I J C T A, 8(5), 2015, pp. 2209-2218

Digital Evidence from Social Networking Sites by using Metadata

P. Krishna Kumaran Thambi*, Krishnaveni S.H.**

Abstract: In this modern age in which we are living, digital images play a vital role in many application areas like social networking websites, for example, Facebook. But at the same time the image retouching techniques has also increased which forms a serious threat to the security of digital images in Facebook. To cope with this problem, the field of digital forensics and investigation has emerged and provided some trust in digital images.

Photograph contain metadata which reveals the details, related to the image such as date and time, geographic location information, make and model of the camera used to take the picture. While uploading a photo the image size will reduce by removing the metadata from it. Because of this mechanism the artifacts which are downloaded from the sites lacks metadata, which is a major challenge for the investigators. This paper proposes a mechanism to recover the image tag which was purposefully removed by the social media. Image tag contains source IP, MAC and timestamp and thereby the system and location.

Keywords: Image metadata, Social networking sites, Digital image forensics

1. INTRODUCTION

In this digital era, the social media is not only been used for entertainment but also for promoting various activities of the business and educational institutions .In such a situation images have become inevitable to convey the ideas. Images, unlike text, represent an effective and natural communication media for people, due to their immediacy and the easy way to understand the image content. Whenever digital images are understood as a means to convey information, it is important to ensure the trustworthiness of this very information. This means in particular that the image has to be authentic, i.e., the image has not been manipulated and the depicted scene is a valid representation. Despite of entertainment the diverse and anonymous nature of social networking websites makes users highly vulnerable to cybercrimes. In such a situation the verification of the integrity of the image had become much more important. Image files can contain information about the content of the images, the image raster, and image metadata. Metadata is becoming increasingly important in this age of digital photos where users are looking for a way to store information with their pictures that is portable and stays with the file, both now and into the future.

Uploading images into social networking media, it's sharing and manipulation had grown into a major problem. While an image is uploading into social networking sites, they remove the metadata since it is not possible to hold such a large volume of data in the server. Because of this reason, even though many tools are available for digital forensics, the recovery of only a little information is possible . The wide spread availability of user-friendly image editing tools also make the investigation a difficult one. An image contain not only the picture information, rather it hold details about the camera signature, geographic location from which the image is taken and sometimes the information about the photographer too. These characteristics allow assessing image trustworthiness independent of the actual image data. It is also possible to link this information to the image and to interpret inconsistencies as processing artefact. The removal of metadata

* Sree Narayana Gurukulam College of Engineering, Kolencherry, Eranakulam-682311 *Email: pkkthampi@gmail.com*

** Noorul Islam Centre for Higher Education, Kumaracoil, Thuckalay, Kanyakumari, Tamil Nadu 629180, *Email: shkrishnaveni@gmail.com*

by social networking sites prevents the exposure of evidence of crimes. Morphed or system generated image cannot be distinguished from the real one.

Digital photography has been well accepted and embraced. Since the professional range cameras are available to the public at affordable rate, the advances of digital cameras and their corresponding technology have become so common. With the increased use of camera in the society, criminals have taken the advantages of this technology. The need and use of digital photographs had grown to a wide range and the real images will posses all the information about its origin and thus the history. Much of today's photo-editing and image management software offers capabilities for embedding and editing image files, and there are also many specialized utilities for working with the digital images. Morphing had grown into a crime which is only a little traceable and a time consuming one to identify that it is not real. Most of the criminal activities than happen in the websites are related to the digital photographs and its manipulation.

Criminal activity based on the image morphing is widely spreading now a days. The lacks of metadata in the images which are uploaded to the social networking sites make the identification of the culprit a challenging task to the investigators. But the other side which should be remembered is that, today digital images have been introduced as evidence to the courts of law. Judging about the trustworthiness of a digital image means to infer the history of that particular image. An image processing experts can easily access and modify the image content and thereby it's meaning, without leaving any traces of tampering. The image which holds the metadata could reveal its history also. Digital image forensics aims at recovering metadata and thus information about its history. The extraction of information from the server log is impractical without the IP address details. Since it gives an idea about the location from which the image had taken and uploaded it could open a way for the future investigation purposes. The reliability of digital visual information has been questioned due to the ease in counterfeiting both its origin and content. Here lies the importance of the proposed system that is the addition of an extra tag which holds the details of the images that a person is uploading.

2. BACKGROUND AND RELATED WORKS:

2.1. Image Metadata

Metadata is "data about data". In the age of digital photos, image metadata makes an easy way for storing information within that. Descriptive information about the picture embedded inside an image refers to image metadata that might include the characteristics of the photo like date and location, camera make and mode, location and copyright information etc. Information stored in an image file is always with the image are in standard formats. IPTC, IPTC-IIM and XMP are the three most commonly used metadata formats for image files. The use of embedded IPTC tags in image file formats became widespread with the use of the Adobe Photoshop tool for image editing. Digital photo containing EXIF (exchangeable image file format) metadata usually are not editable which are auto-generated that includes static information such as the camera model and make, and information that varies with each image. IPTC are mostly "user-entered". Users may also want to upload images from a number of different sources that make use of different subsets of the three image metadata standards supported. The embedded metadata in a digital image is fragile .In some cases, simply uploading an image to a website, or having it processed online to a different size might result in a partial or total loss of metadata. Metadata makes an image valuable. Embedded photo metadata will make it easier to store, find and share information now and in the future. Removing embedded metadata is against the law in the United States under the Digital Millenium Copyright Act.

2.2. Social networking Sites

Over recent years, online social networks (OSNs) have become the largest and fastest growing community on the Internet; Millions of people around the world with access to the Internet are members of one or more social networks. They have a permanent online presence where they create profiles, share photos, share their thoughts with friends and spend hours catching up with what their hundreds of friends are doing with their lives. Human beings are social creatures, and therefore, are dedicated to creating and participating in “social networks” in order to express and share their ideas. Every day over 200 million new photos are uploaded and download images to Facebook may create a chance for making genuinity problem from the side of cyber forensics.

Social networking sites provides a platform for users to upload and share their own images and designed in such a way that it can handle effectively the billions of photos and images upload online each year continue to proliferate. Nowadays users can upload images to the social networking sites direct from the capture device, encode pertinent contextual information automatically and, perhaps most importantly for them, make it available to others and allow them to annotate and generally interact with the image as well. While Flickr is indeed a one of the most popular photo-sharing sites, there are others, including Facebook and Google Picasa Web Albums. Particularly in the case of the former, which specialises in social interaction among its 800+ million users, and for which photos and other media are a secondary focus, the greater richness of social data available in that system.

2.3. Digital forensics

Digital forensics, the recovery and investigation of material is often related to computer crime found in digital devices and aims at validating the authenticity of images by recovering information about their history. Digital image are the useful source of evidence in crime. Two main problems are addressed: the identification of the imaging device that captured the image, and the detection of traces of forgeries. In the digital era images are the main information carries and represent the common source of evidence. The expressive potential of visual media and the ease in their acquisition, distribution and storage is such that they are more and more exploited to convey information.

Digital image forensics (DIF) aims at providing tools to support blind investigation. From existing multimedia security-related research domains like Watermarking and Steganography and exploits image processing and analysis tools to recover information about the history of an image. Two principal research paths evolve under the name of Digital Image Forensics. The first one, image source device identification techniques, includes methods that attempt at answering by performing some kind of ballistic analysis to identify the device that captured the image. The second tampering detection techniques contains group of methods aims instead at exposing traces of forgeries by studying inconsistencies in natural image statistics. Digital Image Forensics has a very precise role among multimedia security disciplines: authenticating images for which no reference is known and no previous integrity protection has been set. For this reason, they are often referred to as passive and blind. This makes DIF techniques the only applicable tools in a large set of practical situations.

2.4. Steganography

Steganography consists in communicating secretly via some media in particular images and videos. Also, one can assume that the stego-picture will not undergo photometric or geometric attacks among the transmission. Steganography sometimes is used when encryption is not permitted. Or, more commonly, steganography is used to supplement encryption. An encrypted file may still hide information using steganography, so even if the encrypted file is deciphered, the hidden message is not seen. The main point for two persons who communicate some information using this technology is to be not detected by a third party. To make the message not detectable, algorithms mix secret information. Steganography today, however, is significantly more sophisticated than the examples above suggest, allowing a user to hide large amounts of information within image and audio files. These forms of steganography often are used in conjunction with cryptography so that the information is doubly protected, first it is encrypted and then hidden so that an adversary has to first find the information and then decrypt it.

2.5. LZW Compression

Lempel–Ziv–Welch (LZW) is a simple compression algorithm which is lossless and adaptive in nature. LZW encodes sequences of 8-bit data as fixed-length 12-bit codes. The codes from 0 to 255 represent 1-character sequences consisting of the corresponding 8-bit character, and the codes 256 through 4095 are created in a dictionary for sequences encountered in the data as it is encoded. At each stage in compression, input bytes are gathered into a sequence until the next character would make a sequence for which there is no code yet in the dictionary.

The encoding algorithm works by scanning through the input string for successively longer substrings until it finds one that is not in the dictionary. When such a string is found, the index for the string without the last character is added to the dictionary with the next available code. The last input character is then used as the next starting point to scan for substrings. In this way, successively longer strings are registered in the dictionary and made available for subsequent encoding as single output values.

To decode an LZW-compressed archive, one needs to know in advance the initial dictionary used. But one of the attractive feature is that the additional entries can be reconstructed as they are always simply concatenations of previous entries.

3. PROPOSED SYSTEM

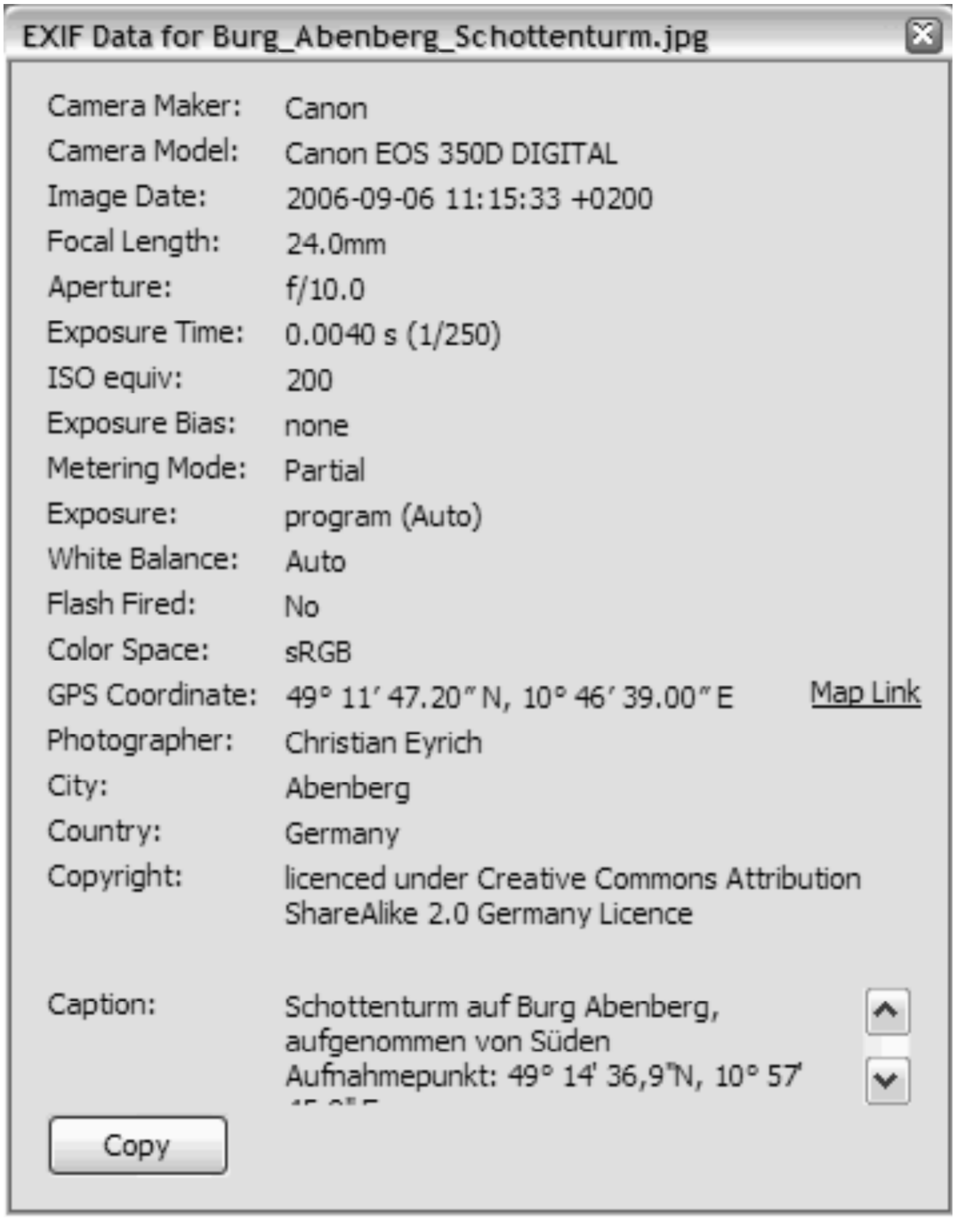
World is getting digitalized and the digital photographs holds an inevitable position in the social networking media for entertainment and for the promotions of various activities of industries and educational institutions. Since the crime rate in this area is increasing, many methodologies have been adapted for forgery detection. From the investigators point of view any details about an image is forensically sound.

Balkirat Kaur etal [6] demonstrate a novel solution to detecting a tampered image by creating and capturing the image signature from the hex file of the image using the values in Jpeg header including Huffman and Quantization values, camera specification values and JFIF format. Find the authenticity of the image is the first part of this paper through which we are trying to protect the privacy of people. Next part of the paper will be to find a way to implement this into the social websites such that a tampered image posted or tagged on the social media can be easily identified and flagged. There are many softwares to edit the metadata of a photograph. If the criminal edit the EXIF header and upload the photo, then there will be no way to cross check the camera signature and the one embedded in the image. So this cannot be suggested as a solution to identify tampered image. This method is used when source camera of posted image is detected. Since it is possible to edit header details, sometimes it may give false output also.

Vijayalakshmi S [7] proposes and evaluates a new framework for website image manipulation and its protection using ASP.NET framework in watermarking technique. The proposed framework shows how to protect the valuable website images from the unauthorized users and to preserve web server’s bandwidth by preventing other unauthorized websites from directly linking to the images of the authorized web server. This framework effectively preserves the quality of the website image by embedding JPEG file metadata as EXIF tags in website images. Uploading the entire metadata is a tedious task lies as a big disadvantage of this method.

In [15] Eric Kee explains that it is often desirable to determine if an image has been modified in any way from its original recording. A camera signature is extracted from a JPEG image consisting of information about quantization tables, Huffman codes, thumbnails, and EXIF format. The camera signatures are simple to extract and offer an efficient method to establish the authenticity of a digital image. The power of our forensic analysis lies in the ability to acquire signatures from a wide variety of cameras and cellphones. This poses significant challenges as new cameras and cellphones are constantly released.

In all the existing system the absence of metadata and the availability of content editing tool lies as a challenge in the field of digital investigation. This paper is focusing on the introduction of the new mechanism while uploading an image into the social networking sites rather than the recovery of the unavailable information. Extraction of metadata is only possible from the remote server and this process is very difficult and challenging. It had pointed out that once we upload an image into a social networking site like facebook, twitter etc it will extract the metadata and keep it in the remote server. The unavailability of important credentials like IP address or MAC address questions the genuinity of the photo. For the analysis of images in the social networking sites it is inevitable to have an evidence to prevent the criminal activities and for the progress of the investigation. Apart from the image data itself, forensic investigators may exploit the rich source of auxiliary digital data, which typically accompanies the image under investigation. Today, the preferred method to organize and store such metadata is specified in the EXIF standard.



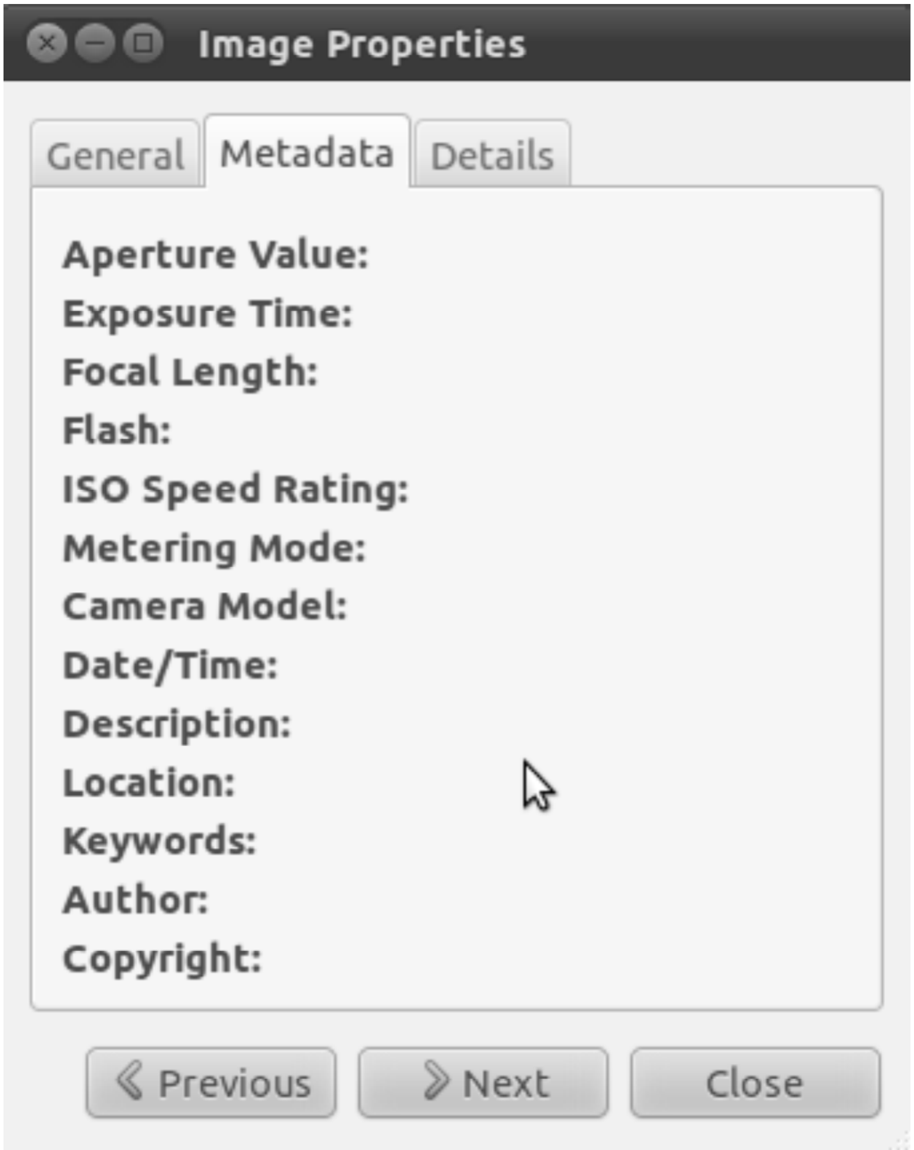
EXIF Data for Burg_Abenberg_Schottenturm.jpg

Camera Maker:	Canon
Camera Model:	Canon EOS 350D DIGITAL
Image Date:	2006-09-06 11:15:33 +0200
Focal Length:	24.0mm
Aperture:	f/10.0
Exposure Time:	0.0040 s (1/250)
ISO equiv:	200
Exposure Bias:	none
Metering Mode:	Partial
Exposure:	program (Auto)
White Balance:	Auto
Flash Fired:	No
Color Space:	sRGB
GPS Coordinate:	49° 11′ 47.20″ N, 10° 46′ 39.00″ E
Photographer:	Christian Eyrich
City:	Abenberg
Country:	Germany
Copyright:	licenced under Creative Commons Attribution ShareAlike 2.0 Germany Licence
Caption:	Schottenturm auf Burg Abenberg, aufgenommen von Süden Aufnahmepunkt: 49° 14' 36,9"N, 10° 57'

Figure 1: Screen shot of Image metadata

Images that are re-digitized from two-dimensional products of an output device like scanners substantially differ from direct projections of real-world. The situations where the genuinity of the image is in question, may occur when it is a screen image or a scanned one. It is impossible to distinguish the photos of this category from the real or original image as people could upload any type of image into the social networking sites. If we analyse any of these photos using any image viewer, to get the metadata, we could only find out the resolution of that image, all other fields go blank. One could extract practically nothing from the server also to answer the problem of genuinity. Another important feature we could find is that the date and time fields of these images get updated when we access it each time. This will overwrite the original information and thereby the origin of the image.

Here it is clear that the fields like accessed date and time and the modified date and time will get update every time when we open that image. Often they will be displayed as same by most of the operating system. That is, in this particular type, a record of the created date or the modified date is not maintained. The criminals take advantage of the lack of the mechanism to solve this problem and this grows as a challenging issue to the forensic investigators. . Although trustworthiness, in particular authenticity is a prerequisite for images being introduced as pieces of evidence to the court, we expect and demand trustworthiness in



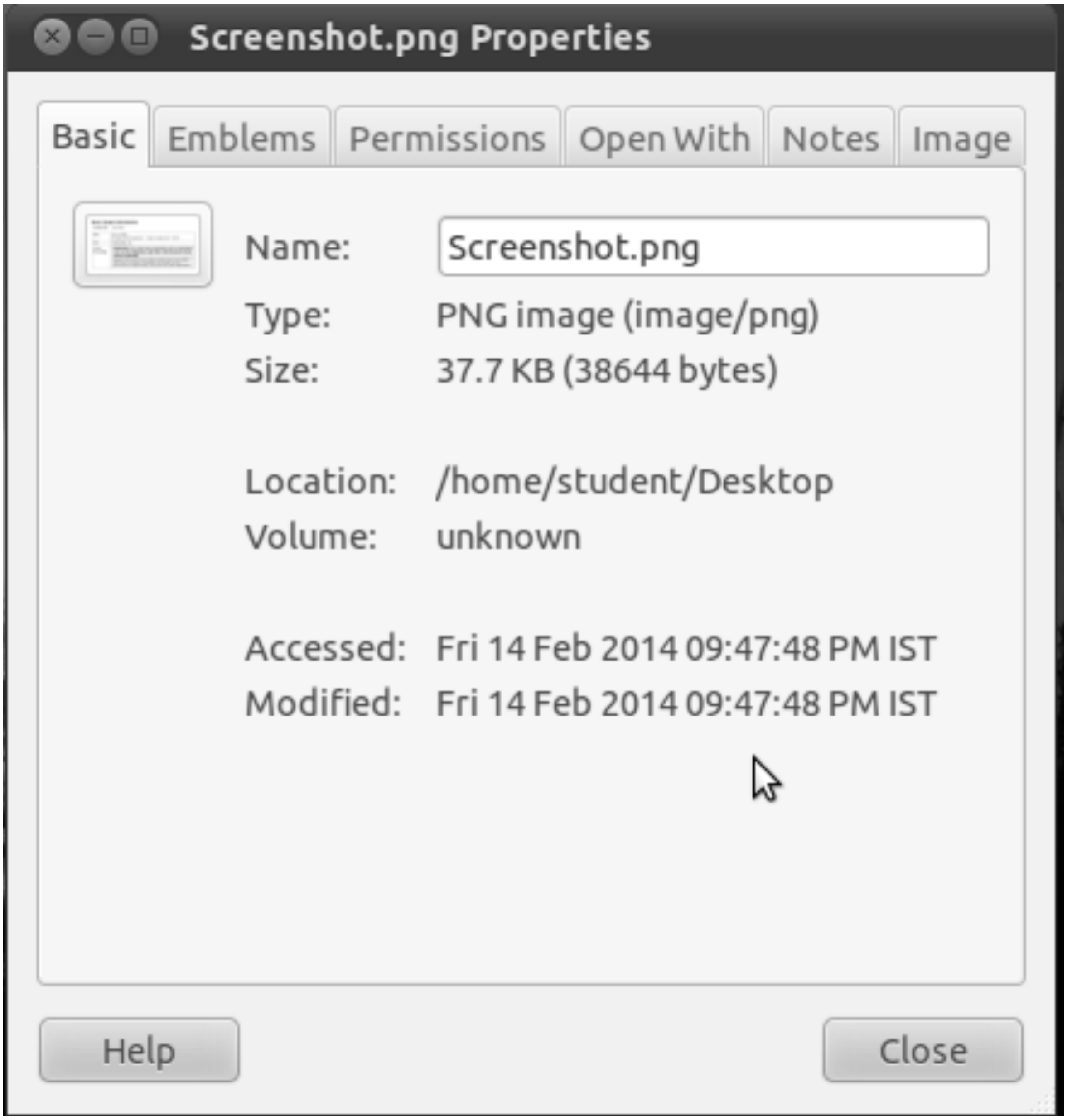
Aperture Value:
Exposure Time:
Focal Length:
Flash:
ISO Speed Rating:
Metering Mode:
Camera Model:
Date/Time:
Description:
Location:
Keywords:
Author:
Copyright:

Figure 2: Metadata of downloaded image

whatever situation we rely upon an image. This is to say that applications of digital image forensics have not only a legal but in general also a very strong social dimension.

When performing a forensic investigation on a computing system, an investigator needs to reconstruct as many events and actions that took place on the system as are necessary to draw reliable conclusions. The main questions a forensic investigator has to answer are: who, what, when, where and why. The "who" question is concerned with who is responsible for certain action on the system and who committed the crime under investigation; "what" addresses the actions that were actually performed on the system; "when" is concerned with the time these actions took place, and "where" question determines where the users were located when they initiated the actions, as well as where are the evidence present on the system.

Metadata of the image is not at all present in the screen shot image or the scanned image. So we could add additional details in order to gain knowledge about the image's origin. To clarify whether an image is real or not, we could add the IP address, MAC address, date of uploading and location as new tags into the image in the encrypted format.



Name: Screenshot.png
Type: PNG image (image/png)
Size: 37.7 KB (38644 bytes)

Location: /home/student/Desktop
Volume: unknown

Accessed: Fri 14 Feb 2014 09:47:48 PM IST
Modified: Fri 14 Feb 2014 09:47:48 PM IST

Figure 3: Properties of print-screen image

IP tag: An IP address is an identifier for a computer or device on a TCP/IP network. Networks using the TCP/IP protocol route messages based on the IP address of the destination. In the field of cyber crime investigation, the discovery of the IP address helps to find out the criminal so soon, since it reveals the network address.

MAC address tag: A media access control address is a unique identifier assigned to network interfaces for communications on the physical network segment. MAC addresses are most often assigned by the manufacturer of a network interface controller (NIC) and are stored in its hardware, such as the card's read-only memory or some other firmware mechanism. The MAC address provide information about the system from which the criminal committed the crime.

Time tag: Time points out to the server time. Because of this reason even though the person change the settings of the system time, it will not get reflected in the metadata. (((IST stands for what??)))

Location tag: Location gives out the knowledge about the place from which the photo is taken. This is to get the latitude and longitude details via Google map. Even though the metadata is removed by the social networking sites, the tag will preserve the information.

History tag(h-tag): it can be described as a flag which get embedded when one upload an image. Its count keep on increasing if the same image is reloaded and the source is the posted image itself.

The addition of the above 4 tags not only solves the problem of genuinity but also provide information about the criminal without extracting the metadata from the remote server. At the mean time the IP address give reference to the server log and one could obtain the detailed image metadata which is helpful for the future investigation. These details will be stored in compressed format and of this reason, even though the metadata is detached, these information remain in the system. LZW compresson algorithm, which is adaptive in nature provides a better privacy of user information and to preserve the metadata details for investigation purpose.

An art of steganography called LSB super imposing to produce the stego-image which is core evidence that store the system information.

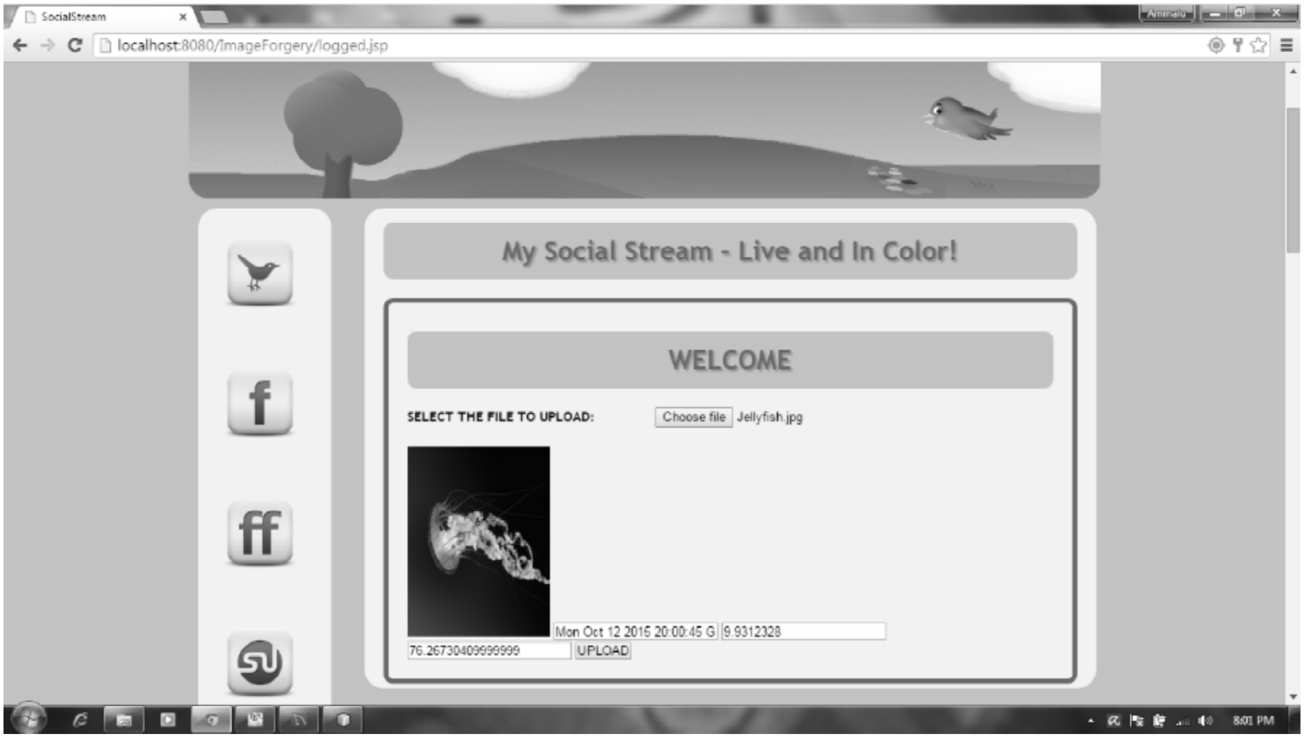
4. IMPLEMENTATION AND ANALYSIS

Problem of genuinity can be solved and the embedding of essential information can be done by the proposed method. The time and location information helps to identify the correct position from which the image uploaded and the time also.

The capturing of the IP, MAC, geo info and timestamp is performed when we upload the image. Proposed architecture performs this task with high efficiency and less time complexity.When we consider the pixel structure of an image we could find that the least-significant bits are the free bits and so the LSB embedding methodology is much more effective . Since the VGA of the posted images are large the image hold high capacity and thus large amount of information can be embedded. Privacy issues can be arise when these captured details get leak out. So in order to provide security compression of embedded facts are performed. Lempel–Ziv–Welch (LZW) is a compression algorithm which is lossless and adaptive in nature. Simplicity in implementation makes it easy to work in any environment.

The downloaded stego-image contains the *h*tag, IP address, MAC address, GPS location and the time stamp. By decoding the LZW compressed image followed by reverse LSB embedding will reveal the system information. The presence of h-tag helps to find the number of times a particular image had uploaded into the social networking site. It could also give details about the people through which an image travelled.

Google API helps to retrieve the location details like latitude and longitude. This helps to plot the approximate or somewhat exact location of the culprit and make further investigation easier. So geo-tag can be described as the factor which helps to make a primary step in the investigation..



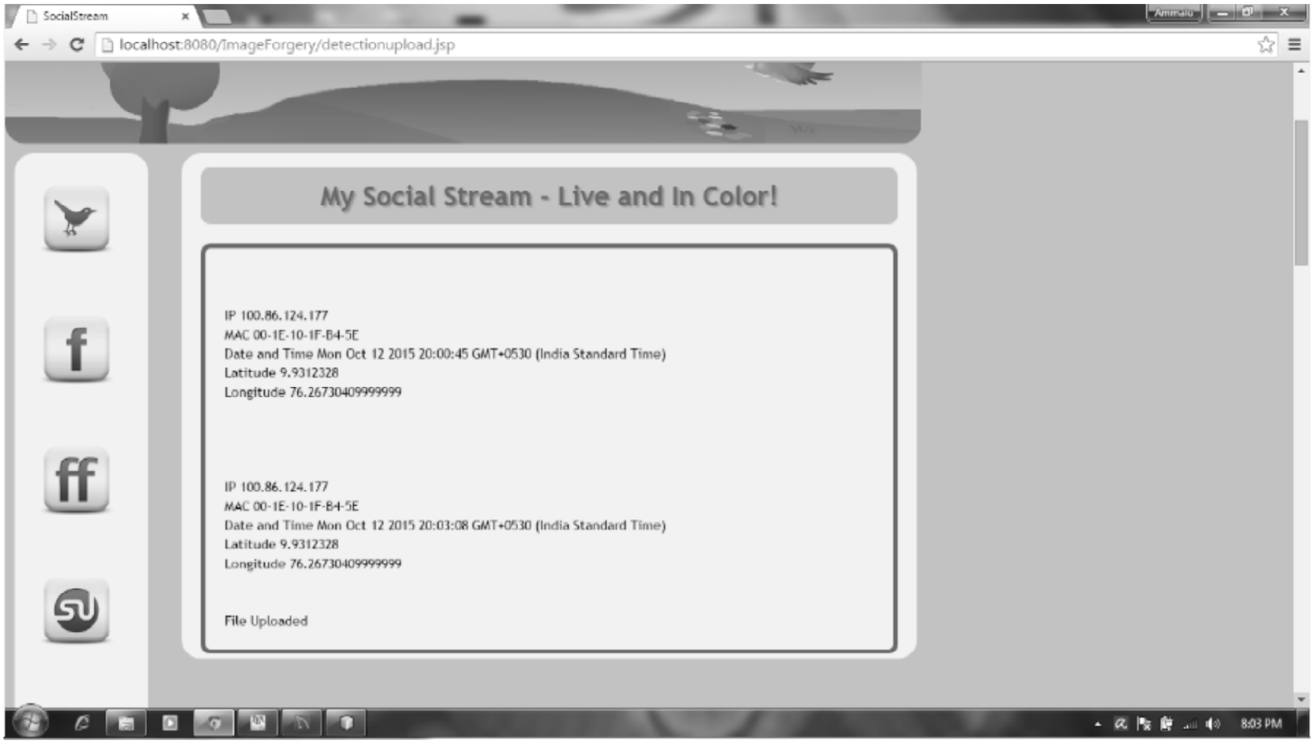
My Social Stream - Live and In Color!

WELCOME

SELECT THE FILE TO UPLOAD: Choose file Jellyfish.jpg

Mon Oct 12 2015 20:00:45 G 9.9312328
76.26730409999999 UPLOAD

Figure 4:Uploading image



My Social Stream - Live and In Color!

```
IP 100.86.124.177
MAC 00-1E-10-1F-B4-5E
Date and Time Mon Oct 12 2015 20:00:45 GMT+0530 (India Standard Time)
Latitude 9.9312328
Longitude 76.26730409999999


IP 100.86.124.177
MAC 00-1E-10-1F-B4-5E
Date and Time Mon Oct 12 2015 20:03:08 GMT+0530 (India Standard Time)
Latitude 9.9312328
Longitude 76.26730409999999


File Uploaded
```

Figure 5: Recovery of Metadata

5. CONCLUSION

In the present scenario, Image processing experts can easily access and modify the image content and thereby it's meaning, without leaving any traces of tampering. The proposed system is capable to provide a substantial evidence in the digital image forensics. The trustworthiness of the image has to be protected since the digital images are valued as evidence in court of law.

Mild change in result may produced by google API can be pointed as the simgle limitation in the system. In this paper we introduce a system to manage the compressed data within the image by using LSB embedding technique, which makes the investigation easier for the cyber forensics department.

References

[1] Howden, C.; Lu Liu; ZhiJun Ding; Yongzhao Zhan; Lam, K.P. Moments in Time: A Forensic View of Twitter, IEEE International Conference on and IEEE Cyber, Publication Year: 2013, Page(s): 899-908.

[2] Saari, E.; Jantan, A. E-Cyborg: The cybercrime evidence finder, Information Technology in Asia (CITA) 2013, 8th International Conference, Publication Year: 2013, Page(s): 1-6.

[3] Dwyer, C.; Hiltz, S.R.; Widmeyer, George, Understanding Development and Usage of Social Networking Sites: The Social Software Performance Model, Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, Publication Year: 2008, Page(s): 292.

[4] Baca, M.; Cosic, J.; Cosic, Z. Forensic analysis of social networks (case study), Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference on Cyber Security, Publication Year: 2013, Page(s): 219 – 223.

[5] Srivastava, A.; Geethakumari, G., Measuring privacy leaks in Online Social Networks, Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on Network and Security, Publication Year: 2013, Page(s): 2095 – 2100.

[6] Hansen, J.A., Adding privacy and currency to social networking, Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference, Publication Year: 2010, Page(s): 607-612.