Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 4th, 2012

Social Network Forensic Tools

Kelvyn Araujo-Valdez, Rhishi Katoch, Rossana Bua, Wil Haywood, and Steve Kim
Seidenberg School of CSIS, Pace University, New York, NY 10038
{ka63869w, rk48856n, rb38340w, wh29732n}@pace.edu

Abstract

Social media has revolutionized how the World Wide Web is used. Computer users once connected to the Internet through networks for the purpose of engaging in e-commerce, research and technology are now part of an interconnected web of relationships that encourage and embrace their commonalities. However, not all users who sign on to the Internet to join social media portals such as Facebook interact with the best intentions. The tremendous increase in Facebook's users also mean increases in the numbers of users who take advantage of the anonymity of the web to commit criminal activity. By employing newer technologies like facial recognition software, law enforcement and government agencies will be able to place a face behind every profile and the public will realize a new level of safety on the Internet. However, the proposed use of this technology is not without controversy. Our research is dependent upon mining for data that is publicly available online; a method that will cause concern for privacy advocates.

1. Introduction

The recent popularity of Social Networking has provided a platform for individual users of the Internet to become connected in ways that they had been unable to during the first stage of the World Wide Web. It is a growing community of millions of people related and unrelated, making connections through their varied interests; which is shared with their family, friends and the general public.

The nature of social media networking allows users to post personal data and content into a personal profile – all linked to that user's interests. Most notable of all social media outlets is Facebook. Boasting an estimated eight hundred million active participants, Facebook has played a large role in keeping more computer users connected than any of its competitors. The networking site is available in more than seventy languages, and seventy five percent of all of its users reside outside of the United States. To reach many of their peers, users download and install applications more than twenty million times each day. Monthly, more than five hundred million people use applications on Facebook or interact with the Facebook interface on other web sites [7].

The growth of social media and the interactive nature of Facebook meant it is inevitable that some users would take advantage of the availability of this technology to conduct criminal activity. As indicated in the public record, the growth of social media use has kept pace with the rise in the amount of crime that is committed over the Internet [2]. To the benefit of law enforcement, forensic tools that have their roots in technology are being employed for the purpose of identifying those who use technology as part of their process to commit criminal activity over the web.

We will look at research conducted by a team at Carnegie Mellon University's Heinz College. The research team was able to successfully identify a sampling of the population on Facebook with publicly available and offline data. Our focus is placed on using face recognition software to provide accuracy in identifying the members of Facebook. We seek to prove that this model works by using this software to identify a small sample of the population and to gather other sensitive data to identify the anonymous.

2. Types of Internet Crime

Cybercrime

There is a pressing need to develop effective tools to combat crime on Facebook and the Internet. Cybercrime involves the action of committing crime using a network and computer. The anonymous nature of the web creates protection for those who develop viruses, keystroke loggers, rootkits and other types of malicious software and those who prey on minors. The internet makes it possible for its users to assume other identities on social media sites. This type of malware often results in an anonymous user taking control of a Facebook account to send spam that directs the account owner to sites set up for identity theft. Such intrusions also expose businesses to theft. By overlooking the need to protect sites such as Facebook and other social media, it can leave companies vulnerable against breaches in security. Corporate servers are being threatened by security breaches that occur through the use of social media.

Phishing

Security group SpamTitan has discovered that seventy percent of companies believe their organization has fallen victim to a spear phishing attack. Spear phishing occurs when an email is sent to one person or several people at a particular company which appears to originate from a person of authority at the same company [15].

Attackers use Facebook to obtain users' information related to their place of work such as email address, phone numbers, physical address and personal information. Then hackers take this info to lure users to websites that look just like their company's website. This allows the attacker to gain important information from the user.

Internet Fraud

Facebook is currently attracting approximately fifteen million users between the ages of 13-17 [3]. These are typically the most vulnerable users of social media due to their lack of knowledge on basic security protocols. Users who conceal their identity have used the internet and social media to commit crimes against minors. Reports have listed crimes involving Facebook as having increased by almost three hundred fifty percent from April 2009 to April 2010 [5] [11].

Such criminal activity costs municipalities, corporation and law enforcement agencies billions of dollars each year [9]. As the "social network du jour," Facebook is a major attraction for those seeking to compromise its network for the purpose of committing cybercrime [5].

3. Privacy and Security

Privacy Related Threats

On Facebook, as with most social networking sites, user profiles can be downloaded and stored over time and incrementally by third parties, creating a digital dossier of personal data. This information can then be used for purposes much different than that intended by the profile owner.

A common vulnerability is that more private attributes which are directly accessible by profile browsing can be accessed by searching by a person's name, phone number or address and is searched on MySpace, Facebook and others, unless default privacy settings are changed and blocked by the user [16].

Risks

The process of interacting online brings daily inherent risk. We disclose personal data knowingly which is viewed on your profile. The creators and network administrators of social networking sites like Facebook have access to more information than is apparent. Data such as time and length of connection, location IP address of connections, profiles visited and messages sent back and forth is already available. Although not just specific to social network services, a users' information is protected to a certain degree with emails, websites, instant messaging systems and internet access points [16].

Information sharing policy

In February 2011 Facebook announced their intention to move forward with a controversial plan to provide third-party developers and external websites access to user's personal information. The information included access to their users' home addresses and phone numbers.

This new feature is also supposed to come with an improved notification and protection for minors. Experts say that this feature could imperil users personal data and increase their risk of being targeted by scams, spam and identity theft.

Although Facebook prohibits applications from selling user's information or sharing it with advertisers and data brokers, spreading phishing scams are not uncommon on the social network. With just a few clicks, an unsuspecting user could potentially hand over their personal information to a scammer who is advertising a new weight loss program or free iPads in an effort to compile credit card data and personal information [16].

4. Facial Recognition Software

The recent highly publicized acquisitions and licensing of companies that produce facial recognition software (FRS) is an indicator of the importance of this technology in the Web 2.0 realm. Facebook's licensing of face.com, a Tel-Aviv-based company that developed software for recognition, is in place to sit at the forefront of future advances. The company was created in 2009 and released two applications that could be used on Facebook for facial recognition. *Photo Finder* searches all images of users as well as their friends and tags those pictures. The second is *Photo Tagger*, which searches many images and bulk-tags faces that appear in multiple photos that are uploaded [12].

The API program began as an Alpha version that allowed third party developers to integrate the Face.com technology, algorithms and databases with their own applications and services. There are some limits to what developers can access but for the most part they can tag and recognize all users from Facebook and Twitter with their own index of images. In February 2011, the program moved from its Alpha to Beta stage, which increased the rate limit of two hundred scanned photos per hour to five thousand photos per hour and is still free of charge. In 2011, the University of Massachusetts conducted a study on the Face.com algorithm and found that it is the most accurate and effective program they have tested [12].

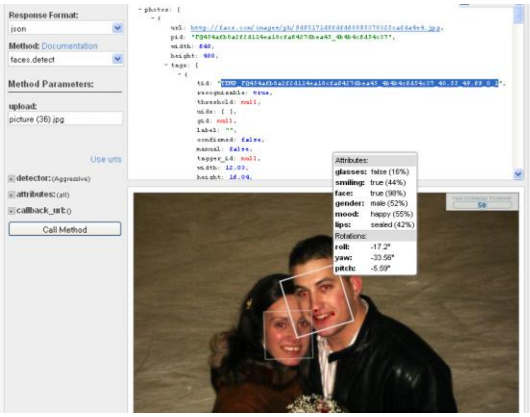
The research conducted by CMU professor Alessandro Acquisti and his team centered around combining publicly available data with face recognition software to re-identify people who are online and offline; and to link online potentially sensitive data to someone's face offline [9]. They ran three different experiments to research the feasibility of combining public online social network data with facial recognition software to find an individual online and off with an image of their face.

The first experiment they searched public profile images from online social networks and compared them to dating sites to find overlaps. This determined just how accurate the face recognizer was in identifying people. This test was conducted with one primary image but since attackers can use different faces, more images have to be tested to increase the ratio of recognition. In their second experiment the researchers took profile images of consenting students walking on campus and compared those to profiles on Facebook. Participants provided information about their personal life through an online survey and were asked to identify images in which they recognized themselves. This resulted in one out of three subjects identified. In their final experiment, the researchers wondered just how much you can learn about a person from their photo.
They gathered personal information not previously known from a face by training an algorithm to id profile owners between experiment two photos and a database of Facebook images. [1]

5. Data

We began our process by using existing data from members of our team. Each individual had online profiles on Facebook. We were required to train the facial recognition software (FRS) to recognize our faces on an image, tag them and then train the software to update our index [8]. We accomplished this by calling on a number of methods provided by the API, including face.detect, tags.save and face.train.

First, we select an application built from the API, request a format (we can use JavaScript object notation or xml), and call the face detect method as you see in Figure 1 below. After the program asks for an image, we call the method and the API presents us with an outline of each face it detects in an image. If we want to save the image and correctly identify the person we need to copy the TID number which is highlighted in Figure 1 as well and will be used in the next step under tags.save.


Response Format: json

Method: Documentation faces.detect

Method Parameters:

upload: picture (36).jpg

```
- photos: [
  - {
       url: http://face.com/images/ph/9495171d9fdfd68895370325cafda4e4.jpg,
       pid: "F@454afb6a2fd4114ea16cfa6427dbea45_4b4b4c6d54c37",
       width: 640,
       height: 480,
     - tags: [
        - {
            tid: "TEMP_F@454afb6a2fd4114ea16cfa6427dbea45_4b4b4c6d54c37_40.55_49.69_0_1",
            recognizable: true,
            threshold: null,
            uids: [ ],
            gid: null,
            label: "",
            confirmed: false,
            manual: false,
            tagger_id: null,
            width: 12.03,
            height: 16.04,
```

Attributes:

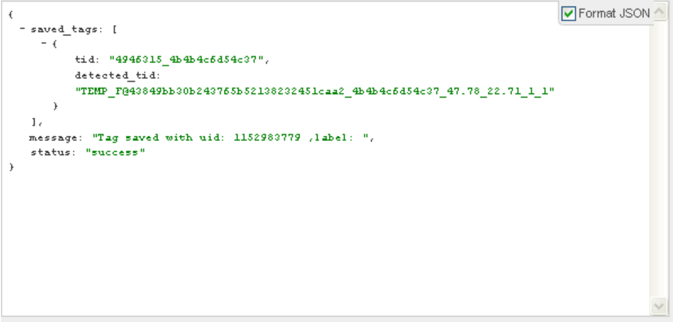
glasses: false (16%)
smiling: true (44%)
face: true (98%)
gender: male (52%)
mood: happy (55%)
lips: sealed (42%)

Rotations:

roll: -17.2°
yaw: -33.56°
pitch: -5.59°

Figure 1: Detecting and saving images

Since we want to identify a specific person, we grab identification information from the code that relates to that person. Switch the method to tags.save and enter the TID's number into the box and find the user id of the person you are trying to identify. Once you have all the correct information you call the method and the API informs the user that a tag has been saved, and is ready for identification as you can see in Figure 2.

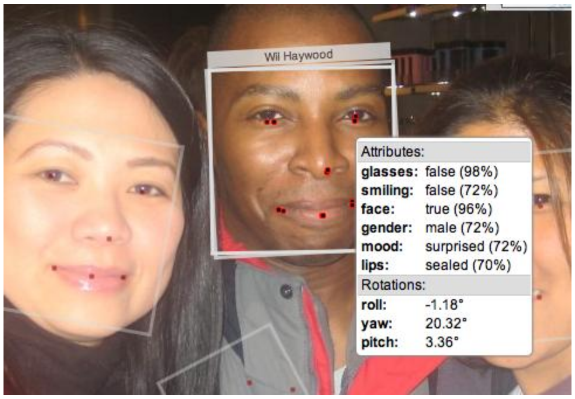

☑ Format JSON

```
{
 - saved_tags: [
    - {
          tid: "4946315_4b4b4c6d54c37",
          detected_tid:
          "TEMP_F@43849bb30b243765b52138232451caa2_4b4b4c6d54c37_47.78_22.71_1_1"
      }
   ],
   message: "Tag saved with uid: 1152983779 ,label: ",
   status: "success"
}
```

Figure 2: Success of tags.save complete

When we call the face.train method, we have an opportunity to name an individual. A sign in is required for access to names in a profile. Providing a username for a particular face, we are then able to train the API to recognize a user of Facebook.

From the image below in Figure 3, we can see that the API identified team member Wil Haywood. The program detects five points on a face with different attributes. Glasses, smiles, gender, mood and lips are given percentages of identification through use of arrays in the code.



Attributes:
glasses: false (98%)
smiling: false (72%)
face: true (96%)
gender: male (72%)
mood: surprised (72%)
lips: sealed (70%)
Rotations:
roll: -1.18°
yaw: 20.32°
pitch: 3.36°

Figure 3: Attributes found when image was scanned

The next photo we can see that the program was able to successfully identify Kelvyn Araujo as one of the photos found on his profile. The person next to him on the image above was identified as unknown because no pictures of this person exist on any of Araujo's Facebook profiles or friend profiles. This poses a problem as far as naming the individual, but they can still be identified and alternate images can be found through Facebook.

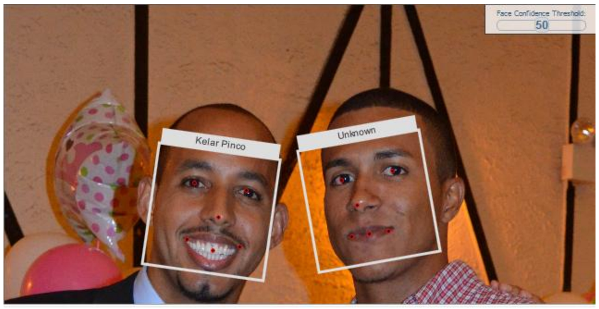


Figure 4: Once recognized name will show on face

After the initial experiment, we turn our focus on how accurate the program was on identifying a person from within a group of people. This experiment demonstrated that a person can be identified by the algorithm regardless of how many people are in the picture. This would allow a detective to locate a person by pinpointing the person among the billions of pictures posted on Facebook and a large amount of people within multiple pictures.

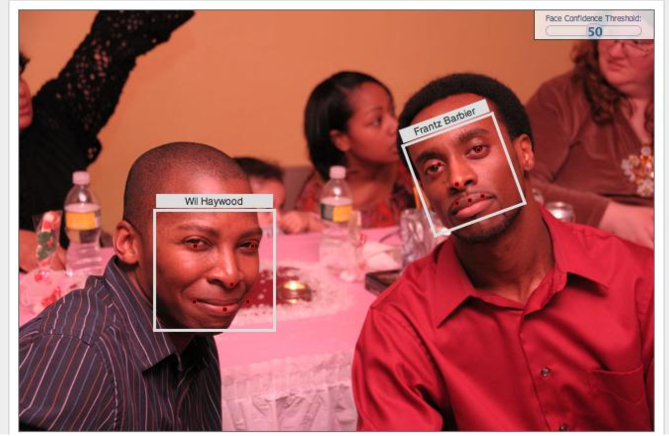


Figure 5: Can identify many faces in one image

In Figure 5 the program scan's many images from our team member's profile of images he had trained with the system, once he ran the facebook.get method it was able to identify both faces in the image.

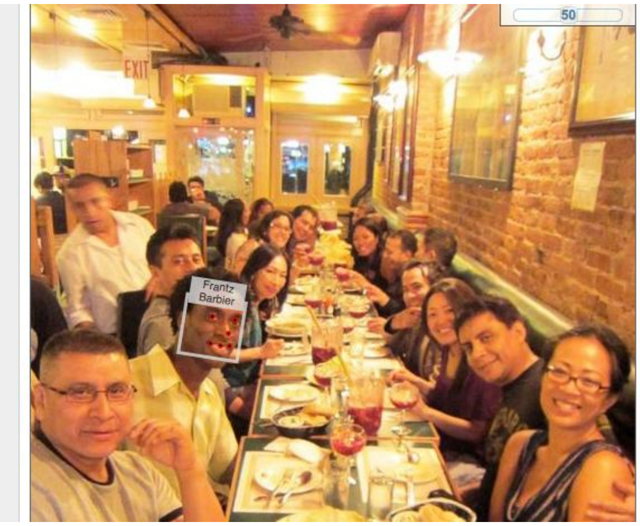
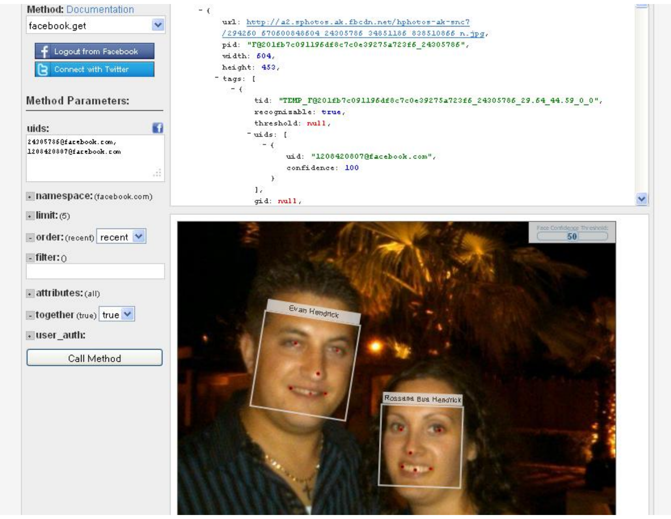


Figure 6: Scanner was able to identify in large group

The image in Figure 6 was able to find the person he trained in the previous image from a large group of people by scanning many images and it correctly found all of images of his friend.

An example of FRS in action would involve its use to track down cyber bullies of a 12-year-old boy who is targeted by someone he identifies as one of his classmates. We can use face.com technology to find any and all pictures on the victim's profile that has been tagged which will later point to who tagged the picture. We can also identify the suspect by checking his or her profile for any tagged pictures of the victim.

In Figure 7 our team member is using multiple Facebook user IDs to find and identify herself in images in which she is together with her husband. She has added the different user id's for each person, the amount of pictures to return and setting the parameter to images that are together to being true.



Method: Documentation
facebook.get

Method Parameters:

uids:
24305786@facebook.com,
1208420807@facebook.com

namespace: (facebook.com)
limit: (5)
order: (recent) recent
filter: ()
attributes: (all)
together (true) true
user_auth:

```
- {
    url: http://a2.sphotos.ak.fbcdn.net/hphotos-ak-snc7
    /294260_670600848604_24305786_34851186_838510866_n.jpg,
    pid: "F@201fb7c091196df8c7c0e39275a723f6_24305786",
    width: 604,
    height: 453,
  - tags: [
     - {
          tid: "TEMP_F@201fb7c091196df8c7c0e39275a723f6_24305786_29.64_44.59_0_0",
          recognizable: true,
          threshold: null,
        - uids: [
           - {
                uid: "1208420807@facebook.com",
                confidence: 100
             }
          ],
          gid: null,
```

Figure 7: Settings for facebook.get

Once we call the method to run the program it brought back five images with both Rossana and her husband in each image which is shown below in Figure 8. Since we set the parameter to true for images together it only brought back pictures that included both user id's that was entered.

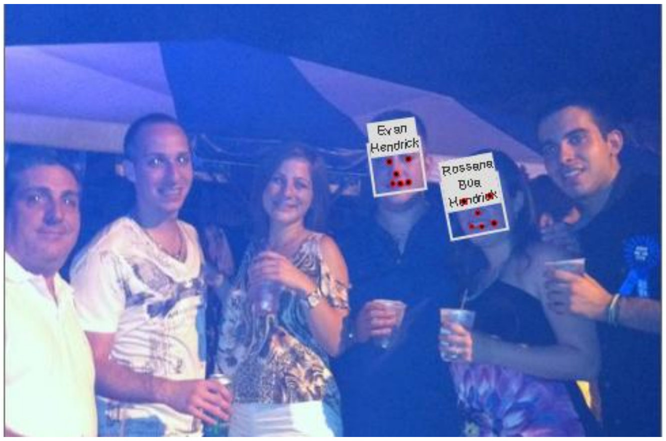


Figure 8: Picked out the correct faces once scanned

We noted that all of these features get us closer and closer to identifying someone on the other side of the spectrum; however the question arises what happens if the suspect is using false id's and pictures. With this scenario alternate technology may be needed (i.e. IP address tracking technology) in order to identify the suspect as the real criminal in front of the PC across a given network.

If we take all images of a high risk suspects such as sex offenders in a database and train the software to ID each, then you can scan images of profiles to find sex offender online.

If we take all images of a high risk suspects such as sex offenders in a database and train the software to ID each, then you can scan images of profiles to find sex offender online. If the program brings up a match then additional checks are necessary to ascertain that they are not using the profile for illegal activity since most sex offenders are banned from social network sites. The image in Figure 9 below shows the process of face recognition and how it will be found.

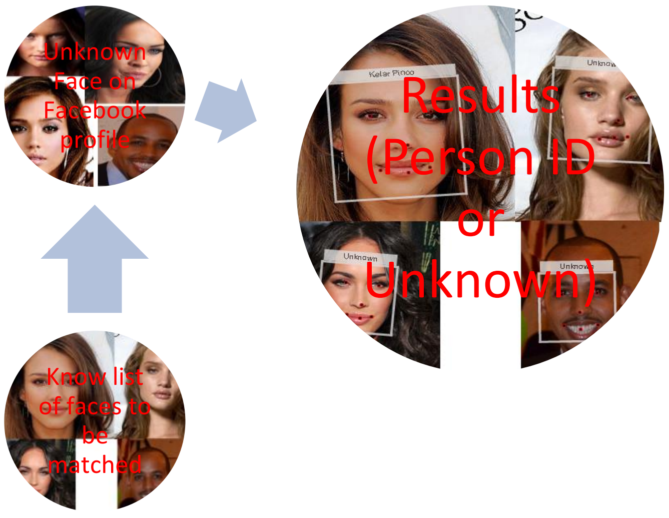

Unknown Face on Facebook profile
Results (Person ID or Unknown)
Know list of faces to be matched

Figure 9: Process of face recognition

6. Implications

There are significant implications for the successful use of FRS.

A future can be conceived where FRS is adopted for a number of domestic uses. With the convergence of FRS and databases, municipalities may be able to identify drivers who repeatedly ignore traffic signals. A signal beamed at the driver's side of a speeding car will capture the image of the driver. This will then be relayed to a database. It is also possible that the need to carry various forms of payment may be replaced due to this technology. Much like iris scan technology that is being considered by the U.S. Department of Homeland Security to replace a fingerprinting database, the FRS would lead to greater efficiency in government process; but this will also raise questions regarding civil liberties and the potential for abuse [10].

The reliability of this technology has garnered the attention of the U.S. Federal government. The U.S. Army has already employed this technology in current overseas conflicts, collecting biometric data from Iraqi and Afghan civilians at checkpoints, workplaces, sites of attacks and door-to-door canvassing [6]. Based on the recommendation of Senator Jay Rockefeller, chair of the Senate Commerce, Science and Transportation Committee, the Federal Trade Commission has recently announced plans to host a workshop in Washington DC in Dec. 2011 to explore this technology and the privacy and security questions raised by its increasing implementation. "While there may be great potential for commercial, personal, and law enforcement users of this emerging technology, its development also raises numerous questions about individual privacy [14]." In announcing the workshop, the FTC has acknowledged the adoption of FRS by "social media networks to digital signs and mobile applications [14]." It is this concern for the violation of basic privacy rights that will be a source of controversy in the use of this technology.

7. Conclusion

The future points to increased use of facial recognition applications such as the one described in this experiment. There will currently numerous companies that are investing in this technology to allow their users access to identify friends and family members. Law enforcement is aware that this technology will see more use in putting a name to the face of, bullies or sex offenders. There are many options for the use of this program and it is not the only one available in the World Wide Web. Of the programs tested, we believe this was the best program that could match images on facebook by uploaded an image you already have and find others throughout facebook.

The nature of social media dictates that there will always be security risks associated with using a portal like Facebook. Using face.com may prevent or minimize incidents to improve security without compromising the benefits of information sharing - thereby increasing the overall social value of social networking sites. Web safety cannot be guaranteed, but as facial recognition technology improves, it will become embraced by law enforcement as an effective tool against cybercrime. However questions regarding privacy will undoubtedly remain. The challenge for society will be to find a way to embrace the technological advance that is FRS, while protecting the rights of its citizenry.

References

[1]A. Acquisti, R. Gross, F. Stutzman ”Faces of Facebook: Privacy in the Age of Augmented Reality http://www.heinz.cmu.edu/~acquisti/face-recognition-study-FAQ/acquisti-faces-BLACKHAT-draft.pdf, August 2011

[2] T. Bradley, “Cybercrime: A Recession-Proof Growth Industry” PC World Business Center http://www.pcworld.com/businesscenter/article/218850/cybercrime_a_recessionproof_growth_industry.html, 2011

[3] K. Burbary, “Web Business” http://www.kenburbary.com/2011/03/facebook-demographics-revisited-2011-statistics-2, 2011

[4] A. Burke, D. Durve, M. Marks, S. Cha, D. Athanasopoulos, “Forensic Evidence Management Information Systems (FEMIS)” *Proceedings of Student-Faculty Research Day, CSIS, Pace University* http://csis.pace.edu/~ctappert/srd2010/c5.pdf, 2010

[5] Daily Mail Reporter “Number of Crimes Involving Facebook ‘leaps 346% in a year.’ http://www.dailymail.co.uk/news/article-1263021/Number-crimes-involving-Facebook-leaps-346-cent-year.html, 2009

[6] Electronic Privacy Information Center, http://epic.org/privacy/facerecognition, 2011

[7] Facebook Statistics http://www.facebook.com/press/info.php?statistics, 2011

[8] Face Recognition How-To http://developers.face.com/docs/recognition-howto, 2011

[9] J. Finkel “Cybercrime Spreads on Facebook” *Reuters* http://www.reuters.com/article/2009/06/29/us-facebook-security-analysis-idUSTRE55S55820090629, 2009

[10] T. Frank, “Homeland Security to Test Iris Scanners” *USA Today* http://www.usatoday.com/tech/news/surveillance/2010-09-13-1Airis13_ST_N.htm, 2010

[11] B. Kraemer, “Facebook User Accused of Sexual Assault” *CRN* http://www.crn.com/blogs-op-ed/the-channel-wire/213201918/facebook-user-accused-of-sexual-assault.htm, 2009

[12] Labeled Faces in the Wild http://vis-www.cs.umass.edu/lfw/results.html, October, 2011

[13] “Researchers Show Power of Facebook Facial Recognition Software”http://latimesblogs.latimes.com/technology/2011/08/facebook-photos-facial-recognition-puts-names-to-faces-at-black-hat-conference.html, 2011

[14] B. Sasso, “Rockefeller asks FTC for report on privacy implications of Facial Recognition Technology” http://thehill.com/blogs/hillicon-valley/technology/188527-rockefeller-requests-report-on-facial-recognition-technology, 2011

[15] Spam Titan http://www.spamtitan.com/core/news.php?NewsID=432&NewsListPage=2, 2011

[16] B. Bosker, “Facebook To Share Users' Home Addresses, Phone Numbers With External Sites” http://www.huffingtonpost.com/2011/02/28/facebook-home-addresses-phone-numbers_n_829459.html, 2011