The Crowd Wisdom for Location Privacy of Crowdsensing Photos: Spear or Shield?

TONGQING ZHOU and ZHIPING CAI*, College of Computer, National University of Defense Technology, China
FANG LIU*, School of Design, Hunan University, China

The incorporation of the mobile crowd in visual sensing provides a significant opportunity to explore and understand uncharted physical places. We investigate the gains and losses of the involvement of the crowd wisdom on users' location privacy in photo crowdsensing. For the negative effects, we design a novel crowdsensing photo location inference model, regardless of the robust location protection techniques, by jointly exploiting the visual representation, correlation, and geo-annotation capabilities extracted from the crowd. Compared with existing retrieval-based and model-based location inference techniques, our proposal poses more pernicious threats to location privacy by considering the no-reference-photos situations of crowdsensing. We conduct extensive analyses on the model with four photo datasets and crowdsourcing surveys for geo-annotation. The results indicate that being in a crowd of photos will, unfortunately, increase one's risk to be geo-identified, and highlights that the model can yield a considerable high inference accuracy (48%~70%) and serious privacy exposure (over 80% of users get privacy disclosed) with a small portion of geo-annotated samples. In view of the threats, we further propose an adaptive grouping-based signing model that hides a user's track with the camouflage of a crowd of users. Wherein, ring signature is tailored for crowdsensing to provide indistinguishable while valid identities for every user's submission. We theoretically analyze its adjustable privacy protection capability and develop a prototype to evaluate the effectiveness and performance.

Additional Key Words and Phrases: location privacy, crowd wisdom, photo crowdsensing

ACM Reference Format:
Tongqing Zhou, Zhiping Cai, and Fang Liu. 2021. The Crowd Wisdom for Location Privacy of Crowdsensing Photos: Spear or Shield?. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 142 (September 2021), 23 pages. https://doi.org/10.1145/3478106

1 INTRODUCTION

The ubiquitous camera-equipped devices of the mobile crowd have present tremendous opportunity for visually sensing, collectively monitoring, and comprehensively understanding the physical world [30][11][10]. There is a great literature devoting to the development and applications of such an emerging paradigm, such as city viewing and perception [28][42], event sensing [5][13], and daily healthcare [2]. From the aspect of the data requester or the sensing platform, the exploitation of crowd wisdom during these tasks is definitely beneficial as it essentially enables the fine-grained data collection in both the spatial and temporal domain [40]. Yet, a successful collection

*Corresponding author.

Authors' addresses: Tongqing Zhou, zhoutongqing@nudt.edu.cn; Zhiping Cai, zpcai@nudt.edu.cn, College of Computer, National University of Defense Technology, Changsha, Hunan, China, 410073; Fang Liu, School of Design, Hunan University, Changsha, Hunan, China, 410073, fangl@hnu.edu.cn.

2474-9567/2021/9-ART142 $15.00
https://doi.org/10.1145/3478106

depends critically on the active response of physically distributed users (participants), which will expose their real-time locations for tagging the photo. Location privacy-preserving techniques [35][51][38] are thus widely studied for motivating the participation of mobile users in crowdsensing tasks.

In fact, even with the raw locations perfectly hidden, one's location information may still be identified from the context [4][51][46]. In particular, the visual content of a shared photo is considered to carry certain patterns for the location at which it is taken [6]. Many research efforts have then been spent on inferring photo shooting locations based on visual matching techniques applied to a large collection of reference photos with explicit geo-tags. Specifically, they propose to either adopt a retrieval-based strategy that searches for one's most similar photos [14][34] or build classifiers based on the references for mapping photos to discrete locations [9][44][20]. However, we point out that there are usually no reference photos for the cases of crowdsensing as it is launched to depict uncharted domains in the physical world with uncertain visual views. For example, existing location inference classifiers (e.g., PlaNet [44]) are only capable to provide city-scale localization with just a few geo-tagged samples per city, while a crowdsensing task generally focuses on discovering dynamics in a finer-grained context (e.g., viewing a campus on the 1st day of school), wherein no prior knowledge is available.

In this paper, we move a step further and investigate a new question from the perspective of the users: Given a crowdsensing task with no reference photos in prior and adequate raw location protection guarantee, is one's involvement in such crowd wisdom a gain or a loss for its privacy? On one hand, joining a crowd of users to execute a crowdsensing task may feel deceivingly safer for a user than making a contribution singly with its own. However, would the accumulation of visual cues from multiple photos increases the risk for their shooting locations to be disclosed? On the other hand, can the crowd of users take active measures collaboratively to avoid possible information leakage?

To analyze the negative effect of the crowd wisdom on privacy, we first propose a more pernicious location inference model that requires no reference geo-tagged photos. The building block is the extraction of three types of crowd characteristics: (1) Feature representation knowledge distilled from *images of a generalized crowd* (e.g., 1.3M ImageNet dataset) can help to discover visual cues of photo location; (2) Visual correlations of the *crowdsensing photos* indicate their co-occurrence possibilities, thus can be used to group photos according to different geographic patterns; (3) Active annotation capability of *latent mobile workers* facilitates a straightforward way to identify a photo's shooting location with no prior knowledge. Through joint exploitation, the adversary model can selectively annotate the locations of some representative/seed photos and use them to infer the rest with a tunable cost.

Using four photo datasets with different domain characteristics and example crowdsourcing surveys, we conduct extensive assessment for the adversary model in terms of its performance on clustering, representativeness, annotation accuracy, inference capacity, as well as computation cost. The results show that the crowd-based representation (i.e., deep features) can facilitate better clustering accuracy and representative selection than using hand-crafted features. Meanwhile, we observe that mobile crowdsourcing can yield accurate geo-annotation for our representative photos, but relatively low performance on the annotation of random photos. This indicates that some photos are geographically inconspicuous for manual identification, while the exploitation of the crowd correlation can lead to more accurate annotation. Further, we highlight that with only a small portion of annotated photos, the model can infer photo location with relatively high accuracy, and more importantly, incurring serious privacy disclosure to the involved users (> 80% users' privacy are threatened). The annotation and inference for photos of event sensing are generally more difficult, while generating more clusters and recruiting more crowd workers can improve the accuracy.

To defend against such threats, we further propose an adaptive grouping-based signing model that can hide a user's track under the camouflage of a user group. The basic insight is that users of the same crowdsensing task share the same privacy concern, thus can form a shield for this crowd by using indistinguishable signatures for their shared photos. During the design, we integrate the general ring signature process in crowdsensing task announcement and registration, and provide adjustable privacy settings for users to balance the risk of privacy disclosure and the signing cost. We analyze the security property of the defense model with theoretical proof. A lightweight prototype is also developed to implement the model and evaluate the performance. Experimental results on the prototype show that the proposal can effectively guarantee users' various privacy requirements with acceptable transmission cost (e.g., 80KB for 300 users) and computation overhead (e.g., $< 90s$ for 300 users), which have significant improvement in terms of the overhead of a raw ring signature-based method.

The main contributions of this paper are summarized as follows:

- We study the pros and cons of the crowd wisdom for the location privacy of photo crowdsensing. Context and domain knowledge of the mobile crowd are extracted to support the discovery and protection of user location information in their shared photos.
- We propose a new line of threat that infers crowdsensing photos' shooting locations without any reference photos. The key design novelty lies in exploring a hybrid effort of crowdsourcing, crowd features, and crowd photo correlation. We examine the feasibility and performance of such threats based on four photo datasets and real-world annotation surveys. Several remarks and implications on the influence and protection of the pernicious information leakage risks are presented.
- We propose a defense model as a technical countermeasure for the identified threats based on an adaptive and indistinguishable signing algorithm. We analyze the privacy-preserving capability theoretically and show the overhead through the implementation of a prototype.

2 RELATED WORK

Location privacy is a major concern that may hinder users from participating in mobile crowdsensing tasks. In this section, we review the general location protection research for crowdsensing and the location privacy issues from sharing photos.

2.1 Location Privacy for Crowdsensing

Plenty of techniques have been proposed to mitigate the potential risks of location leakage in crowdsensing. As stated in a recent survey [27], cloaking is a widely adopted strategy that protects the precise participant locations by hiding them under a coarse area through spatial transformation or dummy locations, e.g., [36]. But such methods are criticized to be sensitive to adversaries' prior knowledge. More recently, differential privacy is introduced for implementing privacy-preserving MCS tasks. In[35][18][39], participants' locations are obfuscated with differential location privacy during the recruitment stage with the overall traveling distance for fulfilling the tasks minimized. An optimal location obfuscation policy for crowd coverage maximization is designed in [38] under certain differential privacy limitations. To address the threats from long-term observation attacks, Niu et al. [24] harness differential privacy and k-anonymity for effective location perturbation with the impacts to usability minimized. Further, some dedicated obfuscation techniques are proposed for data recovery in sparse crowdsensing. For example, Wang et al. [41] focus on reducing the data inference error incurred with differential location obfuscation. Zhou et al. [51] propose a correlation-preserving location obfuscation scheme that provides an effective camouflage without degrading data recovery precision.

Based on these efforts, we thus assume that the raw location information of participants can be well protected and further investigate the tricky threats from the inference category. Namely, the raw visual content and the location of a photo pose no privacy concerns during its submission in our cases. We focus on deducing the intrinsic privacy of photos when they form a large group.

2.2 Photo Location Estimation and Protection

Photo location estimation/inference (a.k.a., reverse geo-tagging, location re-identification) indicates the process of predicting a photo's shooting location for intended privacy disclosing or unintended data analysis [6]. This field is also related to landmark classification and recognition [25][45] in the computer vision community. We can roughly divide the related methods into two categories: retrieval-based and model-based. On one hand, retrieval-based methods determine the location of a query photo by searching for the most visually similar photos in a pre-built geo-tagged dataset [14][34]. Wherein, the matching process can be conducted based on hand-crafted features (e.g., SIFT, SURF) [14] or Siamese network [34]. On the other hand, model-based methods build classifier with machine learning techniques to learn the geographical pattern of different locations from geo-tagged photos. Along this line of work, Fang et al. [9] propose GIANT as an SVM classifier that detects discriminative regions from the training photos and extracts geo-informative attributes for each city. In Google's PlaNet [44], the globe is divided into size adaptable cells with each cell a class for the photos located in it. Then the PlaNet model is trained based on these photos as a multi-classification problem. In [20], researchers consider integrating the context-aware features in learning representation to give more weights to regions that positively contribute to geo-localization. The inference problem is also investigated in other fields, for example, Li et al. [21] design a Bayes' theorem-based model to find location implications from the words in tweets.

We point out these methods all need a well-built training dataset to cover the characteristics for all the candidate locations. As a result, they are merely adequate to infer locations when sufficient reference photos are provided, which does not match the sensing cases for uncharted targets in crowdsensing. In contrast, we study a different problem of photo location inference for fine-grained visual sensing with no reference photo in prior.

To avoid photo location from being identified by the visual cues, traditional image/matrix encryption methods [7][51] can be helpful by allowing access for authorized requesters only. Although effective, the platform couldn't properly perform necessary photo preprocessing (e.g., filtering, summarization) or data analysis, given blind views after encryption. Meanwhile, these methods cannot prevent a curious requester from performing location estimation on receiving sufficient photos. One may also consider protecting photo geo-privacy by directly introducing differential privacy, like the pioneering efforts in numerical crowdsensing data collection and aggregation (e.g., weighted average [17][50], truth discovery [32]). Yet, it is hard to apply differential-privacy-based data publishing methods to the image content as elements in the image matrix share strong correlations [22]. An effective perturbation will add noise on the views and incur accidental viewing quality degradation, which is unfavorable in photo crowdsensing applications. Without disturbing the viewing of photos on the platform, Choi et al. [6] investigate the gain of applying popular photo enhancement on location privacy, while the performance of misleading the inference model by selectively pruning photos from a collection is studied in [47]. Different from these proposals, we present a defense model based on indistinguishable signature with the photos unstained, as well as a theoretical security guarantee. Finally, we believe proper incentive mechanisms [17][48][50] can help to encourage user participation with relatively looser privacy requirements, while is an orthogonal topic deserving independent investigation.

3 NO-REFERENCE LOCATION INFERENCE OF CROWDSENSING PHOTOS

In this section, we introduce our design for inferring location from crowdsensing photos. We first define the problem of no-reference photo location inference. Then we describe the details of a novel adversary model for geo-identification.

3.1 Problem Definition

Traditional photo geo-identification (a.k.a., image geolocation) techniques require context knowledge that builds on tremendous reference photos with geo-tags on the targeted locations. A new photo's location is then predicted based on visual similarity estimation between it and the reference photos or geographic classifiers trained on these references. Yet, from the aspect of crowdsensing, the assumption on the existence of reference photos for all locations does not always hold due to the spatial and temporal diversity of the physical world. Otherwise, it breaks the tenet of crowdsensing tasks for visually discovering the physical world on-demand. For example, the classifier of Google's PlaNet [44] has only static knowledge of city-scale landmarks. For a crowdsensing task that attempts to discover visual dynamics within a city, say a block or a campus, usually no prior knowledge is available for inferring the locations. Therefore, in this work, we focus on a more tricky problem as following:

No-reference photo location inference: Given a collection of crowdsensing photos pho_i ($i \in [1, N_{pho}]$) with no location tag (obfuscated or encrypted) under the required domain $\mathcal{D}$ and time constraint $\mathcal{T}$, we try to identify the location loc_i of photo pho_i based merely on the collection without any previously established reference photo sets.

Compared with the existing geo-identification techniques, such a form of inference can pose a more severe threat to user privacy as no additional metadata or prior context knowledge is required. For the formulations in the rest of the paper, we use blackboard bold to denote sets or vectors (e.g., $\mathbb{U}$ for user set) and use calligraphy to represent general strategies or requirements (e.g., $\mathcal{G}$ for grouping strategy). We denote the number of a specific entity x as N_x (e.g., N_{pho} with subscript pho representing photos).

3.2 'Spears' from the crowd

To infer photo locations from the scratch (i.e., with no reference), we first figure out what resources can be our help.

- Our basic vision is that why don't we simply ask the crowd to manually identify the shooting locations, just as asking the crowd for sharing visual depictions of a domain. Following this idea, a straightforward way is to leverage **mobile crowdsourcing** to put geographic annotation for each photo.

Manually annotation can be very effective for the geo-identification of images. For example, in a famous online challenge for identifying a given photo's shooting location[1], hackers collect and utilize the captured flight information, geometric contexts, historic weather forecasts to successfully discover the hotel where the photographer stands. However, such a brute-force solution is not feasible in terms of the cases of photo crowdsensing for two aspects of practical limitation. On one hand, it cannot scale to frequent and ubiquitous crowdsensing that involves hundreds even thousands of photo collections, as the cost for recruiting workers can be quite large. Meanwhile, it restrains the capability of automatically searching for vulnerabilities in large collections of photos; On the other hand, perceptual characteristics of individual photos may not provide sufficient geographical information for manually identifying their locations. For example, the upper left photo in Fig. 1 presents little clues for its shooting location as only one corner of a building is captured.

We point out that, apart from mobile crowdsourcing, two other 'spears' from the crowd can be further exploited to overcome the above limitations for implementing an effective attack:

- Visual correlations among the crowd of photos provide opportunities to deduce the spatial co-occurrence of them. In other words, we believe that the strength of such correlations can implicitly indicate whether two photos are visuals of the same PoI, as we will evaluate in the experiments of Sec. 4. In this way, although some photos are geographically inconspicuous, we can infer their locations from their neighbors who are representative of them.
- Although we are lack visual samples for the interesting locations during crowdsensing, what we do have is the knowledge distilled from tons of images contributed by a more generalized crowd (e.g., 1.3M ImageNet dataset [29]). As stated in the deep learning community, the internal structure of the end-to-end networks

[1] https://nixintel.info/osint/using-flight-tracking-for-geolocation-quiztime-30th-october-2019/

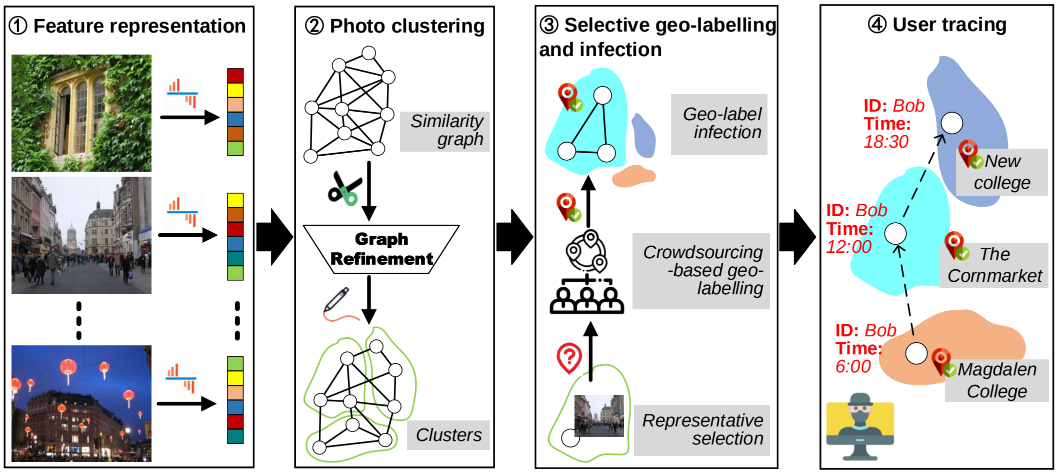


Fig. 1. Framework of no-reference location inference adversary model. Photos in the OXlet dataset are used to present an example task for campus visual sensing. Wherein, on receiving a collection of photos, an adversary (the platform or an authorized requester) infers the location of each photo and traces a specific user (Bob in this example) according to the locations of his photos.

trained on high-level classification tasks with labeled images is useful in investigating image synthesis [49]. In our settings, this representational knowledge can help to discover photos' geographical features and efficiently measure the correlation.

All these three types of crowd characteristics are taken as pitfalls of the crowd in terms of crowdsensing photos' location privacy.

3.3 Adversary Model

We propose to solve the above problem by designing a novel adversary model, whose target situations are typical photo crowdsensing tasks with honest-but-curious crowdsensing platform or requesters.

Task model: The considered crowdsensing tasks consist of three entities, including requesters that are interested in some PoIs, users (a.k.a., participants) that volunteer to join a task and share their visual contents, and crowdsensing platform that regulates users (e.g., recruitment, incentive [17][32][50]), gathers photos, and performs filtering and selection. Initially, a task can be launched by either a requester in an on-demand manner (e.g., Waze[2], Fliermeet [12]) or the platform as a centralized service (e.g., Beautiful China[3]). For the latter cases, requesters can passively fetch and view the photos from the platform.

A task is specified with a tuple in the form of <domain, region, time, number>: 'domain' describes the visual domain of the PoIs (e.g., event, buildings); 'region' and 'time' give the spatial and temporal preferences for collecting photos. The region constraint can be defined with a set of interested PoIs with size N_{PoI}; and 'number' further defines the expected amount of photos for each PoI. A volunteered user then responds to the task with its local photos.

Security assumptions: For the privacy of users' raw locations, we assume effective location obfuscation (e.g., cloaking) will be performed to avoid direct leakage during the sharing. On this basis, the assumed adversaries are

[2] https://www.waze.com
[3] http://www.quanjingke.com/dest

honest-but-curious entities who can get access to all the shared photos. This includes the platform that gathers the photo and requesters that are authorized to view the crowdsensing results. Furthermore, we assume the adversaries, especially the platform, can deliberately decompose a task in the spatial dimension, namely, into sub-tasks with fewer target PoIs (e.g., 10), by actively launching fine-grained sub-tasks or passively distinguishing submissions from different cities (by matching the source IP address with a historic IP-city database[4]). In this way, the pool of candidate locations for each photo can be narrowed down (with a much smaller N_{PoI}), simplifying the inference (curious) without impacting photo collection (honest)[5]. Finally, the portrait privacy contained in the shared photos is not in the scope of this work.

From the aspect of an adversary, an active inference is built on joint exploitation of the above-listed crowd characteristics ('spears'). Specifically, given a collection of crowdsensing photos, a four-step location inference framework is designed as shown in Fig. 1.

(Step 1) Feature representation: In order to effectively analyze photo correlations, we propose to first extract visual features from them as structural representation. We propose two alternatives for this purpose: low-level features and deep visual features. On one hand, we compute the descriptive hand-crafted features using SIFT (the scale-invariant feature transform) for every photo. A group of key points in the photo is detected, each represented as a 128-dimension feature. By performing K-means clustering on all the detected features, we then obtain a fix-size vocabulary (codebook) of visual words depicting representative characteristics of the collection of photos. We set the scale of the vocabulary (i.e., V) to 500 and 2000 to maintain an acceptable efficiency for the adversary, as it would consume significant time for K-means to handle a large number of clusters. Then a photo is represented as a sparse histogram that contains the frequency of occurrence of each visual word in its local features, which is known as the Bag of Visual Word model (BoVW [25]). As $|V|$ clusters are generated, the dimension of a photo's BoVW feature is $|V|$.

We point out that the low-level features may ignore perceptual cues for geographical identification. Meanwhile, it is time-consuming to calculate a vocabulary based on SIFT. Considering that an end-to-end model can discover complex features that generalize to different prediction contexts, we then introduce an advanced representation based on the deep embeddings extracted from dedicated DNN models. In particular, we propose to adopt the internal activations of the MobileNet model [31], which is a popular architecture for image processing in mobile computing applications. Our representation network retains the 20 shallow layers (i.e., 1 fully convolution layer and 19 residual bottleneck layers) from the Imagenet-trained MobileNetV2 with width multiplier $\alpha = 1$. Such a network is very lightweight (8.97MB) in size, which makes it a favorable choice for adversaries, especially a curious requester. Taking a re-scaled and cropped photo with size 224×224 as input, the modified network finally embeds it into a deep feature of 67k-dimension (1280 feature maps with sizes 7×7). We unit-normalize both the BoVW and deep features in the photo dimension, and attain the final representations.

(Step 2) Photo Clustering: In this step, we attempt to cluster photos that share similar visual cues together. In this way, the visual correlation of photos can be translated into geographic co-occurrence possibilities for further location inference. For this, we build a similarity graph of photos, where each node represents a photo and each edge denotes the pairwise similarity between two photos. We propose to use the cosine similarity to measure the correlation between two photos' feature vectors, which equals their inner product after normalized to length 1. A complete graph is then attained for the photo collection. To mitigate the influence of weak correlations between those irrelevant photos, the graph is refined by reserving only the edges between a node and its top-K nearest neighbors. Given the expected number of photos N_{col}^{i} for the i-th location/PoI (determined when publishing the task, see Sec. 5.2.3), we define the threshold for the number of neighbors as $K = max(N_{col}^{i})$.

[4] Only IP-to-city mapping is available with an accuracy of ~50%, so the IP-based side information is just helpful for coarse-grained task decomposition, not location inference. Please kindly refer to https://www.iplocation.net/ for more information.

[5] An extreme decomposition (e.g., a sub-task with less than 3 PoIs) will impair user participation as their privacy are directly disclosed when contributing to such sub-tasks, thus is unreasonable to be adopted in practice.

Photo clustering is then transformed into a graph partition problem, which can be well handled by spectral clustering [23]. Wherein, we point out that the number of clusters N_{cls} is the crux for tuning the clustering performance. Basically, since a PoI/location usually presents multi-views that may share weak similarities, we should set $N_{cls} > N_{PoI}$. For example, the views inside and outside a target building can be very different, thus at least 2 clusters are needed for accurately clustering the photos of this PoI. A larger N_{cls} means smaller clusters and a higher probability for photos in a cluster to be visuals from the same location (i.e., higher precision). Yet, this comes at the cost of more clusters with uncertain geographic locations needed to be identified. To this end, an adversary shall set the N_{cls} as large as the cost is acceptable, which we will analyze in Step 3.

(Step 3) Selective geo-labelling and infection: We perform location annotation for photos through mobile crowdsourcing in this step. Based on the above processing, annotation can be simply performed in a cluster-grained as photos in a cluster are believed to be homogeneous on location. In particular, given a budget N_{lim} on the affordable number of crowdsourcing tasks, we can have $N_{cls} \cdot N_{rep} \leq N_{lim}$, where N_{rep} is the number of representatives in each cluster for manually annotation. Since crowdsourcing results may contain error or discrepancy, increasing N_{rep} can improve the annotation accuracy for the corresponding cluster, while overall clustering precision would be degraded with a smaller N_{cls}. We defer an orthogonal strategy for a balanced inference gain to future work and, w.l.o.g., we set $N_{rep} = 1$ in the following description[6].

For each cluster, the representative is selected as the node with the maximum cut within the corresponding connected component, namely, the photo with the highest sum of visual similarities to the rest photos. The underlying reason is that a photo with high similarities to the others is deemed to be witnessed at their locations. After obtaining the locations for the representatives with crowdsourcing, we use them as seeds to geo-annotate the photos in each cluster the same as its representative. In this way, we attain the possible locations of all the collected photos.

(Step 4) User tracing: Finally, each user's photos are mapped to a series of <time, location> records, together disclosing its historic track during the participation, as shown in the example of Fig. 1. Given the inference results independently extracted from multiple tasks, an adversary may very likely obtain more records of a user at different times of a day. Even worse, a powerful attacker could leverage the proposed model to launch deliberate tasks that covers different locations and push them to a target user, in this way, inferring a fine-grained track for stalking the victim. As a result, the more PoIs and more tasks one contributes to (i.e., an active user), the larger privacy leakage it may experience. This will significantly degrade the enthusiasm for photo sharing.

4 EXPERIMENTAL STUDY ON LOCATION INFERENCE ATTACK

We carry out multiple analyses on the proposed adversary model based on several real-world photo datasets. We study the location inference capability in detail by examining both the performance in each step and the overall performance in terms of inference accuracy, influence, and latency.

4.1 Datasets

In our analysis, we utilize real-world photos with task-specific visual domains (e.g., campus viewing) to simulate photo collection through crowdsensing. Especially, explicit PoI labels for photos are required as the geographical ground truth for performance evaluation. For this, we refer to four publicly available datasets of different scales (i.e., OXFORD [26], Div150Multi [16], EUFLOOD [1], and CUFED [43]) from the computer vision community to build our datasets with multiple levels on numbers of photos, users, and PoIs. Photos with PoI labels are picked out during the pre-processing. Then we manually filter out irrelevant photos from the raw datasets (e.g., photos for a cat named Big Ben in Div), and avoid redundant views or aspects that are shared by the same user,

[6] Note that we will use the number of manually geo-annotated photos and the number of clusters interchangeably as they have equal value when $N_{rep} = 1$.

considering that this can aggravate its ratio of photos easily being geo-identified. The basic information (domain, scale) and statistics of the data are summarized in Table. 1.

Table 1. Basic statistics of the adopted datasets.

Datasets	Visual Domain	Geographic Scale	# Photos	# Users	# PoIs
OXlet	Campus: scenery, scene	Campus	350	60*	12
DIV	Resort: scenery, scene, objects	World wide	7059	493	30
FLOOD	Disaster: events, scene	City	260	13	12
CUlet	Trip: scenery, activities	World wide	887	100*	31

* User id information is not provided in the raw Oxford and CUFED datasets. We generate 60 and 100 users for them and randomly assigning the photos to users as their uploaded contributions.

OXlet. The raw OXFORD dataset consists of photos collected from Flickr for particular Oxford landmarks, wherein the names of the corresponding landmarks are used as the PoI labels. The extracted OXlet dataset is characterized as a task with small N_{pho} and N_{PoI}. Meanwhile, we randomly allocate photos to 60 virtual users to simulate a low per-user contribution (i.e., $\frac{\#photos}{\#users}$).

DIV. DIV is constructed based on the Div150Multi 2015 dataset, which is with photos of 30 famous resorts distributed all over the world. It represents tasks with large N_{pho} and N_{PoI}, and medium per-user contribution. In practice, tourists may be likely to launch tasks for this domain to find interesting places to visit.

FLOOD. EUFLOOD consists of images related to the event of the central European floods in May/June 2013 and has been fetched in July 2017 from Wikimedia. Since PoI labels are not available in the raw dataset, we perform fixed-width clustering for the photos, as a pre-processing step, based on their GPS locations and consider those in the same cluster of the same PoI. In particular, we set the clustering width to $200m$ (i.e., the coverage of a PoI is a circle with radius $\leq 100m$) and filter out clusters with ≤ 10 photos. We check and name the extracted PoIs by searching for the centroid locations on the Baidu map. The obtained FLOOD dataset has the largest per-user contribution with relatively small N_{pho} and N_{PoI}.

CUlet. CUFED contains photos of diverse events and is organized as a series of albums. To simulate a specific visual domain, we select the photos for beach trips from CUFED in our evaluation. The elaborated dataset characterizes a small N_{pho} but large N_{PoI} task context. By manually mapping photos to 100 virtual users, we realize a medium per-user contribution for CUlet.

4.2 Clustering and representative selection performance

We first study the performance of clustering (Step 2 in the adversary model) by examining whether the photos grouped into the same cluster correspond to the same PoI (i.e., clustering precision[7]). We choose five different numbers of clusters (equivalent to the number of geo-labeled photos) for each dataset. The results (averaged on each dataset) for different feature representation techniques are illustrated in Fig. 2. We can see that clustering based on the features extracted from dedicated DNN outperforms the BoVW-based clustering by a large margin (i.e., $\geq 20\%$) on every dataset. Even provided with a larger word vector and N_{cls} (e.g., 2000-dimension features and $N_{cls} = 40$ for FLOOD), the precision of the BoVW-based approach is still lower than 52%. This evaluates our intention of using deep features for geographical correlation exploitation.

More interestingly, we observe that, for similar visual domains, the model presents better clustering performance on the dataset with smaller N_{PoI} (~73% on OXlet v.s. ~50% on DIV), then with smaller N_{pho} (~65% on FLOOD v.s. ~60% on CUlet), as more clusters and photos will both add to the clustering hardness. Besides, with similar N_{PoI}

[7] We don't focus on the performance on recall as it is not relevant to the inference capability (an adversary would expect the labeled representative to 'infect' accurately, instead of widely).

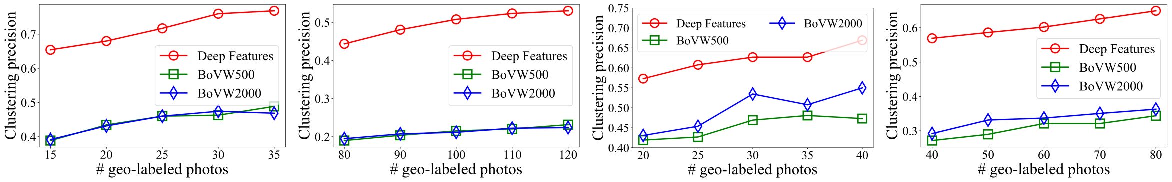


Fig. 2. The clustering precision of the adversary model (From Left to Right: OXlet, DIV, FLOOD, CUlet).

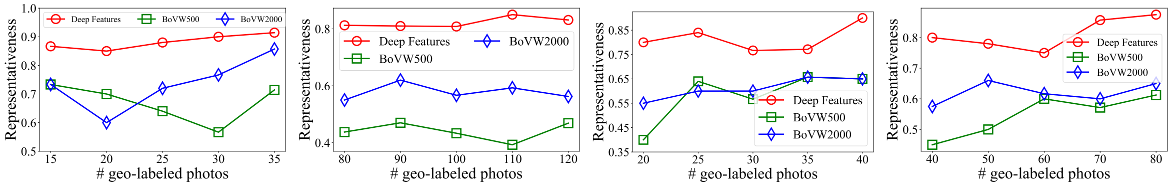


Fig. 3. Performance on selecting representative photos from each cluster (From Left to Right: OXlet, DIV, FLOOD, CUlet).

and N_{pho}, photos of landmarks (e.g., OXlet) tend to be more easily clustered than that of events (e.g., FLOOD) for generally sharing more similarities on visuals. Meanwhile, a larger number of clusters usually brings better precision as the less similar photos in a cluster is gradually grouped into a new cluster, so a larger number of clusters is beneficial for an adversary as long as it is affordable for crowdsourcing.

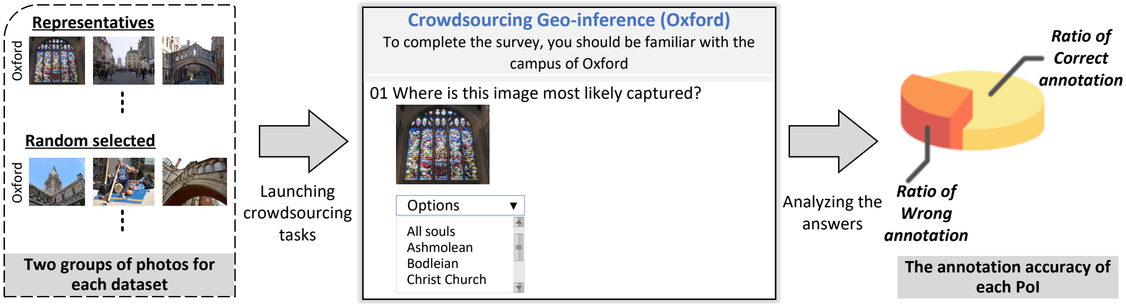
Remark 1: *The feature representation model trained with tremendous crowd-contributed images (i.e., Imagenet) helps to better discover crowdsensing photos' geo-correlation online than referring to the hand-crafted features.*

Then we investigate the performance of picking out representatives/seeds from the generated clusters. In particular, we measure the performance by introducing a metric named *representativeness*, which is calculated as the ratio of photos that belong to the same PoI as the selected representative in each cluster. The results are depicted in Fig. 3. Again, we observe a large margin on the performance of deep feature-based (> 80% for all the tested situations) and BoVW-based approaches, which owes to the differences in clustering precision and similarity measuring. It is also shown that the representativeness doesn't always increase with the number of clusters, which indicates that the selection process is not affected by the size of the cluster.

4.3 Mobile crowdsourcing for geo-identification

Given the selected representative photos, we construct mobile crowdsourcing tasks based on the Tencent Questionnaire and assign them to workers through a social platform to manually infer the photos' locations. Since launching crowdsourcing for every tested situation (we have 60 cases) means too much cost, without loss of generality, we further choose a subset of the representatives elaborated based on the deep features (12, 10, 5, and 10 for OXlet, DIV, FLOOD, and CUlet) to study the crowdsourcing performance. We also randomly select the same numbers of photos from the raw selection, and construct four new crowdsourcing tasks for comparison. As an example, Fig. 4 shows the crowdsourcing-based geo-annotation process for OXlet. During each task, we recruit 15 workers[8] and describe the context knowledge for the domains of different datasets before starting the tasks. Workers are paid with monetary reward through red envelope (about 0.6CNY for each task). The average time needed for each inference question (per photo) of the task on OXlet, DIV, FLOOD, and CUlet are 26s, 18s, 40s, and 32s, respectively.

[8] We do this for evaluation purposes. One may in fact request a much smaller number of workers in practice (e.g., 3).


Representatives
Oxford
Random selected
Oxford
Two groups of photos for each dataset
Launching crowdsourcing tasks
Crowdsourcing Geo-inference (Oxford)
Options ▾
All souls
Ashmolean
Bodleian
Christ Church
Analyzing the answers
Ratio of Correct annotation
Ratio of Wrong annotation
The annotation accuracy of each PoI

To complete the survey, you should be familiar with the campus of Oxford

01 Where is this image most likely captured?

Fig. 4. The construction of mobile crowdsouring tasks for photo geo-identification. The example use several representatives selected with our adversary model and random samples from Oxford, and presents the UI of the corresponding crowdsourcing task. Statistics on the geo-annotation accuracy are then calculated via our real-world survey.

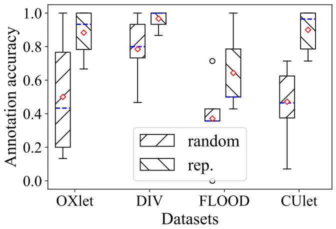


Fig. 5. The accuracy of the PoI labels annotated through crowdsourcing. 'rep.' and 'random' represent selecting photos for crowdsouring geo-annotation based on the proposed adversary model and a random strategy, respectively. The red diamond gives the mean value.

We then examine the accuracy of crowdsourcing annotated locations for the photos by calculating the ratio of correct labels in the 15 answers. The distributions of the accuracy for different photos in each task are depicted in Fig. 5. As shown, the manually geo-annotation accuracy on our model selected representatives are much higher than that of the randomly selected ones, and generally more stable, especially when the geographical context is clear (i.e., in OXlet, DIV, and CUlet). This observation indicates that the correlations among a crowd of photos can lead to more accurate geographic annotation, and further validates the design effectiveness of clustering and representative selection.

Remark 2: *Interestingly, the relatively low performance on the annotation of random photos also demonstrates that brutally annotating all the photos with crowdsourcing cannot yield a favorable accuracy.*

Annotating photos in the event domain is relatively harder and takes a longer time. This is because visuals for an event usually have a weak correlation with the locations where they are captured. In FLOOD, several photos from different locations may turn out to provide similar views on the disaster. Yet, we can still observe an average accuracy of ~65% on it. Note that we estimate the annotation accuracy of each worker independently. In practice, we usually aggregate the answers through majority voting to resolve discrepancies. In this way, the performance of crowdsourcing can be further improved. Empirically, given that the average annotation accuracy on OXlet, DIV, and CUlet are around 89%, 95%, and 90% (Fig. 5), the expected accuracy can reach 96%, 99%, and 97% when an adversary recruits 3 workers. Provided with 5 workers, the annotation accuracy for FLOOD is also expected to be 94%.

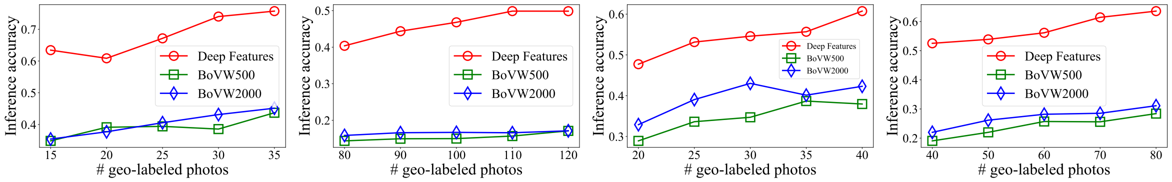


Fig. 6. Inference accuracy for geo-identifying photos to their GT locations (From Left to Right: OXlet, DIV, FLOOD, CUlet).

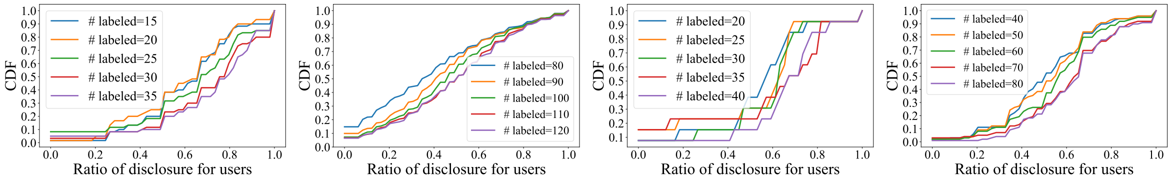


Fig. 7. Ratios of correctly geo-identified photos for each user (From Left to Right: OXlet, DIV, FLOOD, CUlet).

Remark 3: *Leveraging the mobile crowd wisdom can yield accurate geo-annotation for representative photos. For difficult annotation tasks (e.g., visuals of an event), more crowd workers can be helpful.*

4.4 Overall inference performance

Next, we evaluate the overall inference capability by investigating the ratio of photos that are correctly geo-identified (i.e., inference accuracy) and the privacy risks on each user (i.e., the ratio of disclosure). Since we didn't perform crowdsourcing for each tested case, we assign an empirical annotation accuracy factor for each dataset using the statistics above. As shown in Fig. 6, the inference accuracy based on the deep features is much higher than the other two designs. Specifically, we can observe the best performance on OXlet (i.e., ~70%), which shall owe to its small N_{PoI} and N_{pho}. For larger datasets DIV and CUlet, the inference accuracy can reach ~48% and ~60%, which are still much higher comparing with a random guess ($1/N_{PoI} \approx 3\%$). Generally, a crowdsensing photo has nearly $\geq 50\%$ probability of being correctly geo-identified. Note that the single photo geo-identification accuracy of state-of-the-art reference-based inference methods [47] is ~20%.

Remark 4: *Even only manually annotating a small portion of photos, one can infer a single photo's location with a relatively large accuracy using the adversary model.*

In Fig. 7, we present the cumulative distributions for the ratios of locations being identified. As shown, for each test case, over 80% of users get at least one location correctly inferred (see the y-values of the curves for the ratio of disclosure 0). In DIV and FLOOD, some users only submit less than 2 photos, leading to a small portion of low disclosure ratio. The influenced ratios increase to $> 90\%$ given more labeled photos. We claim that, although the inference performance seems modest for single photos, the privacy disclosure risk from the perspective of users is significantly high, even under low per-user contribution (e.g., OXlet, CUlet[9]), indicating serious privacy threats. Meanwhile, compared with the manual annotation results for photos of a random user (Fig. 5), the threat to one's location privacy is obviously higher when jointly considering a collection of photos from many users.

Remark 5: *Contributing submissions together with a crowd of photos will, unfortunately, increase one's risk of being geo-identified.*

[9] It is not hard to see that the privacy disclosure threats would be further aggravated if we reduce the number of users in these two datasets.

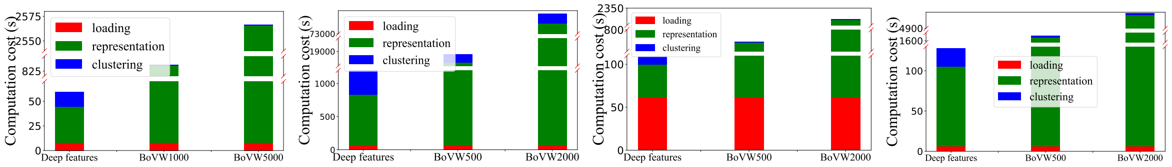


Fig. 8. The time consumption of the adversary model (From Left to Right: OXlet, DIV, FLOOD, CUlet).

4.5 Inference Latency

Finally, we test the latency of the proposed location inference model. The experiments are conducted using Python on a workstation with an Intel Core i7 processor and 64GB RAM. As shown in Fig. 8, the deep feature-based approach is much faster than the BoVW-based approaches (e.g., ×35 on OXlet, DIV, and CUlet compared with BoVW1000 and ×7 on FLOOD compared with BoVW500) for that the extraction and representation of the hand-crafted features are time-consuming. The computation overhead for clustering on deep features is bigger than that on the hand-crafted ones as the former has a larger dimension. Specifically, a large latency for the clustering step can be observed on DIV (i.e., ~1250s) because the computation complexity of spectral clustering significantly increases with an increasing number of nodes (photos). The unusually large loading time for FLOOD is caused by the extra pre-processing step introduced to group photos geographically and assign them the PoI labels.

4.6 Discussion

Adversary preferences for small N_{PoI}. The curious adversaries would favor tasks with small-but-unsuspicious numbers of PoIs (i.e., N_{PoI}) to enlarge the inference accuracy by reducing the clustering hardness (according to the results in Fig. 6 and Fig. 7), and more importantly, to control the crowdsourcing costs with as small budget N_{lim} for that $N_{PoI} \leq N_{cls} \leq N_{lim}$. Hence, adversaries are motivated and capable, as we explain in the security assumptions part, to decompose a large task so that they can realize effective and efficient location inference on some small sub-tasks. Since this process causes no impact on the crowdsensing results, it will not violate the honest-but-curious assumption. Such preferences also justify our choices of datasets with generally 10~30 PoIs.

Limitations. The inference accuracy would be degraded with incompetent crowdsourcing workers. Recall that the adversary model selects some seed photos for manual annotation, and deduces the other photos' locations using them. If the recruited workers are unfamiliar with the targeted region or don't understand the tasks, the seeds' location annotation would be inaccurate, resulting in many inference errors for the rest photos. The annotation errors cannot be totally overcome, one can try to involve sentinels with obvious geographic visuals to distinguish competent workers.

Generalizing the design. As a new line of threat on crowdsensing privacy, the proposed adversary model shall be considered as a base model for more deliberate attacks. As such, our intention is not to compare the proposal with other reference photo-based techniques or to claim who is superior. In fact, based on our design, an adversary can fine-tune the deep model for geo-pattern extraction or even replace it with more complex architecture (e.g., vision transformer) for improved feature representation, thus attaining better inference performance.

Meanwhile, we notice that existing efforts on photo selection (summarization) usually require location tags on the photos to build a spatial coverage model for a photo subset [52]. To this end, the clustering and selection steps in our model facilitate an alternative solution for this problem when only visual content is provided.

Implications for the users. Given the above comprehensive analysis on the location inference capability of a potential adversary, we conclude that characteristics of a crowd (e.g., visual correlation and co-occurrence possibilities) play an important role in successful geo-identification. To this end, we give three practical suggestions for users to securely participate in a crowdsensing task:

- First, a user can attempt to provide a different view for the targeted entity (e.g., keeping a long distance from the building, taking photos from an innovative aspect) to cut down the overlapping visual clues with the other photos, thus avoiding being easily localized from his contribution.
- Second, one may consider contributing to a PoI that is being requested for fewer visual descriptions or yet has fewer contributions to further reduce its correlation with co-located photos.
- Third, one should be aware that sharing photos from multiple sites corresponding to different PoIs may aggravate the risk of being traced, so is suggested to carefully balance its benefits (e.g., monetary incentive) and privacy requirements/expectations.

Note that the crux of the above countermeasures is to reduce the possibilities for one's photo to be identified in a visually correlated group, but definitely not to refuse to join in a crowd of participants. In fact, as we will discuss in the following sections, making contributions together with a large number of peer participants can technically form a natural shield against the inference attacks.

5 DEFENSING AGAINST THE INFERENCE THREATS

As aforementioned, users of a crowd that involves in a visual sensing task can accidentally aggregate the privacy disclosure risks of each other. Although cautious participation can reduce the risk of being geo-identified, *a technical countermeasure is still in need to provide a theoretical guarantee on user privacy.* Hence, in this section, we move a step forward to investigate the possibility of relieving such a pitfall of the crowd by exploiting the cooperation utility of the crowd.

5.1 Basic idea: shield of the crowd

We point out the essence of users' location privacy is the mapping between a specific identity and a series of geographical sites. Formally, the privacy information can be presented as $< u_i, ((t_1, loc_1), (t_2, loc_2), ...) >$. If user identity is blurred, an adversary cannot link the inferred locations from photos to the corresponding user.

A naive solution based on this idea is by adopting anonymous techniques to hide the real identity, which is widely used in nowadays mobile apps. However, it is not appropriate for the campaigned photo collection tasks of crowdsensing. First, a participant (user) of the task may very likely contribute several photos. Given any context information of his/her routine, one can deduce the link between the anonymity and the true identity, rendering the leakage of all the other locations s/he had visited. Second, total anonymity is not favored for the regulation of the task as users and their contributions are not authenticated. As a result, unreliable ingredients from illegitimate users may easily mix into the collection. Forcing each participant to register a unique key for authentication purposes will implicitly expose one's identity to the platform and make the following submissions unprotected.

We argue that, in fact, *the users that participate in a task as a mobile crowd, share the same security concerns on the centralized platform, thus naturally forming a virtual community that provides camouflage for all members.* Intuitively, users can share data privately in the name of the crowd instead of their own identities. By assuring that a submission is made by one user of a legitimate crowd but not knowing which specific user s/he is, such a camouflage facilitates an effective trade-off between the privacy-preserving expectation of the participants and regulation requirements of the platform. In this way, the crowd members work cooperatively to protect each other's identity, which is called the shield of the crowd.

Referring to the field of data security, *ring signature* [3] appears to be a competent solution that caters to our intention of attaining identity privacy and unforgeability simultaneously on a group of users. Given a piece of data signed with ring signatures, a verifier is convinced that the signature is computed using one of the group members' private keys, but the verifier is not able to determine which one. This property can be used to preserve the signer's identity from a verifier. In the following, we will extend the bilinear maps-based ring signature [3] to construct our adaptive defense model.

5.2 Signing photos with ring signature

For the paper to be self-contained, we first briefly describe the security primitives for ring signatures. Then we describe the signature construction details for our photo crowdsensing scenarios with its privacy protection capability theoretically analyzed.

5.2.1 Security primitives: Bilinear map. A multiplicative cyclic group is by definition a group of integers that supports certain multiplicative operations and is generated by a single element. Let G_1 and G_2 be two multiplicative cyclic groups of large prime order p, and g be a generator of G_1. Then any elements in G_1 can be written as g^x with some x. Based on these settings, a bilinear map is a map $e: G_1 \times G_1 \rightarrow G_2$ that holds the following properties [37]:

(1) Bilinearity: For all $u, v \in G_1$ and $a, b \in Z_p^*$, $e(u^a, v^b) = e(u,v)^{ab}$. Wherein, Z_p^* represents integers in the range of $(0, p)$.

(2) Computability: There exists an efficiently computable algorithm for computing map $e(u,v) \in G_2$.

(3) Non-degeneracy: $e(g,g) \neq 1$.

These properties are useful for new cryptographic constructions as it provides basic homomorphic operations for the inputs.

5.2.2 Signature construction and verification. For privacy-preserving, we require each user to register to a task via the crowdsensing platform when deciding to join in. The curious-but-honest platform will share parameters among the users of the same task for them to sign indistinguishably. Specifically, making crowdsensing contributions with ring signature involves four basic steps:

Task Announcement. For each task, the platform chooses generator g_1 and order p and obtains a multiplicative cyclic group G_1. For transmission efficiency, it also introduces a public map-to-point hash function $H: \{0,1\}^* \rightarrow G_1$. The platform will publish these global parameters (i.e., (p, g_1, G_1, H)) together with the general task requirements.

User Registration. For each interested user u_i, he randomly picks an integer $sk_i \in Z_p$ as his private key and computes his public key as $pk_i = g^{sk_i} \in G_1$. Then the users upload their public keys to register to the platform. Wherein, true identity can also be provided together with one's public key for regulation purposes, such as reputation management. After receiving the registrations, the platform publishes the public keys $(pk_1, ..., pk_{N_u})$ to each participant.

Local Signing. An user u_i signs for every photo pho_i^k he shares. For this, pho_i^k is first transformed to a byte message $m_i^k \in \{0,1\}^*$. Then he randomly chooses $a_{ij} \in Z_p^*$ $(i \neq j)$ and computes $s_j = g_1^{a_{ij}}$ to fake all the other users' signature elements. At the same time, he computes his own signature element as

$$s_i = \left(\frac{H(m_i^k)}{\prod_{j \neq i} pk_j^{a_{ij}}}\right)^{1/sk_i} . \tag{1}$$

The signature for m_i^k is then constructed as a ring of the signature elements he generates, namely, $\mathbb{S}_i^j = (s_1, s_2, ..., s_{N_u})$. Finally, user u_i uploads $\{pho_i, \mathbb{S}_i^j, t\}$ to the platform, where t is the time interval that the submission falls in. Since elements in $\mathbb{S}_i^j$ all belongs to G_1, they are indistinguishable to the receiver.

Platform Verifying. On receiving a submission $\{pho, \mathbb{S}, t\}$, the platform first computes $\delta = H(m_{pho})$ and then verify this submission by checking

$$e(\delta, g_1) \overset{?}{=} \prod_{i=1}^{N_u} e(s_i, pk_i),\ s_i \in \mathbb{S}\ . \tag{2}$$

If the above equation holds, the platform believes photo *pho* is signed and submitted by one of the N_u registered users. Otherwise, the submission will be considered unreliable and dropped.

5.2.3 Privacy risk analysis. Based on the security primitives, the above construction process inherits the correctness and unforgeability properties of traditional ring signature. Details of the proof can be found in [37]. Here we analyze the privacy protection property and overhead of the above construction.

Theorem 5.1. *For any inference algorithm $\mathcal{A}$, a sensing task with N_u users, a random user u_i, and its submission $\{pho, \mathbb{S}, t\}$, the privacy risk level $Pr[u_i, loc^*|\mathcal{A}(pho, \mathbb{S})]$ is at most $1 - (1 - \frac{Pr_{inf}^{\mathcal{A}}}{N_u})^{N_{col}^t}$, where $Pr_{inf}^{\mathcal{A}}$ is the probability of inferring pho's correct location using $\mathcal{A}$ and N_{col}^t is the number of co-located photos at time interval t for the inferred location loc^*.*

Proof. For each element s_j $(j \in [1, N_u])$ of $\mathbb{S}$, $s_j \in G_1$. We can have that the distribution of $\mathbb{S}$ is identical to that of $(g_1^{a_1}, ..., g_1^{a_{N_u}})$. According to [37], the probability $Pr[u_i|\mathcal{A}(pho, \mathbb{S})]$ of identifying the owner of *pho* from its signature $\mathbb{S}$ is at most $1/N_u$. Note that merely guessing the identity cannot pose an effective threat to the privacy, an adversary should also infer the correct location of this photo. Given the inference capability $Pr[loc^*|pho] = Pr_{inf}^{\mathcal{A}}$, we can deduce that the probability of disclosing u_i's location from one submission is $\frac{Pr_{inf}^{\mathcal{A}}}{N_u}$.

Crowdsensing tasks usually require retrieving more than one photo for each interested site to attain a comprehensive view, namely, the number of co-located photos $N_{col}^t > 1$ for interval t. Multiple independent submissions for one location will aggravate the threat to each user's privacy. That is, the adversary only has to guess that u_i is one of the owners of the many submissions. Therefore, given the number N_{col}^t, the probability for locating u_i equals to inferring the probability for it to make at least one submission for some location loc^*, i.e.,

$$Pr[u_i, loc^*|\mathcal{A}(pho, \mathbb{S})] = 1 - (1 - \frac{Pr_{inf}^{\mathcal{A}}}{N_u})^{N_{col}^t}. \tag{3}$$

□

Obviously, a larger number of co-located photos will make it easier for an adversary to trace the users. In practice, the platform is required to choose a reasonable N_{col}^t when publishing the task for both collection efficiency and privacy guarantee. Otherwise, the users can refuse to participate considering the disclosure risk of their tracks.

Attaining the above privacy gain incurs additional computation and communication costs. From the aspect of signature generation, although it is conducted on local devices of users distributedly, the time complexity and transmission overhead are proportional to the amount of participated users. Specifically, as shown in the local signing step, a valid signature consists of N_u elements of G_1. Let G_1 be a p-order cyclic group. It means that, even we aggregate different blocks of a photo into the same message when computing $H(m_i^t)$ in Equation (1), a photo's ring signature still requires $N_u \cdot |p|$-bits and N_u multiplication operations. For the centralized platform, $N_u \cdot N_u$ mapping operations in Equation (2) should be performed to verify one submission from each user. As an example, assuming a task with 1000 users and choosing $|p| = 160bits$ as the general setting does, a signature for one photo will reach $160Kbits$. It can be frustrating to both the user and the platform as the size of a signature is comparable to that of a compressed photo [40]. Since the overhead is proportional to the number of users, it also significantly limits the scalability of this signing approach.

5.3 Adaptive grouping-based signing model

To further reduce the overhead while still preserving privacy, we exploit a fine-grained and adaptive grouping strategy, named AGS, for ring signature in this part. According to the analysis in Sec. 5.2.3, when the inference capability is empirically bounded and the co-located number is set, the parameter N_u determines both the privacy protection capability (Equation (3)) and the overhead. Yet, we highlight that the above raw signing approach provides a K-anonymity [33] protection for every user's locations (cannot distinguish the identity from a group of K members) with anonymous (privacy) level $K = 1/Pr[u_i, loc^* | \mathcal{A}(pho, \mathbb{S})] \propto N_u$, which can be unnecessarily high for user's general privacy protection expectation.

In fact, $K = 5$ is generally believed to facilitate sufficient anonymity protection for the corresponding users [15][19], while for a crowdsensing task of $N_u = 500$ (100), the above signature can yield $K > 120$ (25) with $Pr^{\mathcal{A}}_{inf} = 0.8$ and $N^t_{col} = 5$. Meanwhile, mobile users hold different sensitivity on their privacy information (some may consider $K = 2$ acceptable [33]), as a result, simply enforcing them to follow the same privacy protection strategy can be inflexible. These facts give us the opportunity to compromise with the overhead over the guaranteed privacy level.

Algorithm 1: Privacy-aware user grouping for ring signature

```
Input: User set $\mathbb{U} = \{u_i | i \in [1, N_u]\}$, self-defined risk level $k_i$ for $u_i$, number of co-located photos $N^t_{col}$,
       empirical location inference probability $\widetilde{P}_{inf}$;
Output: A set of user group $\mathbb{G}^*$, the group size for users $\{N^{u_i}_{\mathcal{G}}\}$;
Set the initial index of user group $j = 1$ ;
Sort the users in $\mathbb{U}$ in ascending order by their risk levels $k_i$ ;
while $i \leq N_u$ do
    Compute the expected group size $n_i = \frac{\widetilde{P}_{inf}}{1-(1-k_i)^{1/N^t_{col}}}$ for $u_i$ ;
    if $n_1 > N_u$ then
        Issue an alert that privacy requirements cannot be satisfied ;
        return $\emptyset, \emptyset$ ;
    end
    if $i + n_i - 1 \leq N_u$ then
        Build a new user group $\mathbb{G}_j \in \mathbb{G}^*$ ;
        Add users $u_i, ..., u_{i+n_i}$ to $\mathbb{G}_j$, and set $N^{u_i}_{\mathcal{G}}, ..., N^{u_i+n_i}_{\mathcal{G}} = |\mathbb{G}_j|$ ;
        Update the group index $j = j + 1$, and set the next investigated user id to $i = i + n_i$ ;
    else
        Add all the left users $u_i, ..., u_{N_u}$ to group $\mathbb{G}_{j-1}$, and set $N^{u_i}_{\mathcal{G}}, ..., N^{N_u}_{\mathcal{G}} = |\mathbb{G}_{j-1}|$ ;
        break ;
    end
end
return $\mathbb{G}^*$, $\{N^{u_i}_{\mathcal{G}}\}$ ;
```

To this end, we propose to adaptively divide the crowd of users into small groups to tune dedicated $N^{u_i}_{\mathcal{G}}$ (the size of the group that u_i belongs to) for users according to their differentiated preference on the privacy level. Formally, a user is allowed to set his preferred risk level k_i to explicitly require a $(1/k_i)$-anonymity grouping when register to the platform. Wherein, a larger k_i indicates a higher probability of privacy leakage. Without loss of generality, here we denote k_i to be in the range of $[0.1, 0.5]$ (i.e., optional anonymous level $K \in [2, 10]$). After gathering all the preferable risk levels, the platform solves the following overhead minimization problem to find a globally optimized grouping $\mathbb{G}^*$ for each user:

$$\mathbb{G}^* = \arg\min_{\mathbb{G}} \sum_{i=1}^{N_u} N_{\mathcal{G}}^{u_i},\ s.t.,\ 1-(1-\frac{Pr_{inf}^{\mathcal{A}}}{N_{\mathcal{G}}^{u_i}})^{N_{col}^t} \le k_i,\ N_{\mathcal{G}}^{u_i} = |\mathbb{G}_j|\ for\ u_i \in \mathbb{G}_j,\ \mathbb{G} = \cup_j \mathbb{G}_j.$$

The solution to this problem is illustrated in Algorithm 1. Specifically, we first sort the users according to their risk levels in an ascending order (Line 3) to prioritize the grouping of more privacy-sensitive (less overhead-sensitive) users. In this way, those who need a larger group size will be accommodated first with the minimum number of group members (Line 11-13). After several rounds of grouping, the remaining users are those with the lowest privacy expectation and can be easily satisfied by adding them to the last group with the smallest size (Line 16). We can thus strictly guarantee that none of the users' privacy requirements are breached. Note that some users may be grouped into a larger group than he expects. This happens when the number of users with the same risk level cannot satisfy their own privacy requirements, so some users are endowed with a higher level to fill the privacy gap. The time complexity for this process is $O(N_u)$ and only has to be performed once per crowdsensing task.

Finally, the platform publishes the set of groups $\mathbb{G}$ to each user. A user u_i can then construct signatures $\mathbb{S}_{\mathcal{G}}^i$ for his submissions with the public keys of the members in his group, which has a size of $N_{\mathcal{G}}^{u_i} \ll N_u$. Accordingly, the risk level of u_i becomes $Pr[u_i, loc^*|\mathcal{A}(pho, \mathbb{S}_{\mathcal{G}}^i)]$ and its overhead is only $N_{\mathcal{G}}^{u_i}/N_u$ of the raw signing approach.

6 EVALUATION OF THE DEFENSE MODEL

In this section, we first discuss the privacy security property of the adaptive defense model. Then we evaluate its performance based on a prototype we design. As the overhead of both the raw signing approach and our AGS model have been theoretically analyzed in Sec. 5, we will just investigate the run time performance of transmission and computation with experiments on real photo data.

6.1 Security analysis

We propose to mitigate location inference threats on crowdsensing photos by blinding the identities of the involved users with our AGS model. In this part, we will analyze how the proposal can attain proper location privacy protection and discuss the security properties under different situations, respectively.

Location privacy. In AGS, each user that participates in a crowdsensing task (defined by user size N_u and expected co-located photos N_{col}^t) will specify its personalized privacy requirement with privacy risk level k_i, which indicates the largest probability of location disclosure it can accept.

Theorem 6.1. *AGS is able to guarantee that the location privacy requirement of each user is not breached during a task.*

Proof. Given a user's risk level k_i, the minimum group size n_i that can guarantee this specified risk level is calculated with Alg. 1 (Line 5). Following the heuristic strategy in Alg. 1, we group users in the descending order of their privacy requirements, which gradually adds users needing smaller n_i into a group until its size satisfies the requirement of the 1st user in this group. Therefore, for each u_i, we can have $N_{\mathbb{G}}^{u_i} \ge n_i$. According to the identity indistinguishability property of ring signature, the probability of identifying the user from its group based on the signature in its submission is at most $1/N_{\mathbb{G}}^{u_i}$. Then, according to Theorem 5.1, AGS maintains the privacy risk for each user with:

$$Pr[u_i, loc^*|\mathcal{A}(pho, \mathbb{S})] = 1 - (1 - \frac{Pr^{\mathcal{A}}_{inf}}{N^{u_i}_{\mathcal{G}}})^{N^t_{col}} \leq 1 - (1 - \frac{Pr^{\mathcal{A}}_{inf}}{n_i})^{N^t_{col}} = k_i.$$

For situations where $max(n_i) > N_u$, no grouping solution can be achieved and AGS will generate alerts to users with $n_i > N_u$ as their privacy requirements cannot be satisfied. In these cases, the task will not proceed unless more users join ($N_u \uparrow$) or the corresponding users loose their requirements ($n_i \downarrow$). Therefore, we prove that AGS can guarantee that the location privacy of each user is protected aligned with their own expectation. □

Security with a few user involvement. As proved above, AGS can protect privacy in tasks with small user size by strictly matching their requirements to proper groups and suspending the task when the matching problem cannot be resolved. Here we further discuss an extreme scenario with just one user. Note that the user will experience an actual risk level of $Pr^{\mathcal{A}}_{inf}$ for each submission. Namely, the security bound totally depends on the performance of the inference attacks. Interestingly, as the number of collected photos significantly decreases in this case, the inference capability $Pr^{\mathcal{A}}_{inf}$ would also degrade to that of a random selection (refer to Remark 2). If such a level of privacy risk (~0.6 for our tested results in Sec. 4.3) is still acceptable, then the task can proceed with the user's location privacy intact. In fact, we highlight that the situation of fewer users is unfavorable practically from the perspective of the task campaigner, which may act as a curious adversary, because this would deteriorate the quality of crowdsensing. Hence, a reasonable campaigner will proactively recruit sufficient users to maintain the quality, which indirectly avoid the cases with small user participation.

Security against collusion. Collusion is a form of high level attack that happens under strong assumptions of the involved entities. In our cases, the privacy guarantee of AGS may be breached when some users collude with the adversary, wherein the identities of these 'traitors' in their submissions are known by the adversary. Formally, the privacy risk of a honest user would be enlarged to $1 - (1 - \frac{Pr^{\mathcal{A}}_{inf}}{N^{u_i}_{\mathcal{G}} - \epsilon})^{N^t_{col}}$ with ϵ denoting the number of 'traitors' in its group. However, one should see that the collusion cannot commit an effective privacy disclosure as long as $N^{u_i}_{\mathcal{G}} - \epsilon \geq n_i$. This also specifies the security boundary of AGS (i.e., partially secure against collusion).

6.2 Implementation

We develop a light-weight prototype, named **crowdShield**, based on Java to implement our adaptive defense model[10]. We use the Java Pairing Based Cryptography (JPBC) library [8] to obtain the complex cryptographic properties in Sec. 5.2. In particular, the Type A elliptic curve with the form of $y^2 = x^3 + x$ is used as the base to construct bilinear pairing. By locally running the prototype as an entrance for sharing crowdsensing photos, a ring signature will be embedded into the submission for anonymity and authentication purpose. Generally, it allows defining one's preferred risk level before signing the loaded photo (set to be 0.3 by default), and will dynamically print some summarizations for the signing process. We emphasize that the prototype currently performs the setting, signing, and verifying steps all at the user side. However, it can be easily merged into a crowdsensing platform to support visual sensing and privacy protection distributedly.

6.3 Experimental results

Setup. We empirically set the inference capability to be $\widetilde{Pr}_{inf} = 0.8$, which we believe to be an upper bound of possible inference techniques according to the results in Sec. 4. We assume that the number of co-located photos will increase with the number of users that participate in a task. In particular, we set N^t_{col} to $N_u/20$ and $(5 + (N_u - 100)/10)$ when there are < 100 and ≥ 100 participants, respectively. To simulate users with

[10] The implementation is available at https://github.com/anonymous2021-0/crowdShield.

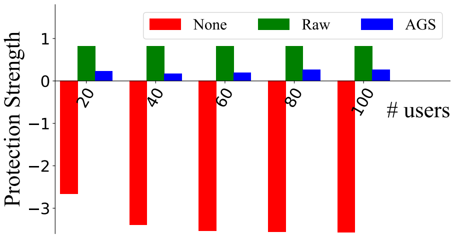


Fig. 9. The protection strength for location privacy of different (or no) countermeasures.

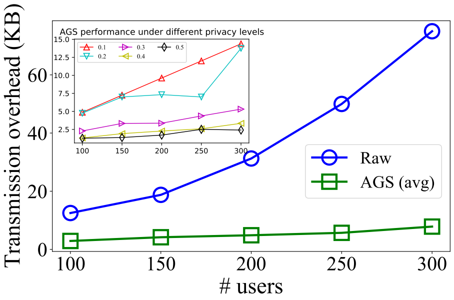


Fig. 10. The signature size as transmission cost under different amount of users.

various privacy preference, each user's risk level (i.e., k_i) is randomly set to be one of the five discrete values in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. Without loss of generality, we make the number of users corresponding to different risk levels to be the same, thus obtaining a uniform distribution. All the experiments are carried out using crowdShield on the same workstation mentioned above and tested with 10 photos from each dataset (i.e., 30 times).

Privacy protection. Besides the theoretical analysis in Sec. 6.1, we further illustratively show the privacy protection effectiveness of AGS by comparing it with crowdsensing without protection (None) and the raw blind signature method (Raw). For this, we introduce a metric named *protection strength*, which reflects the overall privacy-preserving performance for the users and is calculated as $\frac{1}{n}\sum_{i=1}^{N_u}\frac{k_i - Pr[u_i, loc^* | \mathcal{A}(pho, \mathbb{S})]}{k_i}$. The corresponding results are depicted in Fig. 9. Remember that the user-defined privacy risk level k_i specifies its privacy expectation (i.e., $1/k_i$-anonymity), so a positive (negative) strength indicates that the the expectation is satisfied (breached) [11]. As shown, AGS can provide effective protection even when the user size is small (e.g., 20). Raw presents obviously higher protection strengths, which are far beyond user expectation. As we will analyze later, this advantage comes at the cost of efficiency. We emphasize that the design goal here is to provide protection within the self-defined privacy bound with as small overhead. To this end, our flexible grouping strategy (i.e., AGS) that accommodates efficient privacy protection, instead of simply pursuing higher strength at large communication and computation cost, is more favorable by the users.

It is worth noting that AGS doesn't impact the clustering, geo-labelling, or geo-inference accuracy of our adversary model, as visual contents of photos are intact with AGS. Our AGS model takes effects by cutting off the link between a photo to its contributor using the crowd camouflage. In this way, even successfully identifying a photo's location, an adversary cannot tell it is the location of the corresponding contributor or other $k_i - 1$ users.

Transmission overhead. The results on the additional transmission (storage) overhead of the raw signing approach and the AGS model are depicted in Fig. 10. As expected, AGS shows a large margin on the performance compared with our raw signing approach. Specifically, we can observe an obvious increase in the overhead of the raw approach with an increasing number of users (reaching nearly 80KB when $N_u = 300$), while the cost for AGS only increases slightly. This also demonstrates that AGS can *scale to large crowdsensing tasks*. The performance on different risk levels is shown in the small window. Generally, the privacy-sensitive ones cost more as their signatures have more dimensions and larger sizes.

Computation time. Since the defense model involves additional computation steps on both the user device and platform, we investigate the corresponding generation and verification time, respectively. The results are shown in Fig. 11 and Fig. 12. The performance curves of Raw and AGS are similar to that of the transmission

[11] Interestingly, a zero strength means everyone is endowed with exactly $1/k_i$-anonymity as they required.

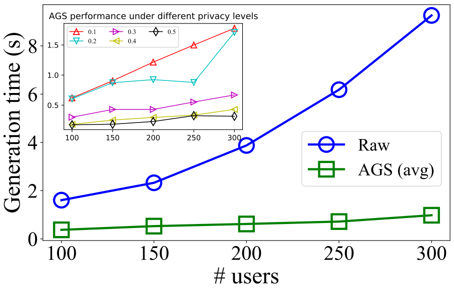


Fig. 11. The signature generation time for the user side with increasing number of users.

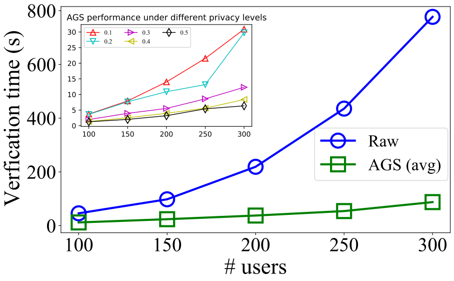


Fig. 12. The submission verification time on the crowdsensing platform with increasing number of users.

overhead. For each signature, the generation step is faster than the verification step (around ×4) due to the adoption of more complex multiplication operations. From the aspect of the task, verification will incur intensive computation resource consumption as the platform has to deal with all the signed submissions. For the raw approach, the cost can be as large as $800s$, which is not acceptable even provided with higher computing capability. Yet, the AGS model can significantly reduce the cost to a relatively reasonable range $< 90s$, which further validates the benefit of adaptive grouping for tuning the performance.

7 CONCLUSION

In this paper, we study the pros and cons of the involvement of the crowd wisdom on location privacy of photo crowdsensing participants. A pernicious no-reference location inference model is first proposed based on three types of crowd knowledge, which are considered pitfalls of the crowd. Experiments on real-world photo datasets and questionnaire-based surveys show that being in a crowd of photos aggravates the risk of a photo being geo-identified. We also observe that the adversary model can yield accurate inference for even the geographically inconspicuous photo with only a small annotation cost incurred, while more geo-annotated photos will usually help to improve the inference capacity. Implications in view of such threats are present to guide cautious participation. Furthermore, the identity hiding capability of a crowd is investigated by integrating ring signature in crowdsensing. We introduce an adaptive grouping strategy that allows users to specify their privacy protection levels and efficiently constructs groups to satisfy their requirements. A prototype is implemented for the defense model and experimental results on it demonstrate its effectiveness.

In future work, for tuning the performance of the adversary model, one interesting direction is to exploit the geo-tagged photos on social platforms, together with the budget-limited crowdsourcing annotation, to efficiently infer the locations of the seed photos. The benefits of calibrating cluster number and per-cluster annotation number for balancing clustering performance and annotation accuracy are also worth investigating. On the other hand, from the defense side, we note that the cloud platform may figure out the regions (not the locations) those active users frequently paying visits to, since their task participation information is known to the cloud. To avoid user profiling and unexpected leakage through such prior knowledge, differential privacy techniques can be introduced to obfuscate the binary participation histories.

ACKNOWLEDGMENTS

This work is supported by the National Key Research and Development Program of China (2020YFC2003400), the National Natural Science Foundation of China (No. 62072465), and the NUDT Research Grants (No. ZK19-38).

REFERENCES

[1] Björn Barz, K. Schröter, Moritz Münch, B. Yang, A. Unger, D. Dransch, and Joachim Denzler. 2019. Enhancing Flood Impact Analysis using Interactive Retrieval of Social Media Images. *ArXiv* abs/1908.03361 (2019).

[2] Joan-Isaac Biel, Nathalie Martin, David Labbe, and Daniel Gatica-Perez. 2018. Bites'N'Bits: Inferring Eating Behavior from Contextual Mobile Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4 (2018). https://doi.org/10.1145/3161161

[3] D. Boneh, C. Gentry, B. Lynn, and H. Shacham. 2003. Aggregate and Verifiably Encrypted Signatures from Bilinear Maps. In *EUROCRYPT*. 416–432.

[4] Hancheng Cao, Jie Feng, Yong Li, and Vassilis Kostakos. 2018. Uniqueness in the city: Urban morphology and location privacy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–20.

[5] Huihui Chen, Bin Guo, Zhiwen Yu, and Han Qi. 2016. Toward real-time and cooperative mobile visual sensing and sharing. In *Proc. of IEEE Infocom*. 1–9.

[6] J. Choi, M. Larson, Xinchao Li, K. Li, G. Friedland, and A. Hanjalic. 2017. The Geo-Privacy Bonus of Popular Photo Enhancements. In *Proc. of the 2017 ACM on International Conference on Multimedia Retrieval*. 84–92.

[7] A. Costanzo, Irene Amerini, R. Caldelli, and M. Barni. 2014. Forensic Analysis of SIFT Keypoint Removal and Injection. *IEEE Transactions on Information Forensics and Security* 9 (2014), 1450–1464.

[8] Angelo De Caro and Vincenzo Iovino. 2011. jPBC: Java pairing based cryptography. In *Proc. of the IEEE symposium on computers and communications (ISCC)*. IEEE, 850–855.

[9] Quan Fang, J. Sang, and C. Xu. 2013. GIANT: geo-informative attributes for location recognition and exploration. In *Proc. of the the 21st ACM International Conference on Multimedia (MM)*. 13–22.

[10] Anhong Guo, Anuraag Jain, Shomiron Ghose, Gierad Laput, C. Harrison, and Jeffrey P. Bigham. 2018. Crowd-AI Camera Sensing in the Real World. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2 (2018), 1–20.

[11] Bin Guo, Chao Chen, Daqing Zhang, Z. Yu, and Alvin Chin. 2016. Mobile crowd sensing and computing: when participatory sensing meets participatory social media. *IEEE Communications Magazine* 54 (2016), 131–137.

[12] Bin Guo, H. Chen, Z. Yu, X. Xie, Shenlong Huangfu, and Daqing Zhang. 2015. FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing. *IEEE Transactions on Mobile Computing* 14 (2015), 2020–2033.

[13] Bin Guo, Y. Ouyang, Cheng Zhang, Jiafan Zhang, Z. Yu, Di Wu, and Yu Wang. 2017. CrowdStory: Fine-Grained Event Storyline Generation by Fusion of Multi-Modal Crowdsourced Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1 (2017), 55:1–55:19.

[14] James Hays and Alexei A. Efros. 2008. IM2GPS: estimating geographic information from a single image. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–8.

[15] Kuan Lun Huang, Salil S. Kanhere, and W. Hu. 2009. Towards privacy-sensitive participatory sensing. In *Proc. of the IEEE International Conference on Pervasive Computing and Communications*. 1–6.

[16] B. Ionescu, A. Gînsca, Bogdan Boteanu, M. Lupu, Adrian Popescu, and H. Müller. 2016. Div150Multi: a social image retrieval result diversification dataset with multi-topic queries. In *Proc. of the 7th International Conference on Multimedia Systems*. ACM, 1–6.

[17] Haiming Jin, Lu Su, Houping Xiao, and Klara Nahrstedt. 2018. Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems. *IEEE/ACM Transactions on Networking* 26, 5 (2018), 2019–2032.

[18] X. Jin, Rui Zhang, Y. Chen, Tao Li, and Yanchao Zhang. 2016. DPSense: Differentially Private Crowdsourced Spectrum Sensing. In *Proc. of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 296–307.

[19] Thivya Kandappu, Archan Misra, Shih-Fen Cheng, Randy Tandriansyah, and Hoong Chuin Lau. 2018. Obfuscation at-source: Privacy in context-aware mobile crowd-sourcing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–24.

[20] H. Kim, E. Dunn, and J. Frahm. 2017. Learned Contextual Feature Reweighting for Image Geo-Localization. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3251–3260.

[21] Pengfei Li, Hua Lu, Nattiya Kanhabua, Sha Zhao, and Gang Pan. 2018. Location inference for non-geotagged tweets in user timelines. *IEEE Transactions on Knowledge and Data Engineering* 31, 6 (2018), 1150–1165.

[22] Chi Liu, Tianqing Zhu, Jun Zhang, and Wanlei Zhou. 2020. Privacy Intelligence: A Survey on Image Sharing on Online Social Networks. *arXiv preprint arXiv:2008.12199* (2020).

[23] A. Ng, Michael I. Jordan, and Yair Weiss. 2001. On Spectral Clustering: Analysis and an algorithm. In *NIPS*.

[24] Ben Niu, Yahong Chen, Zhibo Wang, Boyang Wang, and Hui Li. 2020. Eclipse: Preserving Differential Location Privacy Against Long-Term Observation Attacks. *IEEE Transactions on Mobile Computing* (2020).

[25] S. Papadopoulos, Christos Zigkolis, Y. Kompatsiaris, and Athena Vakali. 2011. Cluster-Based Landmark and Event Detection for Tagged Photo Collections. *IEEE MultiMedia* 18 (2011), 52–63.

[26] J. Philbin, O. Chum, M. Isard, Josef Sivic, and Andrew Zisserman. 2007. Object retrieval with large vocabularies and fast spatial matching. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–8.

[27] Layla Pournajaf, Daniel A. Garcia-Ulloa, Li Xiong, and V. S. Sunderam. 2016. Participant Privacy in Mobile Crowd Sensing Task Management: A Survey of Methods and Challenges. *SIGMOD Rec.* 44 (2016), 23–34.

[28] D. Quercia, N. O'Hare, and Henriette Cramer. 2014. Aesthetic capital: what makes london look beautiful, quiet, and happy?. In *Proc. of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW).* 945–955.

[29] Olga Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Zhiheng Huang, A. Karpathy, A. Khosla, Michael S. Bernstein, A. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115 (2015), 211–252.

[30] Philip Salesses, Katja Schechtner, and C. Hidalgo. 2013. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE* 8 (2013).

[31] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition.* 4510–4520.

[32] Peng Sun, Zhibo Wang, Yunhe Feng, Liantao Wu, Yanjun Li, Hairong Qi, and Zhi Wang. 2020. Towards personalized privacy-preserving incentive for truth discovery in crowdsourced binary-choice question answering. In *Proc. of the IEEE Conference on Computer Communications (Infocom).* IEEE, 1133–1142.

[33] L. Sweeney. 2002. k-Anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 10 (2002), 557–570.

[34] Yicong Tian, C. Chen, and M. Shah. 2017. Cross-View Image Matching for Geo-Localization in Urban Environments. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1998–2006.

[35] H. To, Gabriel Ghinita, and C. Shahabi. 2014. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. *Proc. VLDB Endow.* 7 (2014), 919–930.

[36] Idalides J. Vergara-Laurens, D. Mendez, and M. Labrador. 2014. Privacy, quality of information, and energy consumption in Participatory Sensing systems. In *Proc. of the IEEE International Conference on Pervasive Computing and Communications (PerCom).* 199–207.

[37] Boyang Wang, B. Li, and H. Li. 2014. Oruta: privacy-preserving public auditing for shared data in the cloud. *IEEE Transactions on Cloud Computing* 2 (2014), 43–56.

[38] Leye Wang, Gehua Qin, D. Yang, Xiao Han, and Xiaojuan Ma. 2018. Geographic Differential Privacy for Mobile Crowd Coverage Maximization. In *Proc. of the AAAI Conference on Artificial Intelligence (AAAI).* 1–10.

[39] Leye Wang, D. Yang, Xiao Han, Tianben Wang, Daqing Zhang, and Xiaojuan Ma. 2017. Location Privacy-Preserving Task Allocation for Mobile Crowdsensing with Differential Geo-Obfuscation. In *Proc. of the 26th International Conference on World Wide Web.* 627–636.

[40] Leye Wang, Daqing Zhang, Yasha Wang, C. Chen, X. Han, and A. M'hamed. 2016. Sparse mobile crowdsensing: challenges and opportunities. *IEEE Communications Magazine* 54 (2016), 161–167.

[41] Leye Wang, Daqing Zhang, D. Yang, Brian Y. Lim, X. Han, and Xiaojuan Ma. 2020. Sparse Mobile Crowdsensing With Differential and Distortion Location Privacy. *IEEE Transactions on Information Forensics and Security* 15 (2020), 2735–2749.

[42] Xiong Wang, Lei Ding, Qi Wang, Jin Xie, Tianyi Wang, Xiaohua Tian, Yunfeng Guan, and Xinbing Wang. 2017. A Picture is Worth a Thousand Words: Share Your Real-Time View on the Road. *IEEE Transactions on Vehicular Technology* 66 (2017), 2902–2914.

[43] Yufei Wang, Zhe Lin, Xiaohui Shen, Radomir Mech, Gavin Miller, and Garrison. W. Cottrell. 2016. Event-specific Image Importance. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[44] Tobias Weyand, Ilya Kostrikov, and J. Philbin. 2016. PlaNet - Photo Geolocation with Convolutional Neural Networks. In *Proc. of ECCV.* 37–55.

[45] Tobias Weyand and B. Leibe. 2015. Visual landmark recognition from Internet photo collections: A large-scale evaluation. *Comput. Vis. Image Underst.* 135 (2015), 1–15.

[46] Dingqi Yang, Benjamin Fankhauser, Paolo Rosso, and Philippe Cudre-Mauroux. 2020. Location Prediction over Sparse User Mobility Traces Using RNNs: Flashback in Hidden States. In *Proc. of IJCAI.* 2184–2190.

[47] Jinghan Yang, A. Chakrabarti, and Y. Vorobeychik. 2020. Protecting Geolocation Privacy of Photo Collections. In *Proc. of the AAAI.* 524–531.

[48] Mengyuan Zhang, Lei Yang, Shibo He, Ming Li, and Junshan Zhang. 2021. Privacy-Preserving Data Aggregation for Mobile Crowdsensing With Externality: An Auction Approach. *IEEE/ACM Transactions on Networking* (2021).

[49] Richard Zhang, Phillip Isola, Alexei A. Efros, E. Shechtman, and O. Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 586–595.

[50] Zhikun Zhang, Shibo He, Jiming Chen, and Junshan Zhang. 2018. REAP: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing. *IEEE Transactions on Information Forensics and Security* 13, 12 (2018), 2995–3007.

[51] Tongqing Zhou, Zhiping Cai, Bin Xiao, Leye Wang, Ming Xu, and Yueyue Chen. 2018. Location Privacy-Preserving Data Recovery for Mobile Crowdsensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (2018), 1–23.

[52] Tongqing Zhou, Bin Xiao, Zhiping Cai, and Ming Xu. 2021. A Utility Model for Photo Selection in Mobile Crowdsensing. *IEEE Transactions on Mobile Computing* 20 (2021), 48–62.