International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Encry-Pixel: A Novel Approach Towards Locational Privacy Enhancement in Images

Soumya Shaw[1], Ankur Mukherjee[2], Kartik Joshi[3], Ramya. V[4]

[1,2,4]B.Tech ECE, Vellore Institute of Technology, Chennai, Tamil Nadu, India
[3]B.Tech ECM, Vellore Institute of Technology, Chennai, Tamil Nadu, India

---

Abstract - *With all the liberty and under the hood of providing customization to the end-user, companies are extracting crucial data. Under the big data revolution that is shaping up the world, even though we acknowledge the Variety of Data, we often fail to account for the vulnerability of the Data. It is this 6th V of Big Data that is a concern that needs to be addressed. The variety of data includes multimedia, images, text, voice. With the recent boom in photo-sharing social media platforms like Instagram, there has been a surge in image-location mapped data. While this can be put to very good use in urban city planning, there is a privacy aspect associated with this that is going unnoticed. We attempt to mystify the metadata that includes the location coordinates to address the privacy concern. We put forward an algorithm that performs randomized location hopping to compromise the algorithms targeted to extract sensitive data. The idea is to hide sensitive information that might be used otherwise without user knowledge. We further explore availability-abstraction tradeoff for metadata retrieval such that the user gets to decide about the location sharing.*

Key Words: Geo-propagation, Image privacy, Location Randomization, Pixel Privacy, Privacy Enhancement

1. INTRODUCTION

Being in an era where the usage of Social Networking is transcending unabated, can be seen equivocal in two directions. Global Connectivity on one side and looming insecurities on the other. The technology-driven world has undoubtedly made the world knock our doorsteps but with its concerns as well. Data is everywhere, and one can make use of them in any way the person desires. Jiang *et al*. indicated Twitter, one of the most popular social platforms, produces immense volumes of social media data each year in his study [1]. They also explained how even our ethereal social presence can expose our interest, location, and Social Influence through *Top-k Influential Similar Local Query* (TkISL).

We take a deeper dive into the accessible data and focus on its subsets, i.e., the Images and Geotagging. The Social platforms have witnessed a growing craze for Images. Sites such as Flickr used to publish more than 100 million new images daily back in 2014 [2] [3]. Sites like *'Please Rob Me'* & *'Ready-or-not'* gives an exact status of the threat as well [4]. The advent of geotags unravels a newer dimension of privacy breach through its location, unintentionally. Lee *et al*. used a tag driven approach to exploit the information related to the location and explicitly mentioned the usage, but not limited to business strategies [5]. This exploitation needs a check and will be a pivotal point of discussion in our paper.

2. THE CONCERN

We took a survey of 100 people spreading across different age groups to strengthen our paper's claim and pre-requisite knowledge, the details are summarised in Table 1.

Table - 1: Survey Response Summary

Question	Yes	No	What's the harm?
Do you know that the Photo you click has hidden information like your location, time of capture, camera details, etc.?	91%	9%	-
Do you know that you can be tracked (movement on map) using the information you share on Social Media, mainly Pictures?	76%	24%	-
Do you know that your tagged location on Social Media or Internet can fetch you targeted advertisements based on your location?	62%	38%	-
Do you want them to know your location using your private photos?	98%	0%	2%
Do you know your Social Media Pictures are being used in 3rd party software (that you may have never heard of) to predict paths people take at tourist places?	78%	22%	-

The survey is a pristine reflection of the knowledge the society possesses and how they feel about it. The initial discussion now shifts to how this is important in terms of our society.

The central concern arises with the access of Images world wide, and within the fact that Images contain some hidden information such as metadata [2], also known as EXIF [6] associated with it which is predominantly vulnerable to the owner. Geotag is the social networking version of that information and Social Networking sites collect them all [7]. The survey points out that people expose their crucial information out there in public without even knowing. Users that are unaware that approaches to visual analysis and retrieval can breach their geo-privacy, unintentionally expose themselves to fraud threats or other unexpected effects for sure. Most of the instances include posting pictures directly, but Chen *et al*. suggested Microblogs as the other way pictures go into the public domain [8].

Two questions remain unanswered. First, why is protecting the metadata even required? Second, how do the existing algorithms do it? We will move onto the next sections after exploring the answers first. We encountered many papers that use the metadata to conclude their intrinsic performance. For a quick fact, until December 2016, Instagram witnessed over 282 million selfies [9]. Liao *et al*. used these features to increase the accuracy of Image Classification [6]. [10], [8], [11], [12] applies the same to mine tourist routes & movement patterns. Sreenivasamurthy *et al*. used the data for season prediction of locations [12], and Goldin *et al*. used it for Disaster remediation [13]. Land cover classification is achieved in [2]. Jiang *et al*. took vantage of the social media posts of a user to find the highest influencer within distance r, which can probably cause targeted ads or requests from agencies [1]. Li *et al*. explicitly mentioned that Geotags can successfully reconstruct a complete road map of a city commingled with computational tools, which would prove disastrous [14]. It could expose a strategically significant place that should not be public even though it was unintentional. However, no matter how much roiling some of the uses may sound, the fact that some applications only make a stride in proper usage cannot be denied. Thus, contemplating to annihilate the entire data would be ludicrous.

The metadata contains essential information like latitude, longitude, tags, camera information, time of capture, and some device-dependent information; however, all the algorithms mentioned exploit them. The location is the obvious thing to collect from the metadata. Oba *et al*. classified the photographs based on the keywords user types [2]. The texture of the image is taken into account for [7]. Date & time is extracted for users to get a flavor of the order of time by [10] and [11].

Conclusively, the survey and the details mentioned in this section point to a single notion that the usage of the open-source images in the public domain is a mixture of proper use and 'not so proper' use. Eventually, using public images must be done after permission as 'consent is a must' in these conditions.

3. FORMER ACCOMPLISHMENT

Choi et al. proposes an incredible work that will work in current circumstances but eventually will diminish with time. The paper suggests usage of filters to protect the geo-location information that may be visually matching with a 'geo-propagator' present in the collection if not used [4]. There is a hidden aspect to it as the result may be ephemeral. The algorithm will adapt to the filters soon, and the public will start using the popular filters and tag the geo-location with it, and the situation will again be the same.

There are two ways to tackle this. One, we need an infinite supply of filters that looks good on usage over a picture, and second, we somehow hide our location even if we do not shroud or encrypt the metadata. The first one is quite unrealistic, as developers would require incessant creativity with an exponential increase in complexity. However, on the other hand, the second method is what we are pondering upon in this paper.

4. THE CONCEPT

There is some crucial information that is noteworthy, in developing the algorithm to tackle the situation. The algorithm mentioned by Xie et al. inevitably rejects any Image with Geotag outside the 30 km range from the initial consideration for mining the route [7]. This is the first proof we encounter denoting location hopping will cease the pattern ever being formed. Yin et al. mentioned the importance of Geotags and its necessity to be taken into consideration for Landmark Recognition procedure [11]. In case, the User decides to hide the GPS location of the image and doesn't mention the location as Geotag, the algorithm goes for Visual Matching of that image with the Geo-Propagators, otherwise known as Geo-location Estimation (GLE). The visual matching algorithm/ GLE is a tool with a huge impact factor and the automated training images of Geo-Propagators can help them detect your location of Image with very high accuracy. Hence, your privacy is at question because someone in the world chose not to worry about it anyway.

Using textual information causes an error and not in use now [7] and the same can be achieved for the case of Geotags. Our idea exploits the same loophole to safeguard itself from visual matching algorithms and at the same time provide a fake Geo-propagator for the present

algorithm that can lead to wrong predictions of future Image, hence reduction in accuracy. We randomly replace the original location with a dummy location that's outside the range of consideration as well. As a result, the geotag will also be replaced with a randomized location and thus will point to anywhere else in the world which has no monumental relation with the former. The user can readily decide upon the option of replacing the location with its home location in the Geotag so that it safeguards them from a targeted robbery that can take place if the user explicitly mentions an out-of-town location in their post.

4.1 Random Hop

The world map is precisely assigned a range of latitude and longitude values that can uniquely identify the places all over the world and randomly generating a latitude and longitude value in the range can serve the purpose. Although inspired by [15], we tried going a step forward and made it much more realistic and genuine. The cartesian coordinates are converted into spherical coordinates or vice versa using the following transformations:

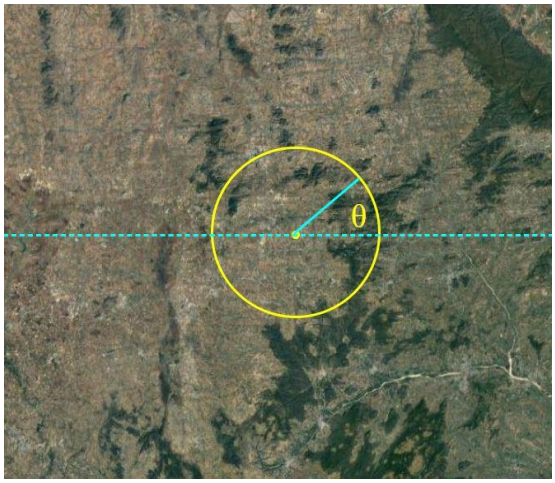
$$\theta = \boldsymbol{tan}^{-1}(\frac{y}{x})$$

$$\boldsymbol{R} = \sqrt{\boldsymbol{x}^2 + \boldsymbol{y}^2}$$

$$\boldsymbol{x} = \boldsymbol{r}\ \boldsymbol{cos}\ \boldsymbol{\theta}$$

$$\boldsymbol{y} = \boldsymbol{r}\ \boldsymbol{sin}\ \boldsymbol{\theta}$$

Now, any point can be defined using these two values. The original location will have a circular ring with a random radius that has the point to be picked and a specific coordinate can be generated by choosing a random angle.



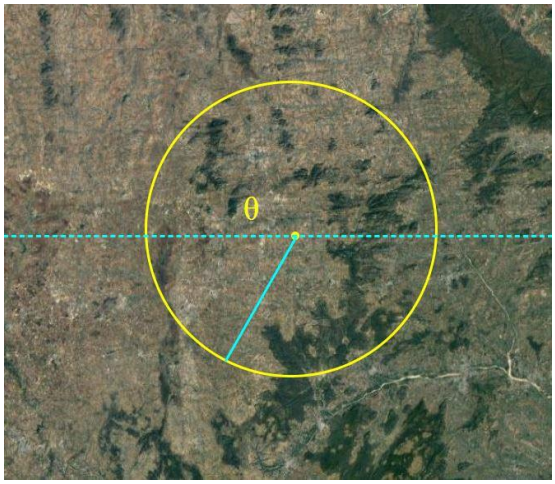


Fig - 1: Location coverage with varying R & θ values

The mentioned algorithm adds an extra dimension of the user choosing the extent of randomization using the R value. The radius is an effective way of categorizing the degree of expanse from the original location. The range selection of the user can also be treated as an optional step for the algorithm. Upon selection, a specific range of R can be treated as a single option and within that range, randomization can take place, making the process user defined yet randomly generated. One fact must be kept in mind though, choosing a very small R value will be as inefficient as not randomizing at all since the relocated coordinates will still be in close vicinity with the original one.

Conclusively, the conversion can be made back to the original coordinate system from spherical one, identical to the former using,

$$X_{new} = r\ cos\ \theta\ + X_{initial}$$
$$Y_{new} = r\ sin\ \theta\ + Y_{initial}$$

If the values go out of range, it can be brought back by using $x = x_i + 180°$ into the standard range, depending whether it is latitude or longitude and range.

4.2 Original Location Disclosure

Original location can be easily obtained once again with the knowledge of the radius and angle. One needs to use the same radius to mark the ring and point exactly opposite to the initial angle, i.e $180° + \theta$.

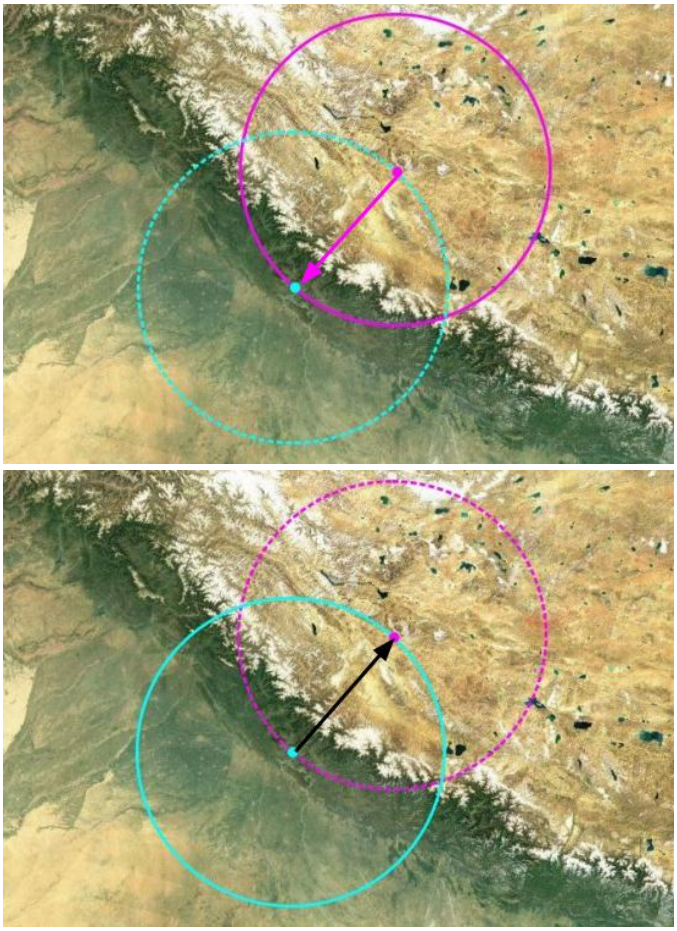


Fig - 2: Random Hop and its Retrieval

5. PERFORMANCE

There is presciently no concrete metric to evaluate the model but to check how the process may take place with and without utilizing the algorithm. We took a journey and the pictures associated with it to deduce the path followed in due course at Singapore, mimicking a Tourist Route mining algorithm. The information extracted is shown in Table 2 and depicted in figure 3 as well.

Table -2: Location Sequence of a Tourist at Singapore

Time Stamp	Coordinates	Location
6 Feb 2020, 13:30	(1.360107, 103.989586)	Jewel, Airport
6 Feb 2020, 15:30	(1.315928, 103.857670)	Mori Hostel
6 Feb 2020, 22:26	(1.286334, 103.852570)	Near Fullerton Hotel
6 Feb 2020, 22:45	(1.286972, 103.854745)	Merlion
6 Feb 2020, 23:58	(1.289015, 103.854912)	Esplanade
6 Feb 2020, 01:08	(1.288497, 103.860475)	Helix Bridge

Fig - 3: Location Sequence over the Map

Proper utilization of the algorithm resulted in the following location hopping sequence which is totally unpredictable and yields a different sequence each time applied.

Table - 3: Randomized Location Sequence

Time Stamp	Random R, θ	New Coordinates	Location
6 Feb 2020, 13:30	46.636485, 71°	(16.543461, 147.953333)	North Pacific Ocean
6 Feb 2020, 15:30	137.462730, 311°	(-88.500407, 0.113231)	Antarctica
6 Feb 2020, 22:26	51.334917, 184°	(-49.923534, 100.271628)	Indian Ocean
6 Feb 2020, 2245	97.595022, 248°	(-35.272767, 13.366216)	South Atlantic Ocean
6 Feb 2020, 23:58	28.278458, 17°	(28.331839, 112.122733)	Taojiang, Hunan, China
6 Feb 2020, 01:08	6.163412, 89°	(1.396063, 110.022948)	Bau, Sarawak, Malaysia

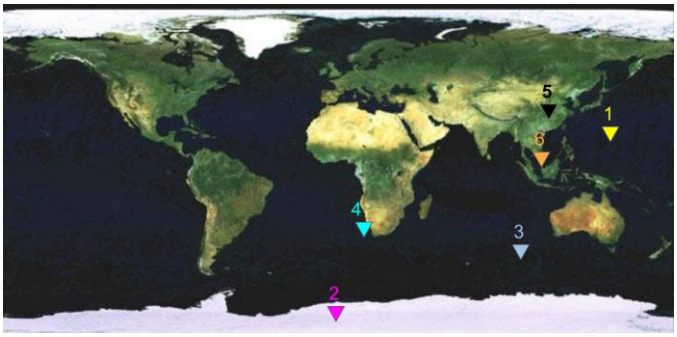


Fig - 4: Randomized Location Sequence over the Map

The sequence of travel generated as a result of the hopping algorithm is undoubtedly illogical in physical reality and doesn't signify any relevance or information worth mining compared to an actual path. The path, if used, will only contribute to accuracy reduction and thus better be left unused. Thus, the sole purpose of having such an idea is fulfilled and proves to be effective. On the other hand, if needed, the true sequence and location can be extracted with the permission and consent of the owner. There's one frail pattern that's evident in the output sequence. Since, the location hop was truly random concerning the world, it couldn't bias between a location on land and ocean. The Earth being approximately covered 71% by Water makes it plausible for the algorithm to hit a location on an ocean by the same probability.

6. CONCLUSIONS

The veracious inception of our idea is based on our belief that 'consent is a must' for any work you carry out with somebody else's information. With the knowledge added from the survey, it is imperative to conclude that the public is unaware of what happens after their eyes have left the digital screen and basically don't want that to happen that way. The requirements may be novel but permission is a prerequisite.

Thus, the random hopping framework was proposed and elucidated with an example. This algorithm stands out in amongst the others as it takes the users' preference as to how much the randomization has to be done, i.e., the radius and the hops are completely stochastic which, in no way, can be used to track the users' movement and if one had to obtain the original set of coordinates, the inverse function can be applied to get back the same with ease. Further, this algorithm can be refined to make it work even better with more efficiency.

Once again, our results are not a claim of Supremacy but an addition to existing features. We share a common goal of making this world much more ethical and immune to crimes and we move a step closer every time we work collectively backing each other's prominent contributions.

7. FUTURE WORK

No design is perfect, albeit opportunities infinite. The idea we put forward can also be improved upon to move a step closer to excellence. The algorithm uses arbitrary hoping to protect privacy by duping. There can be advanced hopping algorithms that prove to be more productive. One such algorithm we suggest is based on the hotspots of social media activity. The hop can be transformed to embed the location having similar photographic activity in comparison to the original. For that purpose, the algorithm must refer to a pre-mapped database of known hotspot details. This will make sure the relocated coordinates don't end up in an ocean just as our approach did.

Many advanced location encryption techniques already made their way to the technological world and look promising in their ability to provide results. Liu et al. suggested some of them, namely, Spatial Obfuscation, Dummy trajectories & Spatiotemporal obfuscation [15].

REFERENCES

[1] J. Jiang, H. Lu, P. Li, G. Pan, and X. Xie, "Finding influential local users with similar interest from geo-tagged social media data," in 2017 18th IEEE International Conference on Mobile Data Management (MDM), 2017, pp. 82–91.

[2] H. Oba, M. Hirota, R. Chbeir, H. Ishikawa, and S. Yokoyama, "Towards better land cover classification using geo-tagged photographs," in 2014 IEEE International Symposium on Multimedia, 2014, pp. 320–327.

[3] A. Gallagher, D. Joshi, J. Yu, and J. Luo, "Geo-location inference from image content and user tags," in 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009, pp. 55–62.

[4] J. Choi, M. Larson, X. Li, K. Li, G. Friedland, and A. Hanjalic, "The geo-privacy bonus of popular photo enhancements," inProceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ser. ICMR '17.New York, NY, USA: Association for Computing Machinery, 2017, p. 84–92. [Online]. Available: https://doi.org/10.1145/3078971.3080543

[5] I. Lee, G. Cai, and K. Lee, "Mining points-of-interest association rules from geo-tagged photos," in2013 46th Hawaii International Conference on System Sciences, 2013, pp. 1580–1588.

[6] S. Liao, X. Li, X. Wang, and X. Du, "Building geo-aware tag features for image classification," in 2014 IEEE

International Conference on Multimedia and Expo (ICME), 2014, pp. 1–6.

[7] Yi Xie, Huimin Yu, and Roland Hu, "Multimodal information joint learning for geotagged image search," in2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2014, pp. 1–6.

[8] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. Zhang, and J. Zhang, "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data, "IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 270–279, 2016.

[9] Mary Lister. 33 mind-boggling instagram stats facts for 2018. [Online]. Available: https://www.wordstream.com/blog/ws/2017/04/20/instagram-statistics

[10] E. Spyrou, I. Sofianos, and P. Mylonas, "Mining tourist routes from flickr photos," in2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2015, pp. 1–5.

[11] H. Yin, C. Wang, N. Yu, and L. Zhang, "Trip mining and recommendation from geo-tagged photos," in 2012 IEEE International Conference on Multimedia and Expo Workshops, 2012, pp. 540–545.

[12] S. Sreenivasamurthy and S. Frank, "Efficacy of season prediction for geo-locations using geo-tagged images," in 2015 IEEE International Conference on Information Reuse and Integration, 2015, pp. 476–484.

[13] S. E. Goldin and K. T. Rudahl, "Mobile geotagged data gathering for disaster remediation," in 2012 IEEE International Conference on Communication, Networks and Satellite (ComNetSat), 2012, pp. 69–73.

[14] J. Li, Q. Qin, J. Han, L. Tang, and K. Lei, "Mining trajectory data and geotagged data in social media for road map inference, "Transactions in GIS, vol. 19, 02 2014.

[15] B. Liu, W. Zhou, T. Zhu, L. Gao, and Y. Xiang, "Location privacy and its applications: A systematic study, "IEEE Access, vol. 6, pp. 17 606–17 624, 2018.

[16] Google. (n.d.). *Singapore*. Retrieved from https://goo.gl/maps/763ANJdGaaTBFFpZ6

[17] Klokan Technologies and OctoGEO. Openmaptiles satellite. [Online]. Available: https://openmaptiles.com/satellite/

BIOGRAPHIES

Soumya Shaw is currently pursuing his B.Tech in Electronics & Communication in the Vellore Institute of Technology, graduating in 2021. He is an IEEE member and interestingly working on Deep Learning, AI, Image Processing. Besides, he is overzealous in the field of Cosmology, Higher Dimensional Mathematics & Quantum Mechanics.

Ankur Mukherjee is studying his B.Tech Electronics & Communication at VIT Chennai. He is interestingly working in the field of Analog and Power Electronics, Signal processing and Embedded systems.

Kartik Joshi is studying his final year Bachelor of Technology (Electronics and Computer Engineering) at Vellore Institute of Technology (VIT) Chennai, India. He is interested in learning new emerging technologies. His research areas of interest also include Ethics in AI, Computer Vision, Natural Language Processing.

Ramya.V is currently pursuing her B.Tech in Electronics and Communication Engineering in Vellore Institute of Technology, Chennai. She is interested in Data Structures, Blockchain Technology, Web development and Information Technology.