ISSN:0975-9646

An Improved Privacy Policy Inference over the Socially Shared Images with Automated Annotation Process

Sangeetha. J [1], Kavitha. R [2]

[1]*M.Phil. Research Scholar, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India*

[2]*Asst. Professor, Dept. of Computer Science, PSGR Krishnammal College for Women, Peelamedu, Coimbatore, India*

***Abstract*— Usage of social media's increased considerably in today world which enables the user to share their personal information like images with the other. This improved technology leads to privacy violation where the users are sharing the large volumes of images across more number of peoples. To provide security for the information, automated annotation of images are introduced which aims to create the meta data information about the images by using the novel approach called Semantic annotated Markovian Semantic Indexing(SMSI) for retrieving the images. The proposed system automatically annotates the images using hidden Markov model and features are extracted by using color histogram and Scale-invariant feature transform (or SIFT) descriptor method. After annotating these images, semantic retrieval of images can be done by using Natural Language processing tool namely Word Net for measuring semantic similarity of annotated images in the database. Experimental result provides better retrieval performance when compare with the existing system.**

***Keywords*— Semantic annotated Markovian Semantic Indexing, hidden Markov model, Hidden Markov Model.**

I. INTRODUCTION

Social media is the two way communication in Web 2.0 and it means to communicate, share, and interact with an individual or with a large audience. Social networking websites are the most famous websites on the Internet and millions of people use them every day to engage and connect with other people. Twitter, Facebook, LinkedIn and Google Plus seems to be the most popular Social networking websites on the Internet.

Today, for every single piece of content shared on sites like Facebook—every wall post, photo, status update, and video—the up loader must decide which of his friends, group members, and other Facebook users should be able to access the content. As a result, the issue of privacy on sites like Facebook has received significant attention in both the research community [1] and the mainstream media [2]. Our goal is to improve the set of privacy controls and defaults, but we are limited by the fact that there has been no in-depth study of users' privacy settings on sites like Facebook. While significant privacy violations and mismatched user expectations are likely to exist, the extent to which such privacy violations occur has yet to be quantified.

Images are now one of the key enablers of users' connectivity. Sharing takes place both among previously established groups of known people or social circles (e.g., Google+, Flickr or Picasa), and also increasingly with people outside the users social circles, for purposes of social discovery to help them identify new peers and learn about peers interests and social surroundings. With the increasing volume of images users share through social sites, maintaining privacy has become a major problem, as demonstrated by a recent wave of publicized incidents where users inadvertently shared personal information. In light of these incidents, the need of tools to help users control access to their shared content is apparent.

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords or descriptions to the image retrieval can be performed over the annotation words. Manual image annotation is time consuming, laborious and expensive to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase social web applications and the semantic web have inspired the development of several web-based image annotation tools.

Automatic image annotation [6] is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database. This method can be regarded as a type of multi-image classification with a very large number of classes large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images.

II. RELATED WORK

In [3] Sergej Zerr propose a technique Privacy-Aware Image Classification and Search [8] to automatically detect private images, and to enable privacy-oriented image search. It combines textual meta data images with variety of visual features to provide security policies. In this the selected image features (edges, faces, color histograms) which can help discriminate between natural and man-made objects/scenes (the EDCV feature) that can indicate the presence or absence of particular objects (SIFT). It uses various classification models trained on a large scale dataset with privacy assignments obtained through a social annotation game.

Learning the Semantics of Words and Pictures [5] present a method which organizes image databases using both image features and associated text. By integrating the two kinds of information during model construction, the system learns links between the image features and semantics which can be exploited for better browsing, better search, and novel applications such as associating words with pictures, and unsupervised learning for object recognition.

In [6] developed an approach Markovian Semantic Indexing (MSI) a new method for automatic annotation and annotation-based image retrieval. The proposed system allows the retrieval technique to benefit from the underlying structure of the annotation data. The proposal is to provide the best image based on the user query with the efficient processing. When the user clicked on the image the indexing is automatically performed and the search result will be displayed. It provides efficient and effective search result.In [7] discussed Markovian Semantic Indexing (MSI) for automatic annotation based image retrieval. This method is suitable for Annotation Based Image Retreival (ABIR) when the per image annotation data is limited.

In the existing work, Adaptive Privacy Policy Prediction (A3P) system is used to help users compose privacy settings for their images. The A3P system consists of two main components: A3P-core and A3P-social. When a user uploads an image, the image will be first sent to the A3P-core. The A3P-core classifies the image and determines whether there is a need to invoke the A3P-social. In most cases, the A3P-core predicts policies for the users directly based on their historical behaviour. A3P-core will invoke A3Psocial when the user does not have enough data for the type of the uploaded image to conduct policy prediction and the A3P-core detects the recent major changes among the user's community about their privacy practices along with user's increase of social networking activities such as addition of new friends, new posts on one's profile etc. In above cases, it would be beneficial to report to the user the latest privacy practice of social communities that have similar background as the user.

The A3P-social groups users into social communities with similar social context and privacy preferences, and continuously monitors the social groups. When the A3P-social is invoked, it automatically identifies the social group for the user and sends back the information about the group to the A3P-core for policy prediction. At the end, the predicted policy will be displayed to the user. If the user is fully satisfied by the predicted policy, he or she can just accept it. Otherwise, the user can choose to revise the policy. The actual policy will be stored in the policy repository of the system for the policy prediction of future uploads. The main disadvantages of the system are (1) Inaccurate privacy policy generation in case of the absence of meta data information about the images and (2) Manual creation of Meta data log data information leads to inaccurate classification and also violation privacy.

III. PROPOSED ALGORITHM

The proposed system introduced Semantic annotated Markovian Semantic Indexing(SMSI), a novel semantic retrieval of images is done based on Hidden Markov model based annotated images. Automatic image annotation phase makes use of a manually annotated training set taken to generate an annotated image database. Annotation based image retrieval phase gets a user query, and then finds similar terms for the query with the help of WordNet. Also discover the similarity between the query and images in annotated image database. Then find the similarity between matching images.

The system carries two major tasks.

- Automatic image annotation
- Annotation based image retrieval

Automatic image annotation phase makes use of a manually annotated training set taken to generate an annotated image database. Annotation based image retrieval phase gets a user query, then find similar terms for the query with the help of WordNet. Also discover the similarity between the query and images in annotated image database. Then find the similarity between matching images. To annotate the images in database, features such as Color and texture feature are extracted by using Color Histogram and SIFT Descriptors methods. Fig.1 and Fig.2 shows the extraction of image features and meta data features.

- **Color Histogram Feature**

Color histogram is simplest and most frequently used to represent color. The color histogram serves as an effective representation of the content. Color is one of the most important features of images. Color features are defined subject to a particular color space or model. A number of color spaces have been used such as RGB, LUV, and HSV. Once the color space is specified, color feature can be extracted from images or regions. An important color features namely color histogram is extracted. Color histograms are frequently used to compare images. In this gray level variations are used to compute the histogram of any image. For this purpose the color image is first converted in to gray level image. Then the histogram values are computed for gray level variations. According to histogram values, images are extracted from the database.

In color histogram the number of pixel of given color is calculated the color histogram extraction algorithm involves following three steps.

- Partition of color space into cells.
- Association of each cell to a histogram bin.
- Counting of number of image pixel of each cell and storing this count in the perspective corresponding histogram bin.

Texture is a very useful characterization for a wide range of image. It is generally believed that human visual systems use texture for recognition and interpretation. This feature has been extracted by using SIFT descriptor. SIFT based analysis involves detecting salient locations in an image and extracting descriptors that are distinctive yet invariant to changes in viewpoint, illumination, etc. To extract these texture features SIFT descriptors are used. With the help of these extracted features, Images are annotated by using

Hidden Markov model. The parameters of the model are estimated from a set of manually annotated images. Each image in a large test collection is then automatically annotated with the a posteriori probability of concepts present in it. After annotating images, images are semantically retrieved based on Natural Language processing tool namely WordNet[9]. Semantic Similarity based Image Retrieval Model is used for discovering similarities between Images in the database with the query image containing conceptually similar terms. These methods are implemented and evaluated using WordNet.

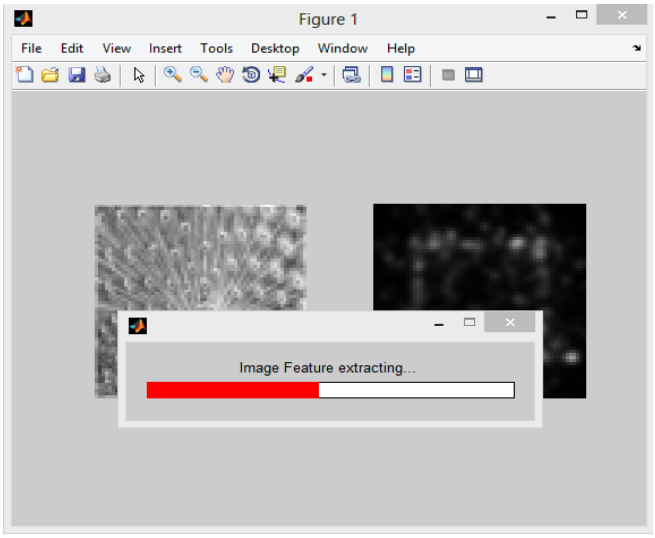

Fig.1 extracting image

- **Texture Feature Extraction**

In this section, Texture feature are extracted by using SIFT (Scale-invariant feature transform) descriptor [10]. Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe texture features in images.

- **SIFT Descriptors**

SIFT based analysis involves detecting salient locations in an image and extracting descriptors that are distinctive yet invariant to changes in viewpoint, illumination, etc. The standard SIFT interest point detector and the standard SIFT histogram-of-gradients descriptor can be used. These 128 dimension descriptors can be thought of roughly as summarizing the edge information in an image patch centered at an interest point. We term the 128 dimension descriptors the local SIFT descriptors for an image. We also compute a single global SIFT descriptor. This global descriptor is a frequency count of the quantized local descriptors. We use the clustering algorithm to cluster a large collection of SIFT descriptors and label each local descriptor with the id of the closest cluster center. The global SIFT descriptor is then computed as

$$SIFT_{GLOBAL} = [t_0, t_1, \ldots, t_{k-1}]$$

where ti is number of occurrences of the quantized Texture features with label i. SIFTGLOBAL is similar to a term vector in document retrieval. The global SIFT descriptors are normalized to have unit length to account for the varying number of local SIFT descriptors per image.

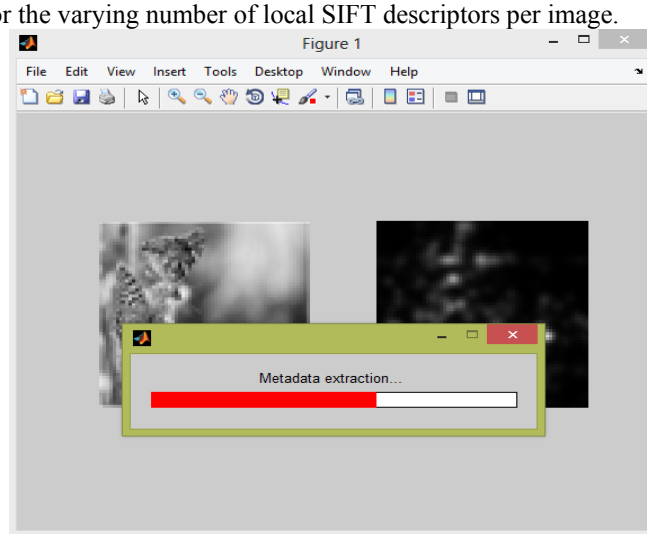

Fig.2 Metadata extraction

- **Database updation with HMM annotated images**

Hidden Markov Model, provides[8]

- Estimating the parameters of the model from annotated image+ caption pairs,
- Aligning image-regions with caption-words in an image+ caption pair, and
- Computing the likelihood of a caption-word being present in an image

Let a collection of image+ caption pairs be provided and consider the problem of developing a stochastic generative model that jointly describes each pair

Let I = <i1….it> denote segments (regions) in an image, and C = {c1. . . cN} denote the objects (concepts) present in that image, as specified by the corresponding label (caption) C. For each image region it, t = 1,. . . T, let $x_t \in$ IRd represent the color, texture, of the region1. Finally, let V denote the global vocabulary of the caption-words cn across the entire collection of images.

We propose to model the {xt}-vectors of an image I as a hidden Markov process, generated by an underlying unobserved Markov chain whose states st take values in C. Specifically, each xt is generated according to some probability density function f(·|st) given the state st, where st itself is a Markov chain with a known initial state s0 and transition probabilities p(st|st−1).

The state sequence is as follows:

$$f(x_1, \ldots, x_T, s_1, \ldots, s_T | s_0) = \prod_{t=1}^{T} f(x_t|s_t) p(s_t|s_{t-1})$$

Since this level of detail is usually not provided in a caption, a hidden Markov model (HMM) is an appropriate formalism for computing the joint likelihood:

$$f(x_1^T, C | s_0) = \sum_{s_1^T \in C^T} \prod_{t=1}^{T} f(x_t|s_t) p(s_t|s_{t-1})$$

where x_1^T denotes the T-length sequence <x1, .., xT >.

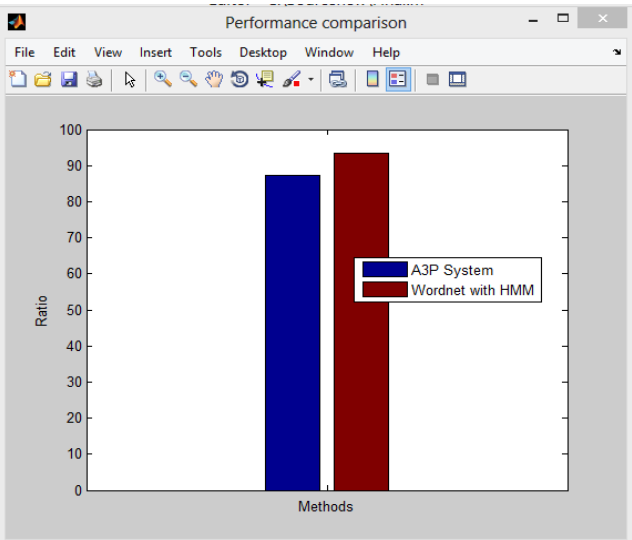

Fig.3 Performance Evaluation

IV. Implementation Details

The proposed system was implemented with MATLAB. A dataset with 200 images, in 4 classes of about 50 images in each class, annotated with keywords, each annotation being a text string of up to 25 keywords is created manually. The system reads images and text in the database extracts features of image and metadata. The features of the images are extracted by using SIFT descriptor and the images are annotated by Hidden Markov Model. The annotated images are semantically retrieved based on Wordnet. After that we can select an image from one of the four image classes. The system provides the security policy for the selected image. Experimental result of automatic image annotation system compared to the existing system is shown in figure 3. The proposed system provides improved accurate generation of the privacy policy for the users rather than the existing retrieval system of framework.

V. Conclusion

The present work proposes Semantic annotated Markovian Semantic Indexing (SMSI) a semantic image retrieval is done and its performance improved by incorporating an automatic annotation system. Automatic annotation of images in database has been done by using a proposed Hidden Markov model which uses the extracted features (color and texture) where all states represent the concepts. Semantic similarity based image retrieval can be done with the use of Natural language processing tool namely WordNet where conceptual similarity between natural language terms were done. Comparative result provides better result for proposed system rather than existing retrieval system of framework.

References

[1] S. Ahern, D. Eckles, N. Good, S. King, M. Naaman,and R. Nair. Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing. CHI, 2007.

[2] B. Krishnamurthy and C. E. Wills. Characterizing Privacy in Online Social Networks. WOSN, 2008.

[3] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova , I Know What You Did Last Summer!:Privacy-Aware Image Classification and Search , Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 2012.

[4] Anna Cinzia Squicciarini, Member, IEEE, Dan Lin, Smitha Sundareswaran, and Joshua Wede, “Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites”, IEEE Transactions on Knowledge and Data Engineering, Vol. 27, NO. 1, January 2015.

[5] K. Barnard and D. Forsyth, \Learning the semantics of words and pictures," in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, vol. 2. IEEE, 2001.

[6] Swarna Kumari. T, Ramadevi Y and Sangeetha A, “Markovin Semantic Indexing Method for Online Image Retrieval System”, International journal of current research, Vol.7, Issue 03, March 2013.

[7] K. A. Raftopoulos, K. S. Ntalianis, D. D. Sourlas, and S. D. Kollias, Mining user queries with markov chains: Application to online image retrieval," Knowledge and Data Engineering, IEEE Transactions on, Vol. 25, 2013.

[8] A. Ghoshal, P. Ircing, and S. Khudanpur, Hidden markov Models for automatic annotation and content-based retrieval of images and video," in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent dirichlet allocation, The Journal of machine Learning Research, Vol. 3, 2003.

[10] F. Yu and H. H.-S.Ip, “Automatic semantic annotation of images using spatial hidden markov model," in Multimedia and Expo, 2006 IEEE International Conference on. IEEE, 2006.