient terms from textual descriptions of visual neighbours searched from web images. In the Joint Equal Contribution (JEC) model proposed in [4], multiple distance functions are computed with each based on a different set of visual features, and the nearest neighbours are determined by the average distance functions. TagProp [7] predicts keywords by taking a weighted combination of tags assigned to nearest neighbor images. More recently, the sparse coding scheme and its variations are employed in [5], [9], [14] to facilitate image label propagation. Similar to the classification method, the search based approaches often fail when the number of training examples is limited.

ISBN: 978-93-86171-02-3

2.2 Tag Ranking Based on Neighbor Search Mechanism

In this section, we propose ranking oriented nearest neighbour mechanism which optimize the ordering of all labelled images for a given image. The top-K ranked results are then selected as the K nearest neighbours. To explain it clearly, we first give some notations. Let χ denote an image collection, and all keywords appearing in the collection are $T=\{t_1 t_{2,\ldots,} t_c\}$where c is the total number of unique keywords. In the image annotation task, we are given a set of n labelled training images, $S=\{x_i \in X | i=1,\ldots\ldots, n\}$,in which each labelled image xi is associated with a c-dimensional binary label vector $y_i \in \{0, 1\}^c$, whose j^{th} element $y_i^{(j)}$ indicates the presence of keyword t_j in x_i, that is, $y_i^{(j)}=1$ if xi is labelled by t_j and $y_i^{(j)}=0$ otherwise. Given a new image $x_{new} \in X$, our goal is to learn a ranking function $H : X \times S \rightarrow R$from the training data, such that $H(x_{new}, x_i)$ can represent the relevance of the labeled image x_i with respect to x_{new}, and x_i is ranked before x_j if $H(x_{new}, x_i) > H(x_{new}, x_j)$[17].

III. EXPERIMENTAL SETUP

Below we can find the experimental setup for image annotation. As described earlier our approach is based on nearest-neighbour–based scheme.

3.1 Dataset

Corel 5k: The Corel 5K data set was first utilized by Duygulu, Barnard, De Freitas, and Forsyth (2002)[8]. Since from that year it created a benchmark for image annotation and that has been widely used in many experiments. As a result, we can compare it directly with the experimental results of our approach against the published results of previous studies. This dataset contains about 5,000 where is image is manually annotated with 1 to 5 keywords. The annotation word vocabulary contains 260 keywords. A fixed set of 499 images are for testing purpose and rest of all images are used for training. Table 1 summarizes the basic information about our data sets.

Table 1: Summary of The Experimental Dataset and Feature Extraction

Data set and Feature Extraction	Description
Corel 5k	5,000 images and 260 keywords
Color histogram	In RGB, LAB and HSV spaces with two spatial layouts; Using L1-distance.
GIST	Using L2-distance.
SIFT	Dense and interest point sampling with two spatial layouts; using $\chi 2$ statistic.
HUE	Dense and interest point sampling with two spatial layouts; using $\chi 2$ statistic.

3.2 Feature Extraction

The representation includes two types of global image features: Sift and color histograms. The color histograms were extracted in three different color spaces: RGB, LAB, and HSV, which are the most commonly used color spaces in computer vision, and have been applied in many proposed image annotation studies (Makadia et al., 2008; Zhang et al., 2012). We divided the color histograms into 16 bins in each color channel, yielding 163 = 4,096 dimensional histograms for each image. These features characterize the image content from a global view. Because local features can capture more semantic content of images than global features, the scale-invariant feature transform (SIFT) and a robust hue descriptor were also adopted as two local features. Both of them were extracted on multiscale grids and around Harris–Laplacian interest points. Each local descriptor was quantized into visual words using K-means clustering.

Furthermore, to capture the spatial layout of images, all histogram features, except Gist, were computed over three horizontal regions of an image, and the resulting three histograms were concatenated to form a new feature. It should be mentioned that the new color histograms were requantized to 12 bins in each color channel, yielding 3 * 123 = 5,134 dimensional histograms for each image. The choice of 12 bins is a compromise between limiting the feature size and avoiding excessive loss of color distribution information. Finally, a total of 15 visual features were extracted from each image.

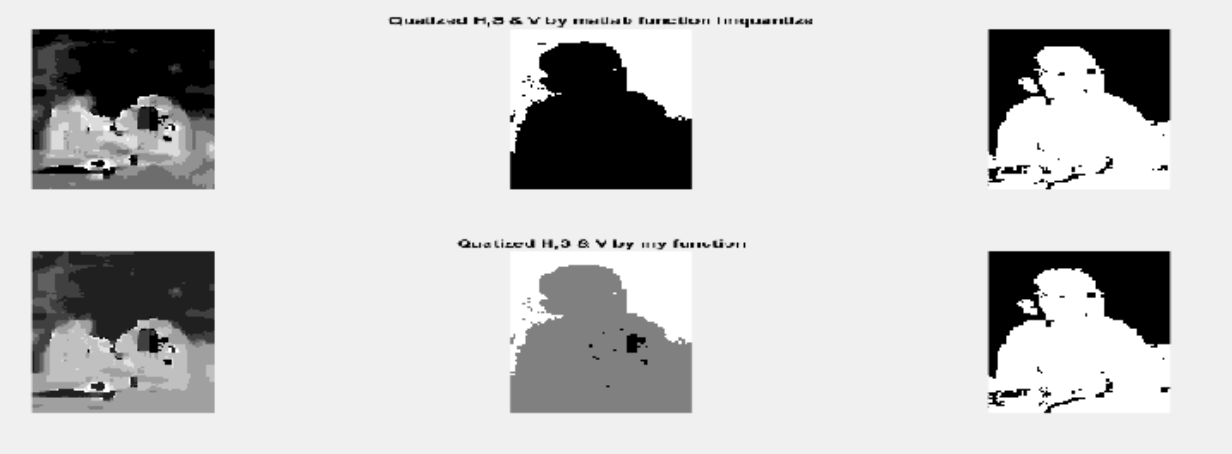


Fig 3: Example Image of A Corel5k Image Dataset Which Extracting Hsv Features

In our approach, the distance in each feature space needs to be defined respectively. Specifically, we used L2-distance as the base metric for Gist, L1-distance for color histograms, and $\chi 2$ statistic for the others.

3.3 Evaluation Criteria

Firstly, to evaluate the performance of automatic image annotation, we adopt the Average Precision (AP@K) and Average Recall (AR@K) as the evaluation metrics, which are defined as where K is the number of truncated tags, nt is the number of test images, $N_c(i)$ is the number of correctly annotated tags for the ith test image, $N_g(i)$ is the number of tags assigned to the ith image. Both average precision and recall compares the automatically annotated image tags. To compare the results of different annotation methods, we adopted the

standard performance measures widely used in previous work, where the quality of the predicted annotations is evaluated by retrieving test images using the keywords or concepts in annotation vocabulary.

$$AP@K = \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{N_c(i)}{K}$$

$$AR@K = \frac{1}{n_t} \sum_{i=1}^{n_t} \frac{N_c(i)}{N_g(i)}$$

Fig4: Feature Extraction Images of Corel5k Dataset.

On the Corel 5K data set, for ease of comparison with published results, we annotated each test image with the five most relevant keywords as done by Makadia et al. (2008)[8], Guillaumin et al. (2009), and Zhang et al. (2012)[9].

IV. EXPERIMENT & RESULTS

In this experiment, we will first evaluate the image annotation performance of the proposed image tagging method with limited training images using nearest neighbour search mechanism. By the end, we randomly sample only 10% of images used for training and the remaining used for 90% for testing. Each experiment is repeated 10 times, differentiating testing and training data. Result of our approach is based on the average Precession and Recall. Below mentioned are state-of-the art approaches .

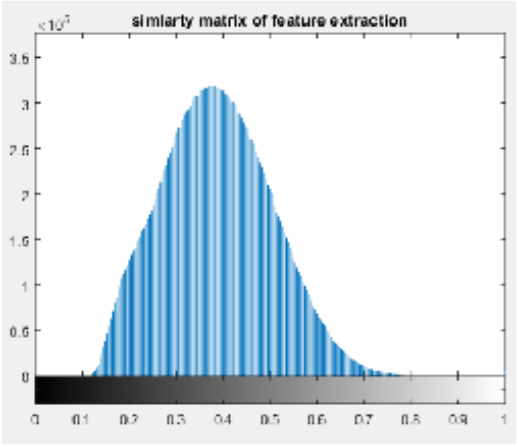
Joint equal contribution method (JEC)[8]: JEC finds appropriate tags to our test image based on a K nearest neighbour classifier that used a combined distance measure which is derived from multiple sets of visual features.

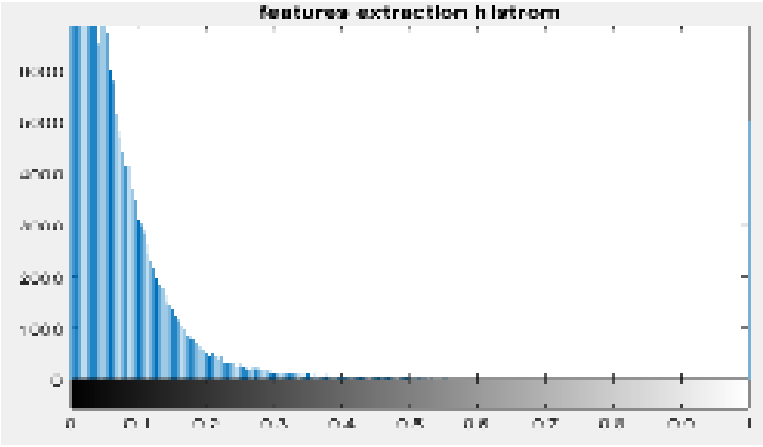
Tag Propagation method (TagProp) [9]. It propagates the tag information from the labeled images to the unlabeled ones with a weighted nearest neighbour graph, where RBF kernel function is used for computing weights between images.

Multi-Class SVM method (SVM) [18]. It simply implements One-versus-All (OVA) SVM classifier for each tag, and ranks the tags based on the output probability values.

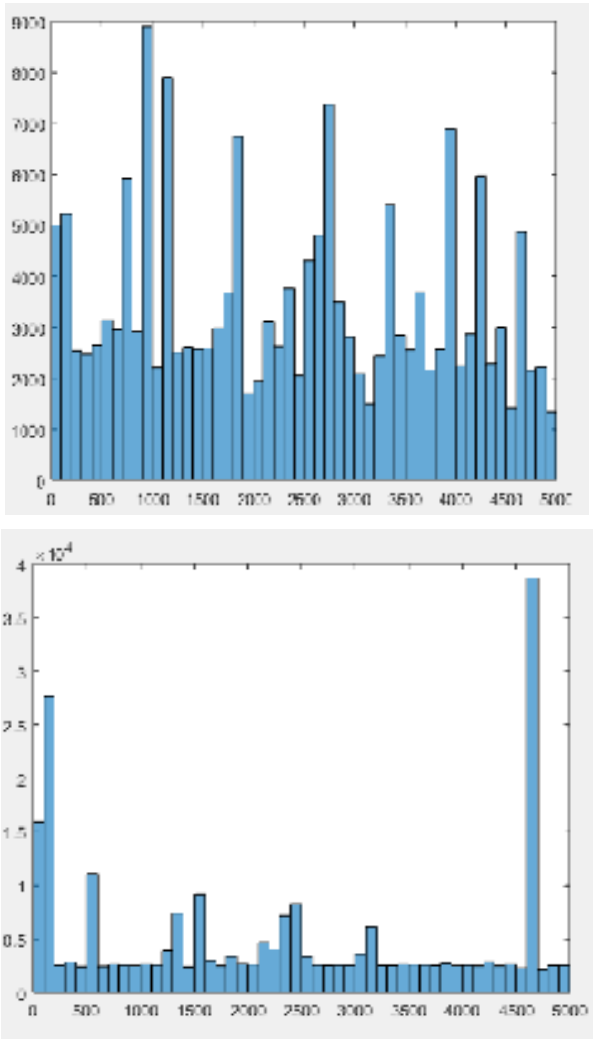
Efficient Multi-Label Ranking method (MLR) [19]. This approach explains the group lasso technique in multi-label ranking which efficiently handles the missing class labels.

ISBN: 978-93-86171-02-3





ISBN: 978-93-86171-02-3



. Fig 5: relevant and irrelevant tags histograms of corel5k dataset.

First, we show the comparison of average precision average recall for the first 5 returned tags of Corel5K dataset. when average precision declines we will observe increasing number of returned tags there by average recall improves. This is also called trade-off phenomenon of precision-recall which is well known in information retrieval [3].

Second, we observe that our method significantly outperforms two nearest-neighbour based methods (JEC and TagProp) on the given datasets since the performance of nearest-neighbour based methods largely depend on the number of training samples.

Table 2: Result of our proposed frame work

	COREL 5K
Classification	0.426
Total Knw	110
Precession	0.187278
Recall	0.221922

 ISBN: 978-93-86171-02-3

Special design of our proposed framework combines the ranking approach with trace norm regularization which avoids binary classification decision, and it is the trace norm regularization that makes our approach robust to the limited number of training examples using nearest neighbour mechanism.

V. CONCLUSION

In this paper, we proposed a novel approach that merges the strength of tag ranking with the power of matrix recovery to improve the training time. Tag ranking is done in descending order of their relevance to the given image, which significantly solves the problem to some extent. Experimental results are provided to prove the efficacy of the proposed methodology.

REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, 2008.

[2] J. Wu, H. Shen, Y. Li, Z. Xiao, M. Lu, and C. Wang, "Learning a hybrid similarity measure for image retrieval," Pattern Recognition, vol. 46, no. 11, pp. 2927–2939, 2013.

[3] X. Li, C. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," IEEE Trans. on Multimedia, vol. 20, no. 11, pp. 1254– 1259, 2009. [19] D. Liu, X. Hua, L. Yang, M. Wang, and H. Zhang, "Tag ranking," in

[4] Z . Wang, J. Feng, C. Zhang, and S. Yan, "Learning to rank tags," in ACM Int. Conf. on CIVR, 2010, pp. 42–49. [21] J. Zhuang and S. Hoi, "A two-view learning approach for image tag ranking," in ACM Int. Conf. on WSDM, 2011, pp. 625–634.

[5] S. Bucak, P. Mallapragada, R. Jin, and A. Jain, "Efficient multi-label ranking for multi-class learning: application to object recognition," in IEEE Int. Conf. on Computer Vision, 2009, pp. 2098–2105.

[6] Z. Li, J. Liu, C. Xu, and H. Lu, "Mlrank: multi-correlation learning to rank for image annotation," Pattern Recognition, vol. 46, no. 10, pp. 2700–2710, 2013.

[7] T. Lan and G. Mori, "A max-margin riffled independence model for image tag ranking," in IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2013, pp. 3103–3110.

[8] A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for image annotation," International Journal of Computer Vision, vol. 90, no. 1, pp. 88–105, 2010.

[9] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation," in IEEE Int. Conf. on Computer Vision, 2009, pp. 309– 316.

[10] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja, "Low-rank sparse coding for image classification," in IEEE Int Conf. on Computer Vision, 2013, pp. 281–288.

[11] Y. Zhang, Z. Jiang, and L. Davis, "Learning structured low-rank representations for image classification," in IEEE Int Conf. on Computer Vision and Pattern Recognition, 2013, pp. 676–683.

[12] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. PAMI, 29(3):394–410, 2007.

[13] J. Fan, Y. Gao, and H. Luo. Multi-level annotation of natural scenes using dominant image components and semantic concepts. In ACM Multimedia, 2004.

[14] S. L. Feng, R. Manmatha, and V. Lavrenko. Multiple bernoulli relevance models for image and video annotation. In CVPR, 2004.

[15] T. Mensink, J. J. Verbeek, and G. Csurka. Learning structured prediction models for interactive image labeling. In CVPR, 2011.

[16] X.-J. Wang, L. Zhang, F. Jing, and W.-Y. Ma. Anno search: Image auto-annotation by search. In CVPR, 2006.

[17] Chaoran Cui, Jun Ma, Tao Lian, and Zhumin Chen, Shuaiqiang Wang, "Improving Image Annotation via Ranking-Oriented Neighbor Search and Learning-Based Keyword Propagation," in Journal Of The Association For Information Science And Technology, 66(1):82–98, 2015.

[18] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. on Neural Networks, vol. 13, no. 2, pp. 415–425, 2002.

[19] S. Bucak, P. Mallapragada, R. Jin, and A. Jain, "Efficient multi-label ranking for multi-class learning: application to object recognition," in IEEE Int. Conf. on Computer Vision, 2009, pp. 2098–2105.