Predictive Tagging of Social Media Images using Unsupervised Learning

Nishchol Mishra
Asstt. Professor
School of IT
RGPV, Bhopal
India

Sanjay Silakari, PhD.
Professor, Deptt. Of CSE
UIT- RGPV
Bhopal
India

ABSTRACT
The popularity of online social media has provided a huge repository of multimedia contents. To effectively retrieve and store this multimedia content and to mine useful pattern from this data is a herculean task. This paper deals with the problems of social image tagging. Multimedia tagging i.e. assigning tags or some keywords to multimedia contents like images, audio, video etc. by users is reshaping the way the people generally search multimedia resources. This huge amount of data must be effectively mined and knowledge is discovered to find some useful patterns hidden in it. Some facts like Facebook which has more than one billion active users, and millions of photos are uploaded daily, YouTube has 490 million unique users who visit every month, People upload 3,000 images to Flickr (the photo sharing social media site) every minute, Flickr hosts over 5 billion images. Apart from their usage for general purpose search, they are also leading towards many diverse areas of research like land mark recognition, tag recommendation, tag relevancy, automatic image tagging or annotation. This paper addresses the problem of automatic image tagging or Predictive tagging of digital images in social network scenario. Predictive tagging aims to automatically predict tags and check the relevancy of tags associated with images. This can be accomplished by using unsupervised learning.

General Terms
Automatic image annotation, content based image retrieval, image annotation, image mining and multimedia data mining.

Keywords
Predictive tagging, tag relevancy, image retrieval, social media analytics

1. INTRODUCTION
The central idea of automated tagging techniques is to automatically learn semantic concept models from image samples, and to use these concept models to label new images. Once images are annotated with semantic labels, they can be retrieved with keywords. The objective is to provide tools for finding relevant visual information from the mass of stored visual data.

Generally, the tags provided by the users are ambiguous and personalized. This is due to fact that each user visualizes a particular image form its own perspective. The cultural background, geographical location, knowledge domain contributes a lot for perspective thinking of the user. So image tagging is usually ambiguous, personalize and subjective [12]. To measure the relevancy of tags, associated with the image is promising and challenging research direction. Hence, a major problem in social image tagging, analysis and retrieval is how to efficiently and accurately learn the relevance of a tag with respect to the visual context the user provided tag is describing.

Many methods to automatically predict or recommend tags or finding the tag relevancy with respect to the describing visual content of the image depends heavily on supervised machine learning methods. On the other hand, the vast coverage of social networking media and the amount of data that has to be learned or trained is so huge that supervised learning methods will not be appropriate. Therefore, in social media tagging environment which has very large and diverse visual content an unsupervised approach which is light weight in nature is required [12]. Beginning with the intuition that if different persons label visually look like similar images using the same tags than these tags reflects the main objective aspect of the visual content the image is describing [6]. This intuition will help finding the relevance of tag associated with the image with respect to the visual neighbors of that image, i.e. relevance of a tag with respect to an image might be approximated from tagging criteria of visual neighbor of that image. Keeping this intuition in mind, a framework is proposed for predictive image tagging and finding out the relevancy of tags associated with images. Predictive image tagging or annotation is the process of assigning tags to labeled image or unlabeled image. This proposed framework will also be helpful in finding the relevancy of tags associated with image.

As there is a semantic gap [5] between low levels content and higher level concepts. "The semantic gap is defined as the lack of coincidence between the information that one can extract form the visual data and the interpretation that the same data has for a user in a given situation" [4]. The overall goal, therefore, remains to bridge the semantic gap using the available visual features of images and relevant domain knowledge. The proposed framework, therefore, will propagate the common tags by using visual similarity, than each tag will receive credits by using voting or by receiving neighbor votes. For experimental purpose, Flickr image data set of twenty thousand images is used to demonstrate the work of tag prediction for labeled and unlabeled images, as well as, to find out tag relevancy of already labeled images.

The rest of the paper is organized as follows. Related work or review on literature has been done in section II. Predictive tagging framework and tag relevancy learning is discussed in section III. Experimental results are demonstrated in section IV. Paper is concluded in section V.

2. REVIEW OF RELATED WORK
Automated image tagging aims to create a computational model to assign words to an image in context to describe its contents. This technique provides an alternative to the cumbersome problem of manual tagging or annotating large

collection of images. In literature, it is generally referred as image auto captioning, automatic photo tagging, automatic image indexing, image auto annotation and multi label image classification. The tags or annotation are generally unstructured textual meta data simple keywords or combination of words that describes the visual content of an image.

Automated image tagging or annotation:-

The key idea of automatic image tagging techniques is to automatically learn semantic concept models from a large numbers of sample images, and use the concept models to label new images. After the images are annotated by using semantic labels, they can be retrieved by keywords which in turn are similar to text document retrieval. The main characteristic of automatic image tagging is that it provides keyword searching based on image contents and has an advantage of both the text annotation and content based image retrieval (CBIR). In CBIR images are automatically indexed and retrieved with low level content features likes color, texture and shape [15]. However, there is a significant gap between the low level content features and semantic concepts used by humans to understand images [5].

To improve image tagging, firstly, consider the *scenario of automated image tagging or annotation for unlabeled image*. In this scenario the widespread methods are discussed which try to predict relevant tag for unlabeled images. Categorization is done based on their model dependence into model based and model free approaches [6].

The *model free* approaches attempts to predict relevant tags for an image by utilizing images on the internet. The basic underlying assumption in these approaches is that there exists a huge labeled database of images such that finding a near visual duplicate for the unlabeled image is easy [6]. The next step is of automated image tagging and this done by simply propagating tags from the near duplicate image to the unlabeled image whose tags are to be estimated. The practical problem with this large database is that it is of unlimited scale with noisy annotations. Therefore, the neighbor search is conducted first to find alike visual neighbor. Next step is to select relevant tags out of the raw annotations of the neighbor. To select relevant tags from raw annotation disambiguation method are used.

In [1] the author proposed a search-based image annotation algorithm that is analogous to information retrieval. They mainly focus on annotation of large personal image collections. The proposed algorithm SBIA has four steps viz. first CBIR technique is used to retrieve a set of visually similar image; second, a text based keyword search technique is used to obtain a ranked list of candidate annotation for each retrieved image. Third, a fusion algorithm is used to combine the ranked lists into a final annotation list. Finally, the candidate annotations are re- ranked using random walk with restarts. In addition to this authors have also provided an annotation rejection scheme to point out the image that their annotation system can not handle efficiently.

The problem in selecting relevant tags out of the raw annotations of visually similar neighbor that number of occurrences of tag i.e. tag ranking is done based on their frequency in neighbor set as done by the authors of [2]. However, if some tag occurrence frequency is more in entire collection of documents then this may dominate the results. So, to overcome this problem the authors of [1] used the concepts of tf-idf (term frequency – inverse document frequency). Tf-idf is a numerical statistics which reflects how important a word is to a document in a collection. It is used as weighing factor in information retrieval. The tf-ids value increases proportionally to the number of times a word appears in the documents. The inverse document frequency is a measure of whether the term is common or rare across all documents. The idf value of a tag is inversely end logarithmically proportional to the occurrence frequency of the tag in the entire collection.

In [6][12][13], authors addressed the problem of Social image analysis and retrieval, which is important for helping people organize and access the increasing amount of user tagged multimedia. Since user tagging is known to be uncontrolled, ambiguous, and overly personalized, a fundamental problem is how to interpret the relevance of a user-contributed tag with respect to the visual content the tag is relating. Naturally, if different persons label visually similar images using the same tags, these tags are likely to reflect objective aspects of the visual content. Starting from this intuition, they propose in this paper a neighbor voting algorithm which accurately and efficiently learns tag relevance by accumulating votes from visual neighbors. Under a set of well-defined and realistic assumptions, they prove that proposed algorithm is a good tag relevance measurement for both image ranking and tag ranking. Three experiments on 3.5 million Flickr photos demonstrate the general applicability of their algorithm in both social image retrieval and image tag suggestion. The results suggest that the proposed algorithm is promising for real-world applications.

The second approach is a *model based* approaches which is generally performs in supervised learning framework that emphasizes on a mapping between high level semantic concepts and low level visual features. In [3], the authors proposed a new approach for modeling multi model data sets, focusing on the specific case of segmented images with associated text. The authors have developed numbers of models for the joint distribution of image regions and words, which explicitly learn the correspondence between regions and words. The results are shown using both an annotation proxy and manually labeled data. In [7], the authors had developed a new annotation method that achieves real time operation and better optimization properties preserving the advantages of the generative modeling approach. By advance statistical modeling and optimization techniques, they can educate computers about hundreds of semantic concepts using example pictures from each concept. They constructed a system called ALIPR which is fully automatic and capable for high speed annotation for online pictures. The results show that a single computer processor can suggest annotation terms in real time and with fair accuracy. In [8], the authors describe an inventive image annotation tools for classifying image regions in one of seven classes. The annotation is performed by a classification system based on a multiclass support vector machine. But these approaches lack in many ways. Currently, only limited number of visual concepts can be modeled effectively. Besides, these approaches are computationally expensive at the cost of manual labeling. Apart from above reason, the accelerated growth of new multimedia data makes the trained model to be outdated quickly.

Now taking into account the *second scenario* i.e. *improving the quality of image tagging for labeled images*. This can be done by removing noisy tags. In [9], authors propose a novel approach that strives to prune irrelevant keywords by the usage of WordNet. To identify irrelevant keywords, they look into various semantic similarity measures between keywords and finally combine outcomes of all the trial together to make final decision using Dempster-Shafer evidence combination. The final result demonstrates that by augmenting knowledge based with classical model annotation accuracy can be improved by removing inappropriate keywords.

Recommending new tags that are relevant to existing can also improve the image tagging. In [10], authors investigated how to assist users in tagging phase. First, they analyze how users tag photos and what kind of tags they provide. Second, they present various tag recommendation strategies to support to the user when annotating the photos. As the incredible amounts of photos are tagged by the users, they have derived relationship between tags, using global co-occurrence matrices. To be precise, by using each of the original tag as a seed, they find a list of candidate tags having the largest co-occurrence with the seed tag. These lists are merged into a single final list and the top ranked tags are recommended for the final tagging.

Resolving the tab ambiguity or by reducing it can also improve image tagging. In [11], an author describes a mean to determine the ambiguity of a set of (user-contributed) tags and suggests new tags that disambiguate the original tags. They introduced a probabilistic framework that allows finding two tags that appear in different context but are likely to co-occur with the original tag set. They have tested their work using geographical, temporal and semantic metadata and a user study. The proposed work is motivated by the findings of [6] [13]. Following the same intuition as discussed by Li et al. [6] in their paper, this work also follows the same and an effort is make to propose a framework for social image tagging.. This piece is of work is an attempt to get insight in the problem of social image tagging.

3. PROPOSED FRAMEWORK FOR PREDICTIVE TAGGING OF DIGITAL IMAGES

3.1 Tag relevancy and Predictive tagging

The scenario considered for this work is fundamental problem in social image analysis and retrieval is how to accurately and efficiently learn the significance of a tag with respect to the visual content the tag is describing. For image retrieval, images significant with respect to user queries should be ranked as high as possible. World Wide Web images and more specifically images those are stored on massive image databases in social networking sites like Facebook, Twitter, Flickr etc. Millions of images are shared on these platforms by many people around the globe. To effectively organise and retrieve these images is a herculean task. The final aim is to develop a framework that can automatically predict the tags for an untagged image that is stored in image database based on the semantics of the image and these tags can help to organise the images effectively for future retrieval. The figure 1 depicts the proposed framework.

The first objective of the work is to effectively retrieve the images based on the content not on captions or labels. This can be achieved by using the concepts of Image Retrieval (IR). As discussed above IR is interrelated to the concepts of image mining. Visually similar images are clustered by using low level features of images for reducing the search space. By using the concepts of IR images are mined related to the query image.

Once the similar images are retrieved secondly, most similar images are displayed first. Once the similar images are retrieved then look for the most relevant images. Here the considerations of user defined tags are taken. Next step is to determine how much the tags are relevant to the images and look for the possibility that if user has provided incorrect tags for that image. By checking the tag relevancy for image correct tags can be provided. The tag relevancy can be determined by using voting algorithm, such that the relevance of a tag given an image might be inferred from how visual neighbours of that image are tagged: the more frequent the tag occurs in the neighbour set, the more relevant it might be. A visual neighbour means the images that are retrieved by using the concept of IR for any particular image. Hence, a good tag relevance measurement should take into account the distribution of a tag in the neighbour set and in the entire collection, simultaneously.

In third phase, proposed work tries to predict relevant tags for unlabeled images. This could be achieved by using data-driven method based algorithm that is unsupervised and takes into the account multiple features. Therefore, in a social tagging environment with large and diverse visual content, a lightweight or unsupervised learning method which effectively and efficiently estimates tags of an image is required. This paper discusses the second and third phase of the proposed work.

To find out tag similarity or relevancy and to predict tag/s for unlabeled image firstly, find out the low level features of an image. Low level features like HSV Histogram, Color correlogram, Texture Features and SIFT (Scale Invariant Feature Transform) are taken into consideration for this work. SIFT is used for extracting distinctive invariant features from images that can be invariant to image scale and rotation. It is one of the most robust feature detection methods [14].

ProposedLow-level visual features capture the most important information encoded in the different color pixels composing the image. From this data information can be gathered about the image using various techniques. Low-level visual features that are computed from images are represented by vector V_L. This vector consists of following features:

HSV Histogram:
$V_{Hi} = \{V_{hi1}, V_{hi2}, \ldots\ldots\ldots, V_{him}\}$ for HSV Histogram descriptor, an m-dimensional vector.

Color Correlogram:
$V_{Cc} = \{V_{cc1}, V_{cc2}, \ldots\ldots\ldots, V_{ccn}\}$ for Color Correlogram descriptor, an n-dimensional vector.

Texture Feature [Tamura]:
$V_{Tt} = \{V_{tt1}, V_{tt2}, \ldots\ldots\ldots, V_{ttp}\}$ for Tamura Texture feature descriptor, n p-dimensional vector.

SIFT Feature:
$V_{Sf} = \{V_{sf1}, V_{sf2}, \ldots\ldots, V_{sfq}\}$ for Sift feature descriptor, an q-dimensional vector.

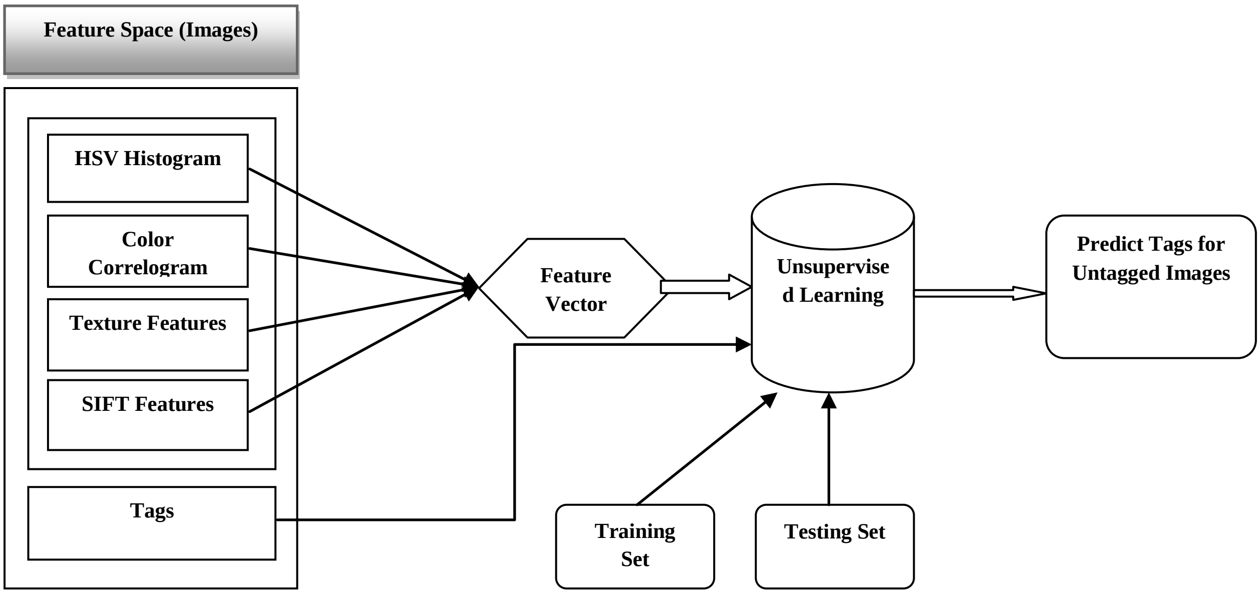


Fig. 1 Framework for predictive Tagging of Digital Images

Combination of all these features will help to find similar images in multimedia database. The similar images will be rerieved based on these features. A matching process between query image features and stored image features will generate most similar image. Matching is done by finding similarity score or by calculating distance. Eucledian distance is genrally calculated for matching but in this work Quadratic Form (QF) distance is used.

Quadratic form distance can lead to perceptually more desirable results than Euclidean distance and histogram intersection method as it considers the cross similarity between colors. Quadratic form distance has been used in many retrieval systems for color histogram-based image retrieval, it is given as:

$$\boldsymbol{D(I,J)} = \sqrt{(\boldsymbol{F_I} - \boldsymbol{F_J})^T \boldsymbol{A}(\boldsymbol{F_I} - \boldsymbol{F_J})}$$

Where *A*= [*aij*] is a similarity matrix, and *aij* denotes the similarity between bin *i* and *j*. F_I and F_J are vectors that list all the entries in *fi*(*I*) and *fi*(*J*).

These distances functions are used to search for the Nearest Neighbor of each vector. This will help to find the most similar images in the image database in respect to result generated for query image. The displayed images will be similar to the query image based on features.

Retrieval Metrics:

Basically two metrics are used for retrieval effectiveness they are *recall* and *precision*. Recall defines the relevant images in the image database that are retrieved in response to a image query. Precision is the proportion of the retrieved images that are relevant to the image query. More precisely, let *A* be the set of relevant items, let *B* the set of retrieved items and *a*,*b*,*c* and *d* are described in Figure 2

In the picture, ***a*** denotes for *'retrieved relevant'* images, ***b*** for *'retrieved irrelevant'* images, ***c*** for *'unretrieved relevant'* images and ***d*** for *'unretrieved irrelevant'* images. Then recall and precision are defined as the following conditional probabilities.

$$\boldsymbol{Recall = P\,(B \mid A)} = \frac{P\,(A \cap B)}{P\,(A)} = \frac{a}{a+c}$$

$$\boldsymbol{Precision = P\,(A \mid B)} = \frac{P\,(A \cap B)}{P\,(B)} = \frac{a}{a+b}$$

With these conditions, image retrieval is said to be more effective if precision values are higher at the same recall values.

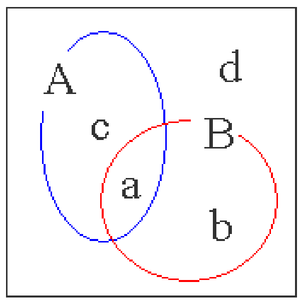


Fig. 2 Precision and recall retrieval effectiveness

3.2 Tag Similarity measurement using ranking/voting algorithm:

Once the similar images are find out with respect to query image using content based image retrieval method, their associated user defined tags are taken into consideration for calculating relevancy of tags provided by user. All the similar images retrieved with respect to query image and their associated tags are taken for voting. The voting procedure is based on Modified Borda Count method. In a modified Borda count (MBC), the number of points given for a voter's first and subsequent preferences is determined by the total number of candidates they have actually ranked, rather than the total number standing. To implement the voting algorithm, it will be needed to define visual similarity between the images and then search visual neighbours from image database. To search thousands of images by content, efficient indexing methods are essential for speed-up. Here K-means clustering is used for its practical success. K-means clustering is performed based on visual features of image. For indexing the entire dataset is divided into smaller subsets by using K-means clustering. Each subset is indexed by a cluster centre. Then for query image, neighbours are calculated whose centres are the nearest to the query. Hence the search space is reduced.

The voting procedure is like all the user defined tags of similar images retrieved in respect to query image are listed and will participate in voting. The similar images obtained are grouped on the basis of visual features of query image. The maximum number repetitions of tags of the similar images retrived are taken at priority. The following arithmetic expression finds out the maximum occurrences of tags based on image similarity. Here t_nc_{hn} represents tags of images obtained by finding visual similarity based on color histogram feature, t_nt_{xn} represents tags of images obtained by finding visual similarity based on texture similarity feature, t_nh_{sn} represents tags of images obtained by finding visual similarity based on hsv histogram feature, t_nc_{on} represents tags of images obtained by finding visual similarity based on color correlogram feature, t_ns_{in} represents tags of images obtained by finding visual similarity based on SIFT feature.

$$[\{t_1c_{h1}, t_1c_{h2}, \ldots t_1c_{hn}\}, \{t_2c_{h1}, t_2c_{h2}, \ldots t_2c_{hn}\}, \ldots \{t_nc_{h1}, t_nc_{h2} \ldots . t_nc_{hn}\}] \cap [\{t_1t_{x1}, t_1t_{x2}, \ldots t_1t_{xn}\}, \{t_2t_{x1}, t_2t_{x2}, \ldots t_2t_{xn}\}, \ldots . \{t_nt_{x1}, t_nt_{x2}, \ldots t_nt_{xn}\}] \cap [\{t_1h_{s1}, t_1h_{s2}, .. t_1h_{sn}\}, \{t_2h_{s1}, t_2h_{s2} \ldots t_2h_{sn}\}, \ldots \{t_nh_{s1}, t_nh_{s2}, \ldots, t_hs_{hn}\}] \cap [\{t_1c_{o1}, t_1c_{o2}, \ldots, t_1c_{on}\}, \{t_2c_{o1}, t_2c_{o2}, \ldots, t_2c_{on}\}, \ldots, \{t_nc_{o1}, t_nc_{o2} \ldots, t_nc_{on}\}] \cap [\{t_1s_{i1}, t_1s_{i2}, \ldots, t_1s_{in}\}, \{t_2s_{i1}, t_2s_{i2}, \ldots, t_ns_{in}\}, \ldots ., \{t_ns_{i1}, t_ns_{i2}, \ldots, t_ns_{in}\}].$$

The maximum numbers of repetition of tags are taken in priority. The tags hence shortlisted have maximum occurrences. This is defined by following expression

$$Max\ [T_{CH}] \cap Max\ [T_{TX}] \cap Max\ [T_{HS}] \cap Max\ [T_{CO}] \cap Max\ [T_{SI}]$$

The ranking of tags are done on the basis of occurrences. The preferential order list is now matched with the user provided tags of the query image. If the user defined tags of the query image matches with the maximum of the occurrences of the visually similar images user defined tags arranged in preferential order then the tag relevancy of the query image and its associated user defined tags are correct. Hence relevancy is matched, i.e. tags are relevant to the image that is provided by the user. The proposed algorithm is defined as follows:

Step1: Collection of Images and their user provided tag are taken into consideration. (Image Database and Tag Vocabulary)
Step2: Query image is passed as input
Step 3: Calculate low level visual features of query image.
Step 4: Matching between query image features and stored image features.
Step 5: For effective indexing and speed-up k-means clustering are done based on visual features of image
Step 6: Each subset is indexed by cluster centre.
Step 7: For query image visual similar neighbours are calculated whose centres are nearest to query.
Step 8: All visually similar images related to query are listed, based on low level features.
Step 9: The similar images retrieved will be considered.
Step 10: The tags of similar images will participate in voting.
Step 11: The maximum number of occurrences of tags are taken at priority.
Step 12: $Max\ [T_{CH}] \cap Max\ [T_{TX}] \cap Max\ [T_{HS}] \cap Max\ [T_{CO}] \cap Max\ [T_{SI}]$.
Step 13: For calculating tag relevancy user defined tags of query image matched with priority tags obtained by voting.
Step 14: If match is found then relevancy of user defined tags (query image) is satisfactory.
Step 15: If match is not found then tags that has maximum occurrences are suggested for query image (Predictive tagging).

4. RESULTS AND SNAPSHOTS

This section discuss about results of implementation. The dataset used for experimental purpose is 20k social-tagged images from Flickr social media (photo sharing) site. A visual index is built for each feature by separating the entire collection into minor subsets by K-means clustering. Following are some snapshots of the implementation and results.

4.1 Similar Images retrieval snapshots

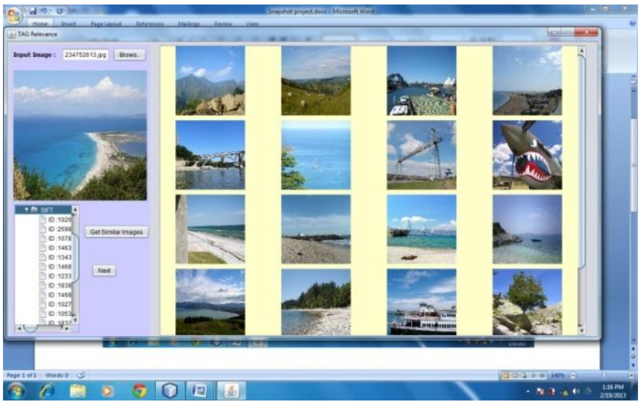


Fig 3. Query image type *beach*, Similar images retreived with respect of query image based on low level feaures.

Fig. 4 Query image type: *bridge*, similar images retrieved with respect of query image based on low level features.

Fig. 5 Query image type: *Tiger*, Similar images retrieved with respect of query image based on low level features

4.2 Tag Relevancy Snapshots

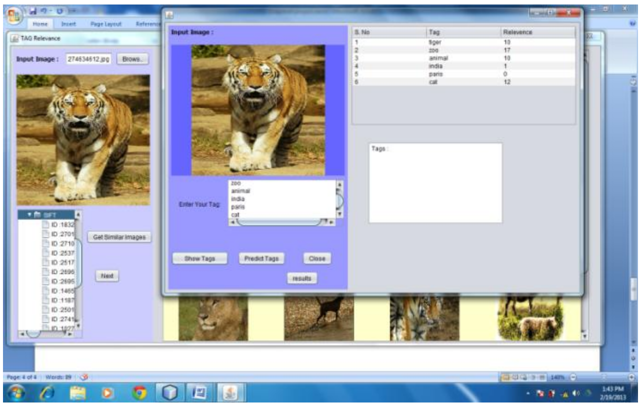


Fig. 6 Displaying results of relevancy of tag with respect to query image, tags like *tiger, zoo, animal, india, paris, cat* were checked to know tag relevancy of these tags in respect to query image.

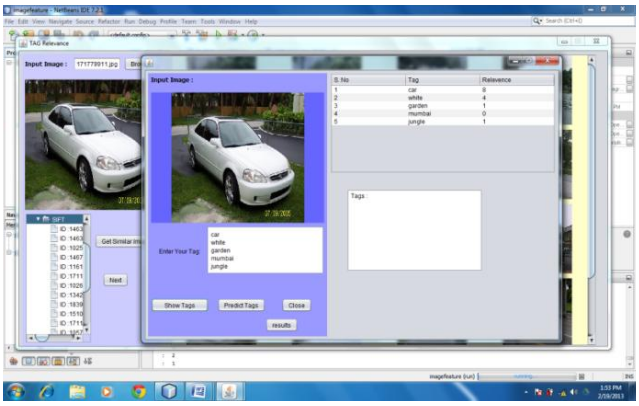


Fig. 7 Displaying results of relevancy of tag with respect to query image, tags like *car, white, garden, mumbai, jungle* were checked to know tag relevancy of these tags in respect to query image.

4.3 Predictive Tagging Snapshot

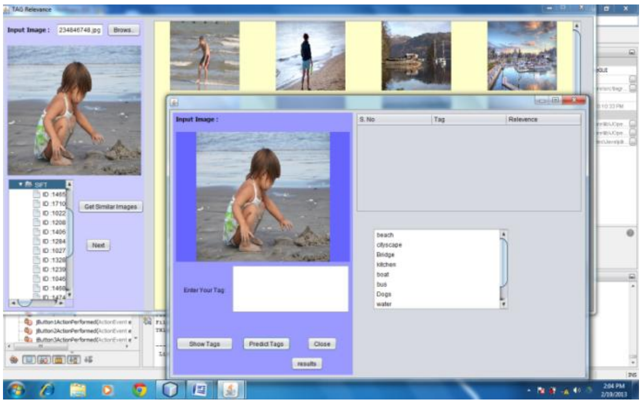
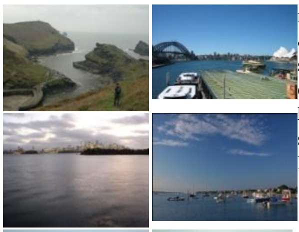


Fig. 8 Snapshot demonstrating if unlabeled image is given as input the proposed framework will predict its relevant tags based on voting.

Table I Examples of Tag prediction for labeled images by proposed method

User-tagged images		**Predicted new tags**
Image	Tags	*PredictiveTagging*
	sunset mountain	mountain sunset *cloud* *day*
	city building mexico night	city *cityscape* night
	bridge river beach	bridge sydney harbor *ocean* beach river
	boat ocean	boat *river* ocean *sea*

Table II Examples of image retrieval, tag similarity and tag prediction for unlabeled query image.

Query Image	Similar Images retrieved	Tags of similar images retrieved	Tag relevancy score out of 20	Predicted tags of query image
		mountain sea ocean harbor sydney london	12 8 11 11 5 4	*mountain* *ocean* *harbor* *sea*
		aeroplane airplane air white 737 boeing	14 12 5 2 1 2	*aeroplane* *airplane*
		tiger cat killer 2006	15 8 2 1	*tiger* *cat*

5. CONCLUSION

The popularity of online social media sites and users, all over the world, label or tag images on social media sites such as Facebook, Flickr, YouTube etc. habitually. The process of tagging is known to be highly personalized and biased in nature. Hence, the elementary drawback in social image analysis and retrieval is how to efficiently understand the significance of a tag with respect to the visual content the tag is relating. In this paper a voting algorithm is proposed as preliminary step towards sorting out the problem. The main focus is to learn the relevance of tag related to the image and to predict tags for unlabeled image based on visual neighbor search of that image. To verify the work experiments were conducted on 20k Flickr images which are tagged by different users around the globe.

6. RERERENCES

[1] C. Wang, F. Jing, L. Zhang, and H.-J.Zhang, "Scalable search-based image annotation," *Multimedia Systems*, vol. 14, no. 4, pp. 205–220, 2008.

[2] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," IEEE Trans. PAMI, vol. 30, no. 11, pp. 1958–1970, 2008.

[3] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," Jour. Machine Learning Research, vol. 3, no. 6, pp. 1107–1135, 2003.

[4] R. Datta, D. Joshi, J. Ali, J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys, Vol. 40, No. 2, Article 5, 2008, pp. 5:1-5:59.

[5] Arnold W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, 2000, pp.1349-1380.

[6] Xirong Li, Cees G.M. Snoek, and Marcel Worring, "Learning Social Tag Relevance by Neighbor Voting", in IEEE Transactions on Multimedia , volume 11, issue 7, page 1310-1322, 2009.

[7] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," IEEE Trans. PAMI, vol. 30, no. 6, pp. 985–1002, 2008.

[8] C. Cusano, G. Ciocca, and R. Schettini, "Image annotation using SVM," in Proc. SPIE, 2004, pp. 330–338.

[9] Y. Jin, L. Khan, L. Wang, and M. Awad, "Image annotations by combining multiple evidence &Wordnet," in Proc. ACM Multimedia, 2005, pp. 706–715.

[10] B. Sigurbjornsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in Proc. WWW, 2008, pp. 327–336.

[11] K. Weinberger, M. Slaney, and R. van Zwol, "Resolving tag ambiguity," in *Proc. ACM Multimedia*, 2008, pp. 111–119.

[12] Xirong Li, Cees G.M. Snoek, and Marcel Worring, "Learning Tag Relevance by Neighbor Voting for Social Image Retrieval", in Proceedings of the ACM International Conference on Multimedia Information Retrieval (MIR), Vancouver, Canada, October, 2008.

[13] Li, Xirong, Cees GM Snoek, and Marcel Worring. "Unsupervised multi-feature tag relevance learning for social image retrieval." In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pp. 10-17. ACM, 2010.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, vol. 60, no. 2, pp. 91–110, 2004.

[15] Long, Fuhui, Hongjiang Zhang, and David Dagan Feng. "Fundamentals of content-based image retrieval." *Multimedia Information Retrieval and Management* 17 (2003): 1-26.