

# PRIVACY POLICY AND DEDUPLICATION OF USER UPLOADED IMAGES ON SOCIAL WEBSITES

By

J. SANTHIYA \*

K. NIROSHA \*\*

N. KALAIVANI \*\*\*

R. JAYANTHINISHA \*\*\*\*

K. KAMATCHI \*\*\*\*\*

\* Assistant Professor, Department of Information Technology, A.V.C College of Engineering, Mayiladuthurai, India.  
 \*\*\_\*\*\*\*\* UG Scholars, Department of Information Technology, A.V.C College of Engineering, Mayiladuthurai, India.

## ABSTRACT

Usage of social medias have increased now-a-days. User share personal information and images through social network. While maintaining the privacy has become a major problem and also duplication of images have reduced the capacity of the databases. The authors a two level framework for maintaining privacy and securing the photo share. Users who want to maintain privacy are usually provided with access control. The canny edge detection technique is used for the deduplication of images. It increases the storage space capacity of the database. Watermarking is used for every image that is shared on the website for the copyright protection and restricting the images on other websites.

Keywords: Privacy, Online Content, Social Website, Image, Edge Detection.

## INTRODUCTION

Images are widely shared across the world by the social medias. Social media means communicating through the internet. Social media is broadly defined as an inexpensive and widely accessed electronic tool that can permit anyone to publish images and content on social website. Users are able to share the photos and text on the social website [9]. Many forms of social website are available. It allows people from all parts of the world to interact with each other. Through the social media, a user can upload and download the photos [8], post their thoughts on blogs and even participate in online discussions.

Social media contains virtual world content and media sharing sites. Virtual world content sites offer game like virtual environment in which, the users interact. Media sharing sites allow the users to post videos or photographs. Protection of the images is a major part while creating a website. Adaptive Privacy Policy Prediction (A3P) system aims to provide a user friendly privacy settings experience by automatically generating these policies. The A3P system handles the user uploaded images and factors by the following criteria for privacy settings of images: The

impact of social and personal characteristics with the user profile and relationship provide useful information for the privacy setting. The role of image's content and metadata is considered for the image classification. Similar images often have similar privacy preferences [10].

One may upload several photos of their kids and allow only the family members to see those photos. The A3P-core focuses on analyzing the user's own image and metadata, while the A3P-Social offers privacy setting recommendations for a user's preferences. The authors design an interaction between the two building blocks to balance the benefits of the user's personal characteristics.

## 1. Related Work

Dong Liu et al., [1] proposes image retagging scheme for improving the quality of the tags associated with content relevance. Applying knowledge based method to differentiate visual content of the images. To improve this, the user is provided with tag by tag filtering (content related & content unrelated). The techniques used for images are Tag filtering and Tag refinement. Tag filtering from the part-of-speech we only considers nouns. The

noun tags were further analyzed and an automatic process is adopted to detect visual property of the tags. High level categories including "organism", "artifact", "thing", "color" and "natural phenomenon". Tag refinement is relevant to the image, it is labeled as positive, and otherwise it is negative. Observation mentioned above is mainly applicable for content related and content unrelated tags. Those tags have resemblance with the visual content. Content Based Annotation Refinement (CBAR) provides the tags for image.

Ricardo Da Silva et al., [2] proposed a content based image retrieval to retrieve the images based on user-defined specification (e.g., shape, size, and sketch). Their goal is to support image retrieval based on content properties (e.g., shape, color, texture) encoded into the feature vectors. CBIR approach is used as an automatic retrieval process, instead of traditional keyword-based approach. Images in these systems have the image processing algorithms (usually automatic) to extract feature vectors, that represent image properties such as, color, texture, and shape. One of the main advantages of this approach is the possibility of automatic retrieval process, contrasting to the effort needed to annotate images.

Heather Richter et al. [3] proposes privacy setting in facebook for images to improve privacy management in Facebook and other sites by structuring privacy settings. It reduces the inadvertent sharing of personal information. Designing a new interface for managing the privacy settings on facebook focusses on audience point of view. First iteration, created and examined the audience view without any mechanism for modifying settings, similar to Orkut's interface. They examined the effects of this design and informed the next prototype iteration of the user setting. Main concepts like comparing the standard Facebook privacy settings interface with their prototype using audience view interface.

Mitra Basu [4] proposed the edge detection for the deduplication techniques. It is mainly concentrated on the images with gray scale with Gaussian-based edge detection, which presents the hybrid detector for the

edge detection and the noise variation has to be estimated. The separation of a scene into object is carried out by the human visual system. The edge segments are combined from different scales. Also how to choose the size of the filters and how to combine the edges by multistage analysis. The multi resolution the edge is incorporated by the Gaussian filter.

Ching-man Au Yeung et al. [6] proposed an access control for the online photo albums about the semantic web and linked data. The semantic web is designed to describe data with machine readable metadata. Tagging is referred as the act of assigning descriptive keywords to the online resources. It proposes a access control system by the decentralized authentication protocol. It allows creating expressive policies.

Katherine Strater et al. [5] proposes strategy and struggles with privacy for the social networks for creating facebook like social websites with the privacy preference compared with the A3P method. It describes about the public profiles, private profiles and the search profile. The strategies are about the privacy with the one-time event based on the report of the participation.

Su et al. [7] proposed a collaborative filtering for the data. The major challenges are the sparsity of data, synonym and the privacy protection, etc. It is widely adopted and the assumption made by similar behaviors like watching, listening and buying. These are done with the memory based CF, Model based CF, and Hybrid based CF techniques. For better prediction, they used a model-based Collaborative filtering. The CF systems find the degree of variability and predict the performance by the synonym.

## 2. Proposed Approach

A3P system provides the privacy preferences automatically to the images, when they are uploaded on the website. The automation process is done by the core and the social part of A3P technique. Images classification is also done by the A3P for the deduplication of uploaded images on the database. System architecture is shown in Figure 1.

The privacy preferences are applied on the website and

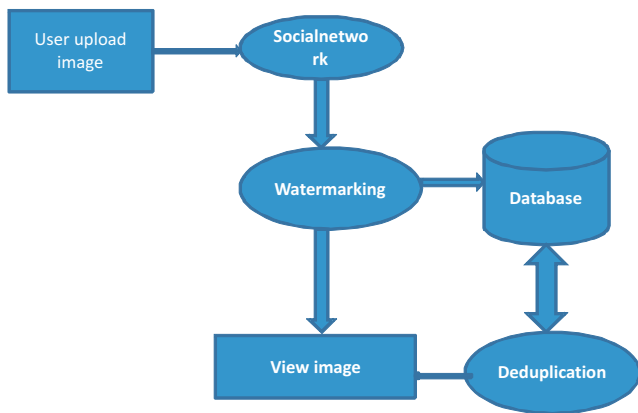


Figure 1. System Architecture

the unnecessary access is restricted by the tagging concepts. If the tagging is for family or friends, they can view and download the images. Strangers cannot have the the option to download the images.

*Policy Mining:* The privacy policies are the first privacy preferences expressed by the users. By applying different rules, the policy mining deals further. It follows the order in which a user defines a policy and decides what rights must be given to the images. This hierarchical mining approach starts by looking the subjects and their actions in the policies and finally for condition.

Watermarking technique is proposed for every image that is uploaded on the site. It is recommended for the protection of the images. A digital watermark is a signal permanently embedded into data i.e. image. Digital watermarking techniques are derived from steganography, which means covered writing. Digital Watermarking works by the embed information within digital data, and it cannot be detected without any special software for the purpose of making sure that the concealed data is present in all copies.

Deduplication technique is used for the image repetition on the database. Canny edge detection technique is used for the deduplication technique, which reduces the storage space of the image.

### 3. Module Description

#### 3.1 API Implementation of Social Network

- *User Registration*

New user sign up - new database is created with an

unique id for the user.

Existing user sign in - check the database for the id and password references.

The user registration is done with the name, date of birth, gender, email id and password. The details are stored and further logins are only checked with the id and the password. The existing user logged on the page by the e-mail id and the password. It is checked on the database and the welcome page is opened.

#### 3.2 Privacy Preferences of User Uploaded Image

- The user creates relationships with other user as a friend, family and workers.
- User have the options like message and post or upload a photo.

While uploading a photo, he or she can tag friends or family members and give privacy preferences.

Images are stored in the user databases.

#### 3.3 Image Classification

- The image classification is done for the deduplication technique. It is done by the content based and the metadata based classification.

#### 3.4 Watermarking

- Every time a photo is uploaded by the user, watermarking is applied on the image. The digital watermarking technique is automatically applied on the image.

Secure the image by water marking:

- The authors create a website with a condition that, the watermarked images are not uploaded by the user. Because it is considered as hacking of the image without the owner's permission.

### Conclusion

Social network is an upgrading media for the information sharing. It provides content sharing as both text and the images. Hence protection of the image is concentrated. For those issues, the authors proposed a Adaptive Privacy policy prediction for the privacy setting of a user for their images. The image are watermarked and then stored. The metadata based classification is done for the data deduplication technique. The two level frameworks

describe the social behaviors of a user and thus, predicts the privacy options automatically.

## References

- [1]. D. Liu, X.S. Hua, M. Wang, and H.J. Zhang, (2010). "Retagging Social Images Based on Visual and Semantic Consistency". in *Proceedings of 19<sup>th</sup> ACM International Conference on World Wide Web*, pp.1149-1150.
- [2]. R. Da Silva Torres and A. Falcao, (2006). "Content-based Image Retrieval: Theory and Applications". *Revista de Informatics Teoricae Aplicada*, Vol.2, No.13, pp.161-185.
- [3]. H. Lipford, A. Besmer, and J. Watson, (2008). "Understanding Privacy Settings in Facebook with an Audience View". in *Proceedings of Conference Usability, Psychology, and Security*.
- [4]. Mitra Basu, (2002). "Gaussian-Based Edge-Detection Methods-A Survey". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.135, No.3, pp.252-260.
- [5]. Katherine Strater and Heather Richter Lipford, (2008). "Strategies and Struggles with Privacy in an Online Social Networking Community". *Proceedings of the 22<sup>nd</sup> British HCI Group Annual Conference on People and Computers: Culture, Creativity, and Interaction*, Vol.1, pp.111-119.
- [6]. Ching-man Au Yeung, Lalana Kagal, Nicholas Gibbins, and Nigel Shadbolt, (2009). "Providing Access Control to Online Photo Albums Based on Tags and Linked Data". Retrieved from <http://dig.csail.mit.edu/2008/Papers/AAAI%20SSS%202009/paper.pdf>
- [7]. X. Su and T. M. Khoshgoftaar, (2009). "A Survey of Collaborative Filtering Techniques". *Advances in Artificial Intelligence*, Vol.2009, pp.19.
- [8]. Anna Cinzia Squicciarini, (2015). "Privacy Policy Inference of User-Uploaded Images on Content Sharing Sites". *IEEE Transactions on Knowledge and Data Engineering*, Vol.27, No.1.
- [9]. Anna C. Squicciarini, Mohamed Shehab, and Federica Paci, (2009). "Collective Privacy Management in Social Networks". Retrieved from <http://www2009.eprints.org/53/1/p521.pdf>
- [10]. Sattarova Feruza Y. and Tao-hoon Kim, (2007). "IT Security Review: Privacy, Protection, Access Control, Assurance and System Security". *International Journal of Multimedia and Ubiquitous Engineering*, Vol.2, No.2, pp.17-32.

---

## ABOUT THE AUTHORS

J. Santhiya is currently working as an Assistant Professor in the Department of Information Technology at A.V.C. College of Engineering, Mayiladuthurai, India. She obtained her B.E Degree from Annamalai University and M.Tech Degree from SRM University, India. Her area of research includes Network Security and Web services.

N. Kalaivani is currently pursuing B.Tech Degree in the Department of Information Technology at A.V.C College of Engineering, Mayiladuthurai, India.

K. Nirosha is currently pursuing B.Tech Degree in the Department of Information Technology at A.V.C College of Engineering, Mayiladuthurai, India.

R. Jayanthinisha is currently pursuing B.Tech Degree in the Department of Information Technology at A.V.C College of Engineering, Mayiladuthurai, India.

K. Kamatchi is currently pursuing B.Tech Degree in the Department of Information Technology at A.V.C College of Engineering, Mayiladuthurai, India.