Children and Geotagged Images: Quantitative Analysis for Security Risk Assessment

Joanne Kuzma
Faculty of Computing, University of Worcester
Henwick Grove, Worcester UK WR26AJ
Phone 01905 542023
j.kuzma@worc.ac.uk

BIOGRAPHY

Dr. Joanne M. Kuzma earned her Ph.D. in Information Systems at Nova Southeastern University in Fort Lauderdale, Florida and has over 20 years of computer consulting background along with university teaching experience. Currently, she is a Senior Lecturer in Computing for the Worcester Business School at the University of Worcester, Worcester, UK.

Abstract

This paper investigates the levels of geocoding images with children pictures, and discusses privacy and safety issues that may affect children. This study analyzed the number of geocoded images of children's pictures on Flickr, a popular image-sharing site. For 50 of the top most expensive residential zip codes in the U.S., the number of images that had gelocation tags was counted. Results showed significant number of images with children's faces that had geotagged information. The location information could possibly be used to locate a child's home or other location based on information publicly available on Flickr. Publishing geolocation data raises concerns about privacy and security of children when such personalized information is available to Internet users who may dubious reasons for accessing this data. People should understand the implications of this technology and post only appropriate data in order to protect themselves and their children.

Keywords Flickr, geocoding, geolocation, geotags, privacy, security

1. Introduction

The rise of image and video web hosting service sites such as Flickr and Photobucket has allowed millions of worldwide users the ability to upload, store and organize their digital images. Most sites allow the ability to set privacy controls to regulate private or public image viewing. Privacy controls allow images to be accessed by group members, such as friends or family, while non-members who have Internet connections can view images tagged as 'public'. Yu (et al, 2009) indicates that photo-sharing sites have attracted millions of people, whose common area of interest revolves around image and video topics, thus promoting social interactions among the globe.

Along with image sharing, many sites enable the publication of geo-located content in real-time (Vicente, et al., 2011) by use of geotagging. Geotagging is the process of storing longitude and latitude data within an image (Vamosi, 2010). This functionality allows other Internet users to use geographic positioning systems (GPS) and a variety of Internet mapping sites to determine the physical location of where the picture was taken. However, like many other technologies, geolocation has dangers along with its beneficial functionality. Privacy advocates warn that when a user posts a geotagged resource that location can be easily accessed by millions of Internet users to determine the location of the person posting the

resource (Vicente, et al, 2011). Prior studies have reviewed various aspects of geolocation privacy and security issues, but no studies so far have concentrated on the safety aspects related to children.

There were two major aims addressed in this study:
1. Determine the range of children's images posted with geolocation tags among 50 high-income US zip codes
2. Review security and privacy issues that could be raised with geocoding and suggest methods for better protection

The study starts with a literature review of geocoding issues and prior studies, followed by a methodology and results discussion. Finally, this study highlights privacy implications that users should consider when posting pictures of children to Web sites.

2. Literature Review

Geotagging and the growth of image sharing sites

The popularity of image sharing sites is growing rapidly, especially because of the increased use of location-enabled devices like mobile phones and iPads (Ozer, et al, 2010). The author quotes a study that found that one in four US adults had use a location based service, and two-thirds of iPhone users access these services at least once a week.

Technical advances in photographic equipment and software to enable geotagging functionality into images and video. Images and video can use 'tags' which can contain geospatial information for general mapping purposes, and online applications, such as Google Maps, has ensured that online mapping has become commonplace (Burton, 2010). Joshi (2008) indicates that photo-sharing sites such as Flickr have realized the need to tap into geolocation functionality to provide more service options for their users. For example, Flickr allows users to provide geolocation information or coordinates for their pictures with the help of a map interface. Besides gelocation tags, users can also enter a plethora of other user annotation information, which could also be used to infer the location of pictures with good accuracy (Joshi, et al, 2010).

Legal Protection

Some privacy and security legislation exists that offers limited protection to online consumers, but the strength of the laws is piecemeal and depends on a myriad of legislative factors depending upon the locality and region. Burton (2010) contends that information privacy laws, which were founded on the notions of fair information practices and principles, may have a limited impact regarding the privacy problems arising from invasive geolocation technologies. He argues that the historical development of first generation privacy laws highlighted the collection, storage and use of personal information by data collecting organizations that use this information for business or statutory purposes. However, new technologies, such as Web 2.0, have distorted this balance because individuals have now entered into the domain of data collection and use. The author indicates that individuals now have the same capacity to cause privacy and security problems to online users that were once the exclusive domain of organizations. At this point, most legislation has done little to address problems arising from actions of individual users.

A 2010 survey of 1,600 social networking users who owned geolocation-capable devices found that over half were worried about loss of privacy due to geotagging ability. Of the UK respondents, 52 percent admitted they tagged their whereabouts in a photograph online (Douglas, 2010). Ozer (et al, 2010) indicates that 55 percent of consumers using location based services are very concerned about loss and risks to their privacy. However, those risks are exacerbated by outdated laws that do not adequately address technology issues raised by these service sites. The author states that privacy protection laws have not yet accounted for the fact that services and technologies can reveal intimate details and locations about a person's life. Many of the existing online privacy statutes were written over a decade ago when social networks and image sharing sites. Ozer (et al, 2010) states that there are many complexities with geolocation technology that shapes future legislation. For example, does the information transmitted to a location-based service without the device-owners informed consent fall under the domain of the service provider? If a person takes a picture with a child's image in the background (say on vacation) and posts it to a site, should children in the picture have to give informed consent? The number and complexity of these types of questions contributes to the slow pace of the legal process (Ozer, et al, 2010).

Although no legislation now exists that addresses the technology issues of geolocation privacy and security, some U.S. legislators are starting to address the problem. In 2010, Congressman Bobby Rush introduced a privacy protection bill, "Best Practices Act of 2010" which address "precise geolocation information and any information about the individual's activities and relationships associated with such geolocation" (U.S. House, 2010). In late 2010, German politicians called for comprehensive data protection regulations including protection legislation on geotagging (The Local, 2010). Lawmakers in California proposed a set of bills to require sellers of mobile devices with geotagging capabilities to disclose those capabilities to a consumer (Butcher, 2011).

The U.S. Federal Trade Commission FTC has noted the rise in the use of geolocation applications and issues with companies collecting personal data about consumers and sharing it with other entities. They issued a recommendation that companies should seek affirmative express content before they collect this sensitive data (FTC, 2010).

There is a patchwork of global laws that address some forms of consumer's online privacy, but few laws, especially outside the U.S., specifically address legal protection of children when dealing with online activities. In 1998, the U.S. Congress passed the Children's Online Privacy Protection Act (COPPA), which set rules for online collection of information on children (Bowie and Jamal, 2006). The law requires online sites to secure parental consent before collect personal information from children under 13, and forbids release of such information if it has been collected (Bhasin, 2006). However, the law is dated when it comes to newer technologies such as geotagging. Kevin Pomfret, executive director for the Centre for Spatial Law and Policy, indicates that expansion of geotagging and gelocation technology is starting to being addressed in discussion among the U.S. Congress, and within review of the COPPA Act (Lofing, 2010).

Privacy and Security Threats

Vicente (et al., 2011) warns of three types of privacy dangers if one exposes their geolocation: location, absence and co-location privacy. A location threat example would be if a parent takes a picture of a child in their home, and posts the geolocation information along with the image to various Web sites. A stalker could easily use Google maps to fix the

geolocation coordinates to the exact physical location of the person's home. Vicente (et al., 2011) explains that absence privacy could occur if a family immediately posts pictures of them on vacation in a foreign country, and a burglar determines that this family may be gone for an extended period. This would indicate the family house is empty and rife for a burglary. The authors give an example of co-location privacy, where an adversary could gain information about the relationship of the same people in one image.

Ozer (et al, 2010) mentions other situations that could be raised by geotagging images. People reading the location and know that a person in a picture is visiting a doctor's office, shopping at a specific store, interviewing for a new job or engaging in a gun rally. The geolocation can be used to infer regular habits and routines, or deviations from these routines. Problems with online privacy can be especially acute when well-meaning parents or friends upload pictures of children to these sites without understanding the implications of geotagging. In addition, it is possible that total strangers who take pictures of children (for example at a sports game) and may upload pictures of children without permission.

Prior Studies

Prior studies in the area of geolocation and geotagging have shown issues related to privacy and predictions of user behavior. Clements (et al., 2010) performed a study where they collected geotags of 36,242 Flickr users who used geolocation tags. They were successfully able to predict user's favorite locations within a city based on Flickr geotags of other cities. Friedland and Sommer (2010) performed a study on the degree to which image and videos are geotagged. They found 4.3% of images in Flickr are geotagged, 1.3% of images in Craigslist had GPS location and 3% of YouTube videos have geolocation. The authors indicated users are unaware of the full scope of the threats they face when publishing information with gelocation tags.

Vamosi (2010) quotes an example of a site, icanstalu.com, which uses a Perl script to scrape some 20,000 images each day from sites such as Twitter. The script is able to read the geotags and the site then posts information about specific pictures along with supplying information about the longitude, latitude, street location and location on a visual map. Vamosi gives another example of PleaseRobMe.com (now defunct) which also posted personal location data online. The preponderance of these types of sites adds to the risk of security and privacy problems that can be encountered by well-meaning consumers who innocently post pictures of themselves containing geotags containing personal data.

A variety of prior studies do exist which show the usage of geolocation for image and video sharing sites. However, the preponderance of this research is based on generic types of images, and does not divide these into specific categories. No studies at the time of this paper deals with a review of geotagged images relating to children's pictures, thus this study contributes to new originality in this field.

3. Methodology
The research was accomplished through analyzing 50 of the most expensive residential zip code areas within the Flickr site to determine levels of geocoding of children's images. The project consisted of the following phases:
   1. Choosing an image hosting service to test
   2. Choosing a sampling of images within Flickr
   3. Compiling the results

Choosing an image hosting service

The first phase of this project involved choosing an image hosting Web service to test. There are many of these sites, but for the purpose of this research paper, the service needed to provide specific functionality: a) images would contain geolocation tags b) images could be searched by user-defined tags c) application would be connected to mapping function and d) a sizable number of searchable photos (at least 1 million worldwide). The first service reviewed was Photobucket.com, where 23 million users upload 4 million images per day (Photobucket, 2011). Although the images allow geographical mapping, the site requires a login account to use. A second site "Free Image Hosting" was a smaller site and did not have mapping capabilities (Free Image Hosting, 2011). A third option, Flickr, owned by Yahoo!, and possesses over 142,272,905 geotagged items, and has a function allowing users to search images by a variety of user tags and locations (including US zip codes) (Flickr, 2011). Therefore, because of the large number of items and versatility of the service, Flickr was chosen as the image hosting service for this project.

Choosing a sampling of images within Flickr

The first part of this sampling was to determine the 50 neighborhoods with the highest incomes in the United States, using the site "The Higley 1000." Based on this site, residential neighborhoods with the highest incomes were sorted in order of income level. The researcher then determined the zip code for each of the neighborhoods, and this zip code would eventually be used within the Flickr search criteria. The reason high-income residential areas were chosen was twofold. First, photographs from a residential area may be more likely to contain photos where the residents actually live, as opposed to when they are on vacation, thus increasing the likelihood of pictures being near their homes. Second, higher income consumers may have greater access to newer model digital cameras with geolocation functions.

For the list in the Higley site, some neighborhoods needed to be combined because of similar zip codes. For example, the eighth most expensive residential neighborhood on the Higley list was "North Beverly Hills-Peavine Canyon" and number 32 was "Trousdale Estates", both with the same zip code of 90210. Therefore, only the first occurrence of a zip code was used, but in all, 50 of the most expensive neighborhoods were chosen for testing.

Flickr functionality allows someone who has uploaded a photo image to tag the photo with user tags to create user definitions for each of the photos. For example, users could add tags such as 'child', 'school', 'family', etc. in order to better explain the purpose of that picture and to allow Internet searches of that images based on the specified tags.

For this research, it was deemed appropriate to start narrowing the search photos with children by first searching for specific user tags that may be more consistently used with pictures or portraits of children's faces. In this case, the researcher decided to start the search with the terms 'family,' 'face,' 'portrait,' 'child' or 'school.'

Compiling the results
Along with the composite search of the terms most likely to result in pictures of children, the researcher inserted each of the 50 zip codes to run a search of all images for that specific zip code. In all 50 searches, a number of images with and without children's pictures resulted.

Figure 1 shows the results of a typical search. A series of images is displayed, and a user can choose a specific picture, of which Flickr map will then show the exact geographical location of where that picture was taken.

For this paper, the number of children's images was counted manually from the initial results. Only those pictures of children whose approximate age was under 18 years were counted, although sometimes this was an approximation. If there was a question on the age, then the image was not counted. Faces of children had to be clearly visible. Pictures of groups of people were counted as long as there was clearly a picture of at least one child within the group.

4. Results

Table 1 shows the results of the Flick search on geocoded images of children of the top 50 most expensive residential neighborhoods in the US. Column 1 indicates the specific neighborhood, with Columns 2 and 3 showing the associated zip code and average home price for the area. Columns 4 and 5 show the number of images found for the search term 'family or face or portrait or child or school' and the number of children's portraits.

All zip code locations had geotagged images of children found within the search criteria. The average number of images was 1160 with each location containing approximately 116 portraits with a child's face that could be easily geolocated within Flickr map function.

It should be noted that there were several instances where security and privacy were problematic. In one zip, there were over 30 related pictures of a family gathering at a home for Rosh Hashanah. Other pictures showed family gathering in front of homes, at home parties, pictures of new babies in their homes or children's sports events. All of these pictures had geocoded locations that could be superimposed over a map, thus posing issues for security and privacy.

Table 1: Results of Flick Goecoding Searching

Neighborhood	Zip	Avg Price	Images	Children
Holmby Hills	90077	$585,925	207	7
Buell Mansion-Cherry Hills Park	80113	$582,129	128	6
Round Hill-North Greenwich	06831	$541,565	538	194
St. Louis Country Club	63124	$464,046	138	58
Merrywood-Knollwood (Short Hills)	07078	$461,340	57	7
Indian Hill Club-Woodley Road	60093	$459,070	1492	179
Jupiter Island	33455	$448,796	129	6
North Beverly Hills-Peavine Canyon	90210	$440,913	965	20
Midtown	10021	$434,824	7906	145
Belle Harbor	11777	$432,207	56	12

Rumson Country Club-Waterloo	07760	$424,994	398	25
Johns Island	32963	$424,938	243	10
Bonnie Briar Country Club	10538	$412,587	163	11
Everglades Golf Club	33480	$411,652	209	6
The Highlands	98177	$410,500	238	53
Fisher Island	33109	$404,240	8	0
Woodhill Country Club	55391	$402,985	129	12
Murray Hill-Heathcote	10583	$402,565	95	19
Volk Estates-Windsor Place	75205	$396,786	916	409
Orienta	10543	$394,293	586	29
Rockdale Estates-Cutler Bay Estates	33190	$388,671	2	0
Northside Summit	07901	$386,517	136	22
Brentwood Park	90049	$384,140	1331	76
Fremont Place	90005	$381,500	362	1
Hunting Valley, Cleveland	44022	$373,181	140	9
Mockingbird Valley	40207	$372,538	531	186
Co Club at Castle Pines	80108	$372,010	1315	148
Rob Roy-Davenport West	78733	$368,743	81	11
Villanova East	19085	$366,904	591	9
Country Club of New Canaan	06840	$360,440	175	40
C. Portola Valley-Alpine Hills	94028	$358,559	111	25
Devonwood-Bayridge Estates	33156	$354,102	382	6
Southeast Section	60521	$352,306	199	11
Purchase	10577	$350,953	281	111
Washington Park	98112	$350,900	1665	104
Hewlett Bay Park	11557	$347,898	48	28
Chestnut Hill (Estate Section)	19118	$344,651	505	36
Lost Tree Village-Seminole Landing	33408	$343,089	99	44
Ponte Vedra Beach Oceanfront	32082	$340,595	374	73
Westover Hills, Fort Worth	76107	$338,225	1259	107
Country Club of the South	30022	$335,500	352	101
Winnetka Southwest-Crow Island Park	60043	$335,213	33	2
Eastover	28207	$333,412	197	6
Coco Plum-Gables Estates	33143	$333,309	217	63
La Jolla Country Club	92037	$331,500	2404	45
Bronxville East	10708	$331,107	284	171
Bradley Manor-Longwood	20817	$330,211	437	32

Langley	22101	$329,416	453	76
Glencoe Lakefront	60022	$327,820	332	16
Shoal Creek-Stonegate Farms	35242	$326,500	208	47
Average			1160	112

5. Implications and Discussion

This study has shown that for high-income locations within the US, a large number of geotagged images of children's faces were found on Flickr. As the literature review and prior research shows, this sort of personal data could lead to compromised children's security, as people with dubious intentions may be able to ascertain a child's physical location with unfortunate consequences. Protecting children's online privacy is a multifaceted approach, requiring many stakeholders to be involved in mitigating this type of online risk. Online sites who host and display personal information are at the forefront of privacy and security protection, but individual users and governments all need to become more involved.

Burghardt (et al, 2009) suggest that although various privacy mechanisms exist for location based services, it is unclear which mechanisms humans really find useful and how to make use of them. Their study indicated that no single mechanism meets all privacy needs. Thus, services such as Flickr could provide a plethora of privacy protection mechanisms that individual users could chose from in order to select privacy and security levels that meet specific needs. Grensing-Phphal (2011) suggested that although balancing functional needs versus privacy can be complex and challenging, sites could implement methods to better register users and keep track of IP address within their systems.

Technical solutions are only one part of the solution to protect children's security. Friedland and Sommer (2010) indicate even tech-savvy users find it difficult to accurately assess the risk they face. Thus, the authors suggest that the security and privacy community take a more active role in both developing technology to protect individuals, but also to educate Internet users on how to more effectively protect themselves. They suggest the Internet community should aim for a consensus in what constitutes acceptable levels of privacy for location-based services. It may be appropriate to consider that privacy and security controls for children be aimed at a more stringent level to afford a higher level of protection.

Legal safeguards for the online protection of children are problematic. While online privacy laws do exist in various countries to some extent, there is little that specifically addresses the needs of younger individuals. In the U.S., the Children's Online Privacy Protection Act (COPPA) only addresses Websites that are directed at children under 13 (Miyazaki, et al., 2009). Legal measures specifically designed to protect children's online data, such as geotagged information, should be implemented to provide adequate security protection for children.

Further research may be done in this area to understand more implications with children's privacy and security on the Web. First, this study did not ascertain whether the people who uploaded the pictures actually understand the technology behind geotagging and whether they actually know how much personal data they are uploading to a public domain. Second, this research analysed geotagged images for high-income locations within the US. It may be interesting to analyse whether the same percentage of geotagged children's images are found among lower-income locations within and outside the US.

6. Conclusion

This paper had two goals: to determine the range of geocoded children's images and to raise concerns and methods to address the issues. First, the study found that a significant number of children's images with geolocation information were found on Flickr for specific high-income U.S. postal areas. This finding could show the adverse effects on children's safety, as Internet surfers could possibly determine the exact location of where the children live, thus posing serious concerns with physical security. In order to mitigate some of these issues, a multidimensional approach should be implemented. Better technology could be used by location sites to protect privacy of information. Legal mandates could be implemented to specifically address the unique needs of younger individuals. Finally, the industry needs to better inform parents and individuals who post pictures to public Web sites that geolocation information can have both advantages as well as repercussions, as safety must be a prime priority.

7. References

Bhasin, M. (2006) 'Guarding Privacy on the Internet', *Global Business Review*, Vol. 7, No. 1, 137-156.

Bowie, N. and Jamal, K. (2006) 'Privacy Rights on the Internet, Self Regulation or Government Regulation', *Business Ethics Quarterly,* Vol. 16, No. 3, pp. 323-342.

Burdon, M. (2010), "Privacy Invasive Geo-Mashups: Privacy 2.0 and the Limits of First Generation Privacy Laws", *University of Illinois Journal of Law Technology and Policy*, Vol. 1, pp. 1-50.

Butcher, D. (2011), "Legal and regulatory risks associated with the growth of mobile marketing", Mobile Commerce Daily, http://www.mobilecommercedaily.com/2011/04/13/legal-and-regulatory-risks-associated-with-the-growth-of-mobile-marketing

Clements, M., Serdyukov, P., Vries, A., and Reinders, M., (2010), "Using flickr geotags to predict user travel behaviour", in *Proceeding of the 33rd International ACM SIGIR conferences on Research and development in information retrieval*, 19-23 July 2010, Geneva, Switzerland, pp.851-852.

Douglas, H. (2010), "Handsets provide a new way of connecting to the internet, but also provide new privacy problems for users", *ITPro*, http://www.itpro.co.uk/625108/mobile-users-wary-over-geotagging

Federal Trade Commission (FTC), (2010), "Protecting Consumer Privacy in an Era of Rapid Change, A Proposed Framework for Business and Policy Makers", http://www.ftc.gov/os/2010/12/101201privacyreport.pdf

Free Image Hosting (2011) "Free Image Hosting FAQ", http://www.freeimagehosting.net/faq.php

Flickr (2011), "Flickr Maps", http://www.flickr.com/map

Friedland, G. and Sommer, R., (2010), "Cybercasing the Joint: On the Privacy Implications of Geotagging", In *Proceedings of the Fifth USENIX Workshop on Hot Topics in Security* (HotSec 10), August 2010, Washington, D.C., http://usenix.org/events/hotsec10/tech/full_papers/Friedland.pdf.

Grensing-Pophal, L. (2011) "Security and Identity: Balancing Privacy, Flexibility and Ease of Use", EContentMag.com, Vol. 34, No. 1, http://www.econtentmag.com/Articles/Editorial/Feature/Security-and-Identity-Balancing-Privacy-Flexibility-and-Ease-of-Use-73858.htm

The Higley 100 (2011), "The Elite 100: America's Highest Income Neighborhoods", http://higley1000.com/archives/44

Joshi, D., Gallagher, A., Yu, J., and Luo, J., (2010) Exploring user image tags for geo-location inference. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),* 14-19 March 2010, Dallas, Tx, pp. 5598-5601.

Joshi, D. and Luo, J., (2008), "Inferring generic activities and events from image content and bags of geo-tags", In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, 7-9, July 2008, Niagara Falls, ON, Canada, pp. 37-46.

Lofing, N. (2010), "Facebook's new geotagging ability bring potential for abuse", *The Sacremento Bee*, http://www.sacbee.com/2010/08/24/2977615/facebooks-new-geotagging-ability.html#storylink=misearch

Miyazaki, A., Standaland, A., Lwin, M. (2009), "Self-Regulatory Safeguards and the Online Privacy of Preteen Children", *Journal of Advertising*, Vol. 38, No. 4, pp. 79-91.

Ozer, N., Conley, C., O'Connell, H., Ginsburg, E. and Gubins, T. (2010), " Location-Based Services: Time for a Privacy Check-In", ACLU of Northern California, http://www.aclunc.org/issues/technology/asset_upload_file533_9627.pdf

Photobucket, (2011), "About Photobucket", http://photobucket.com/about.

The Local, (2010), "Politicians call for 'geotagging' laws ahead of data summit", http://www.thelocal.de/sci-tech/20100920-29937.html

U.S. House, 111th Congress. "H.R. 5777, Best Practices Act." (Version 1: Version Date: July 19, 2010), http://www.house.gov/apps/list/press/il01_rush/h_r_5777_the_best_practices_act_2010.pdf

Vicente, C., Freni, D., Bettini, C. and Jensen, C. (2011), "Location-Related Privacy in Geo-Social Networks", *IEEE Internet Computing*, Vol. 15, No. 3, pp. 20-27.

Yu, J., Joshi, D. and Luo, J., (2009), "Connecting people in photo-sharing sites by photo content and user annotations", In *IEEE International Conference on Multimedia and Expo 2009*, 28 June-3 July, 2009, New York, NY, pp. 1464-1467.

Burghardt, T., Buchmann, E., Muller, J. and Bohm, K. (2009), "Understanding User Preferences and Awareness: Privacy Mechanisms in Location-Based Services," *Lecture Notes in Computer Science*, Vol. 5870/2009, pp. 304-321.

Figure 1: Results of Flickr Search

1,118 results matching "family or portrait or child or school " here