From Uncertain Photos to Certain Coverage: a Novel Photo Selection Approach to Mobile Crowdsensing

Tongqing Zhou*†, Bin Xiao†, Zhiping Cai*, Ming Xu*, Xuan Liu‡§
*College of Computer, National University of Defense Technology, China. Email:{zhoutongqing, zpcai, xuming}@nudt.edu.cn
†Department of Computing, The Hong Kong Polytechnic University, Hong Kong. Email: csbxiao@comp.polyu.edu.hk
‡College of Computer Science and Electronic Engineering, Hunan University, China. Email: xuan_liu@hnu.edu.cn
§State Key Laboratory for Novel Software Technology, Nanjing University, China.

***Abstract*—Traditional mobile crowdsensing photo selection process focuses on selecting photos from participants to a server. The server may contain tons of photos for a certain area. A new problem is how to select a set of photos from the server to a smartphone user when the user requests to view an area (e.g., a hot spot). The challenge of the new problem is that the photo set should attain both photo coverage and view quality (e.g., with clear Points of Interest). However, contributions of these geo-tagged photos could be uncertain for a target area due to unavailable information of photo shooting direction and no reference photos. In this paper, we propose a novel and generic server-to-requester photo selection approach. Our approach leverages a utility measure to quantify the contribution of a photo set, where photos' spatial distribution and visual correlation are jointly exploited to evaluate their performance on photo coverage and view quality. Finding the photo set with the maximum utility is proven to be NP-hard. We then propose an approximation algorithm based on a greedy strategy with rigorous theoretical analysis. The effectiveness of our approach is demonstrated with real-world datasets. The results show that the proposal outperforms other approaches with much higher photo coverage and better view quality.**

I. Introduction

The proliferation of sensor intensive mobile devices and the demand for pervasive sensing lead to a new sensing paradigm, known as mobile crowdsensing (MCS) [1]. In MCS, individuals use their own mobile devices instead of specialized sensors to sense and collect real-time environment data [2].

Photo mobile crowdsensing is a widely adopted technique in the MCS paradigm due to its informativeness, and fuels a plenty of applications [3]. An illustrative example is the street view application that helps people see the city online. While industries employ a war-driving way to provide a coarse grained and infrequently updated street view, mobile crowdsensing can yield a more detailed, on-demand and specific view for an interested area with the crowd's (e.g., visitors') eyes [4][5]. A sketch of photo mobile crowdsensing process is shown in Fig. 1, where the sensed photos undergo two processes of selection. Photo *pre-selection* usually takes place among the participants during the participants-to-server stage, while photo *selection* is performed by the server during the server-to-requester stage. Most previous work focuses on the pre-selection process for servers to obtain crowdsensing data.

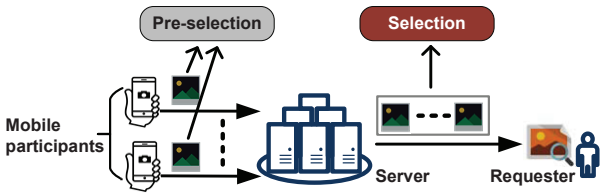


Fig. 1. The traditional participants-to-server photo pre-selection and the server-to-requester photo selection.

How to attain a proper set[1] of photos for the requester with photo selection is a new problem. Unlike the traditional pre-selection process that aims to reduce the redundancy for transmission efficiency [6][7][8], the photo selection process will find a photo set to satisfy the requester's expectation in understanding a target area. Therefore, the challenge faced in this process is that the selected photo set should attain both *photo coverage* and *view quality* (a.k.a. certain coverage in this work). Namely, the requester favors a photo set that captures as many Points of Interest (PoIs) of the target area with clear and accurate view (i.e., not a blocked, blurred, or wrong direction view). In fact, tons of photos are still aggregated to the server even with pre-selection. As a result, a selection process to transmit limited photos to the requester, if not being carefully performed, would ruin the informativeness of mobile crowdsensing, and degrade requester experience.

Traditional participants-to-server pre-selection approaches may not be helpful to handle the new challenge in the selection process as they fail to consider the uncertainty properties of crowdsensing photos. From the aspect of photo coverage, existing approaches (e.g., [7][9]) require varied geometric data (including location, shooting direction, etc.) to be provided during a metadata collection stage, and use these information to formulate exact coverage model for collected photos.

[1]In fact, a photo subset of the raw photo collection is selected. We use photo set for short in this paper.

However, the required shooting direction information is in fact not accessible via the built-in camera applications of mobile devices. The photo coverage of an area becomes unsure with uncertain directions, since photos at a location may have different views when facing different directions. As to view quality, previous approaches [10][11][12] in MCS usually assume the existence of a quality criterion for the sensory data such that they could use data analysis techniques (i.e., EM, clustering) to classify the collection. However, the view quality of photos is specific to the potential PoIs in the target area. It is difficult to set up certain photo criteria with no reference photos. Therefore, the photo selection for certain coverage remains challenging in terms of the photo uncertainty.

In this paper, we propose a novel photo selection approach to the new server-to-requester selection problem in mobile crowdsensing. The approach takes crowdsensing geo-tagged photos in the server as inputs, and selects a representative set of photos for offloading. The geo-tagged photos have simple location information but without any additional metadata. During the selection process, we should be able to assess the coverage capability of a photo set with merely photos' location information, and guarantee photos view quality without any reference photos.

In view of these challenges, this paper first designs a utility measure to quantify the contribution of a selected photo set. The measure leverages two factors to quantify photo coverage and view quality, respectively. First, a spatial factor exploits the distribution entropy of photo locations to meet the coverage expectation. The underlying idea is that two photos located in different subareas capture either two different PoIs or different aspects of one PoI. By finding an uniformly distributed photo set, we would obtain a better coverage on potential PoIs. Second, a content factor is designed based on the fact that useful photos of one PoI are similar while ones with awful view quality are distinct. Then we exploit the visual correlation between a photo set and its complementary set, and use the correlation degree as a metric for the set's view quality (i.e., a set containing more useful photos should show a larger correlation with its complementary set, and vice versa). By jointly considering these two factors, our approach then selects the photo set with the maximum utility (greedy algorithm) to obtain a certain coverage.

Our approach only needs the generalized location metadata and the visual content in the selection process, which makes it immune to the photo uncertainty properties. We conduct experiments on real-world datasets to evaluate the performance of our design and demonstrate its effectiveness. The main contributions of this work include:

1) We analyze the certain coverage challenge and photo uncertainty properties of a new server-to-requester photo selection problem in mobile crowdsensing. To the best of our knowledge, it is the first attempt to attain good photo coverage of a target area from uncertain crowdsensing photos.

2) Our approach leverages a novel utility measure for photo coverage and view quality. It is effective even with simple metadata information (i.e., GPS location). Finding the maximum utility is proven to be NP-hard, and a $(1\text{-}1/e)$ approximation greedy algorithm is proposed accordingly.

3) We evaluate the approach with two real-world photo datasets. Experimental results show that our proposal outperforms existing clustering-based approach [6] and a random selection approach by an average of 87% and 84% on photo coverage, and 35% and 32% on view quality, respectively.

II. Photo Selection Process

A. The Goal of Photo Selection

During photo mobile crowdsensing, photo selection is performed by the server to obtain a photo set with certain coverage considering the offloading constraint. In Fig. 2, we use an illustrative example to describe this selection process in a MCS task. Briefly, the downward steps prepare a photo collection for selection, and the upward steps perform the selection strategy and transmit the results to the requester.

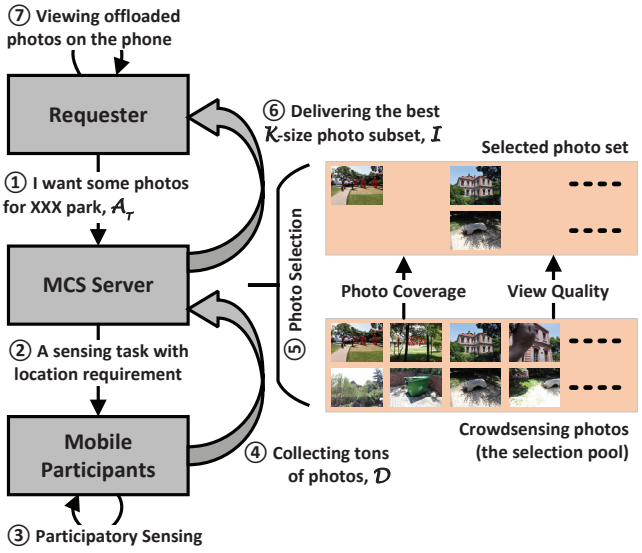


Fig. 2. An example of typical photo mobile crowdsensing task in a park scenario.

Initially, the requester releases a sensing task regarding a target area via the MCS server. Mobile crowd can choose to join the task (participatory sensing with or without monetary incentives [1]), and use the built-in camera applications on their own mobile devices to collect and report photos of PoIs in the area. In fact, the submission can be either a photo just captured or a relevant photo stored in the album. Each photo p_i is tagged with the location L_i where it is taken, which can be extracted from its EXIF header[2].

The server would collect tons of photos with even only one photo from each participant. Offloading all these photos is not efficient for the requester, so a representative subset is preferred. As shown in the 5th step in Fig. 2, the collected

[2]Location annotation is an available function of the built-in camera applications in both Android and iOS systems. In contrast, the shooting direction information required in [7][13][9] cannot be obtained with the camera applications.

photos together capture a group of PoIs and have varied view quality (e.g., clear view, blocked view). Thus, in order to satisfy requester's expectation in understanding the area, the selection process should attain both photo coverage to capture as many PoIs and view quality to provide useful summations (i.e., certain coverage).

B. Challenges in Photo Selection

We formulate the photo selection problem for certain coverage and describe the challenges as follows. We denote $\mathcal{D}$=$\{p_1, ..., p_n\}$ as the raw crowdsensing photo collection, K as the maximum number of photos that could be offloaded, and $\mathcal{I}$ ($\mathcal{I} \subset \mathcal{D}$, $|\mathcal{I}| \leq K$) as the selected photo set (a.k.a. photo selection). Within the sensing area A_T, a group of uniformly distributed objects (i.e., PoIs) are captured with photos in $\mathcal{D}$, where photos of the i-th PoI constitute subset $PoI_i \subset \mathcal{D}$, and $PoIs$=$\{PoI_1, ..., PoI_m\}$. Each photo in PoI_i can be considered as a unique aspect of the i-th PoI. For a subset $\mathcal{I}$, we say an aspect of the i-th PoI is covered by $\mathcal{I}$ if we have $PoI_i \cap \mathcal{I} \neq \emptyset$.

Definition 1: **(Photo Coverage)** Assuming the PoIs in A_T are known, photo coverage considers the number and aspects of PoIs that are captured by the photos in $\mathcal{I}$, and is formally defined as,

$$C(\mathcal{I})\text{=-}N_{cov}^{\mathcal{I}} \cdot \sum_{i=1}^{|PoIs|} \frac{|\mathcal{I} \cap PoI_i|}{|PoI_i|} \cdot \log_2(\frac{|\mathcal{I} \cap PoI_i|}{|PoI_i|}) \quad (1)$$

where $|\mathcal{A}|$ denoted the number of photos in set $\mathcal{A}$, and $N_{cov}^{\mathcal{I}}$ is the number of PoIs covered by $\mathcal{I}$. Obviously, a photo set that can cover as many PoIs uniformly is preferred.

Definition 2: **(View Quality)** A photo is determined to be of good view quality if it does not provide a blocked (the PoI is blocked by an obstacle, e.g., the 4th photo in Fig. 2) or blurred view, or shooting at a wrong direction (no object is captured, e.g., the 6th photo in Fig. 2). Then the view quality of a selection $\mathcal{I}$ can be defined as,

$$Q(\mathcal{I})\text{=}1\text{-}\frac{N_{in}^{\mathcal{I}}}{|\mathcal{I}|} \quad (2)$$

where $N_{in}^{\mathcal{I}}$ is the number of unexpected photos with awful quality in $\mathcal{I}$.

Definition 3: **(Certain Coverage)** We define certain coverage as an integrated performance metric of photo coverage in space and view quality in content,

$$V(\mathcal{I})\text{=}(\|C(\mathcal{I})\|\text{+}\|Q(\mathcal{I})\|)/2 \quad (3)$$

where $\|\cdot\|$ denoted normalization operation.

However, challenges in finding a photo selection with best certain coverage are that $C(\mathcal{I})$ and $Q(\mathcal{I})$ cannot be directly assessed in the server due to the uncertainty properties of crowdsensing photos:

- The shooting direction information of crowdsensing photos is not recorded in the built-in camera applications of mobile devices. The distribution of potential PoIs is unknown for an uncharted sensing area. As a result, the server cannot figure out the geographical relations between $\mathcal{I}$ and PoI_i in Eq. 1.
- There are no reference photos of A_T for comparison in photo mobile crowdsensing applications. Hence, we cannot simply assess $Q(\mathcal{I})$ by comparing photos in $\mathcal{I}$ with a predefined photo criterion and counting $N_{in}^{\mathcal{I}}$.

In view of these challenges, existing approaches require the participants to provide additional metadata information [14][7] and manually verify the view quality [15]. In contrast, we design a more generic approach which leverages a novel utility measure to quantify the photo coverage and view quality performance of a photo set.

III. Novel Photo Selection Approach

Our selection approach alternatively looks for the best utility photo subset to solve the intractable selection problem. Briefly, the approach takes the crowdsensing photos as inputs, and jointly exploits the spatial and content characteristics of a photo set to facilitate a utility measure for its contribution in covering an area. Then the photo set with the maximum utility while satisfying the offloading constraint is selected and delivered to the requester. To ease the presentation, we summarize some important notations in the paper in Table I.

TABLE I
Frequently Used Notations.

A_T	the target sensing area
$\mathcal{D}$	a mobile crowdsensing photo set
$\mathcal{I}$	a selected photo set (or photo selection) for offloading
K	the maximum number of photos that can be offloaded
$p_i^{\mathcal{A}}$	a photo in set $\mathcal{A}$
$SD(\mathcal{A})$	spatial diversity of photo set $\mathcal{A}$
$CI(\mathcal{A})$	content influence of photo set $\mathcal{A}$

A. Photo Set Utility

The utility measure consists of a spatial diversity factor and a content influence factor. Spatial diversity represents the distribution of photos' locations, which implicitly measures its photo coverage capability for potential PoIs. Content influence assesses the correlation between a photo subset and its complementary set, which reflects the content representativeness of the photos in that subset regarding the rest ones. A better content representativeness indicates a better view quality. An example of the spatial distribution and content correlations of photos is illustrated in Fig. 3. We then introduce how to quantify these two factors to facilitate a utility measure in the following sections.

1) Spatial Diversity Factor: The requester would favor the selected photos to cover as many potential PoIs. Recall that a photo mobile crowdsensing task focuses on collecting visual descriptions of an area, which covers a range of objects (i.e., PoIs). For two photos located significantly different in A_T, they could either be visual summations of two PoIs or views from different aspects of one PoI. Therefore, by introducing the spatial diversity factor, we intend to select a uniformly

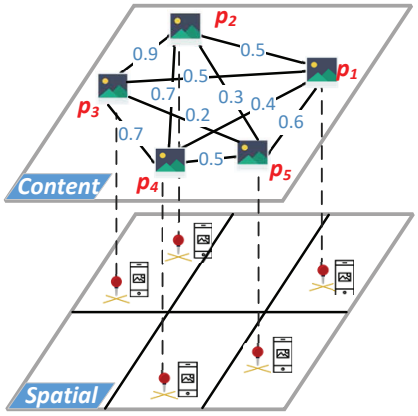


Fig. 3. An example of spatial distribution and content relations of a crowdsensing photo collection. Each photo captures a unique aspect of an object in content, and is located in a grid in space. As a result, photos are correlated with each other via their visual similarities in the content layer, and featured with their location tags in the spatial layer.

distributed photo set, which would enlarge the coverage of PoIs and their aspects during photo selection.

We propose to model the spatial diversity of a photo set as the location distribution entropy of the contained photos. The rationale of this model is that entropy owns the following properties: if the locations of a photo set is close to a uniform distribution, then the entropy is high; if the distribution is skewed (e.g., all photos located in the same cell), then the entropy is low. Without loss of generality, we regard A_T as a rectangle area, and record the ratio of its height and width as r. Then we propose to partition A_T into n_w=$\lceil|\mathcal{I}|/r\rceil$ columns horizontally, and n_h=$\lfloor|\mathcal{I}|/n_w$+0.5$\rfloor$ rows vertically. In this way, the partition granularity would be adaptively tuned to capture the uniform degree of a group of photos' locations. Given the grids, we calculate one photo set's spatial diversity with,

$$SD(\mathcal{I})\text{=-}\sum_{i=1}^{n_h\cdot n_w}\frac{n_i}{|\mathcal{I}|}\cdot\log_2(\frac{n_i}{|\mathcal{I}|}),\ n_i\neq 0 \tag{4}$$

where n_i is the number of photos in $\mathcal{I}$ taken in the i-th grid. The n_i=0 terms are ignored.

We use the PARK dataset (see Section IV-A) to examine the effectiveness of our spatial diversity model. We consider four constraint conditions with the offloading constraint ranging from 20% to 35% of the size of the raw collection. For each case, we enumerate all the possible selections[3], and calculate their prior spatial diversity (Eq. 4) and posterior photo coverage (Eq. 1). Then we sort each selection in the ascending order of its spatial diversity, and examine the mean photo coverage of selections under each diversity level. Fig. 4 shows the results where spatial diversity is normalized to range $[0, 1]$. As shown, a photo set with bigger spatial diversity shows higher photo coverage. This validates the proposal of seeking photo coverage with location distribution entropy.

[3]Enumeration is only used here to analyze the relation between spatial diversity and photo coverage.

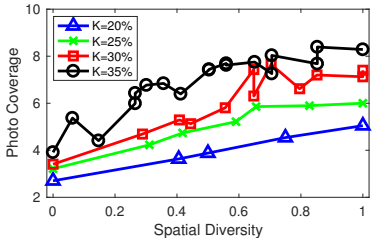


Fig. 4. The relation between spatial factor and photo coverage of a selection. K=$x\%$ means that the maximum number of offloading photos is $x\%\cdot|\mathcal{D}|$.

We further describe this factor with the spatial layer in Fig. 3. If the offloading constraint is K=3 photos, then the area is partitioned into 6 grids. During selection, $SD(\mathcal{I}_1)$ of $\mathcal{I}_1$=$\{p_1, p_2, p_3\}$ is about 0.92, while the entropy for $\mathcal{I}_2$=$\{p_1, p_2, p_5\}$ is about 1.6. From the aspect of photo coverage, we would select $\mathcal{I}_2$ in this case.

2) Content Influence Factor: The consideration of spatial diversity factor could provide a series of selections with proper photo coverage. To improve certainty on the coverage, we further adopt the content influence factor to measure the view quality of a photo set.

The content of a photo reveals a potential PoI from a unique aspect. If two photos share similarities on their contents, then they are likely to be visual descriptions of the same PoI. We denote such similarity between two photos as their content influence to each other. Intuitively, useful photos regarding one PoI are more likely to be similar with each other, while photos with low view quality (i.e., blocked, blurred, wrong direction) are supposed to have their own defects, thus distinct on their content. Based on this idea, we regard photos in a selection as 'candidates' for offloading, the rest photos in the raw collection as 'voters', and the similarity level between 'candidates' and 'voters' as the content influence of this photo selection. In this way, a photo set with higher content influence is deemed to show better view quality as more remaining photos vote for (or support) it, and vice versa. Specifically, we formulate the content influence factor based on a visual similarity metric.

We first calculate the content influence of one photo p_i regarding a set of photos $\mathcal{A}$ as,

$$INF_i^{\mathcal{A}}\text{=}\sum_{p_j\in\mathcal{A}}S(p_i,p_j) \tag{5}$$

where $S(p_i, p_j)$ is the visual similarity function, and reflects the content correlation between two photos. Many well-studied image feature extraction methods (e.g., SIFT, SURF, GIST, pHash) can be used to explore the similarities between images. Here, we propose to use SIFT to extract features and calculate the similarity between two photos as,

$$S(p_i,p_j)\text{=}\frac{|M(F_i,F_j)|}{\frac{1}{2}\cdot(|F_i|\text{+}|F_j|)} \tag{6}$$

where F_i denotes the SIFT features of p_i, and $M(F_i, F_j)$ denotes the matching level of two photos' feature spaces.

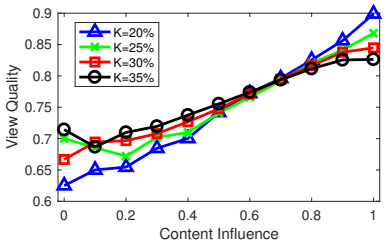


Fig. 5. The relation between content factor and view quality of a selection.

The above influence model is for one photo. For a selected photo set $\mathcal{I}$, the content influence is defines as,

$$CI(\mathcal{I})=\sum_{p_i\in\mathcal{I}} INF_i^{\mathcal{D}\text{-}\mathcal{I}} \tag{7}$$

In fact, $CI(\mathcal{I})$ measures the correlation between subset $\mathcal{I}$ and subset $\mathcal{D}$-$\mathcal{I}$. For the photos left behind in $\mathcal{D}$-$\mathcal{I}$, a stronger correlation means that the selected photo set $\mathcal{I}$ can better represent them visually, which in turn indicates the higher view quality of $\mathcal{I}$ as photos in $\mathcal{D}$-$\mathcal{I}$ give more support to it. It is straightforward to see that $CI(\mathcal{I}) = CI(\mathcal{D}\text{-}\mathcal{I})$. Given two subsets $\mathcal{I}_1,\ \mathcal{I}_2 \subset \mathcal{D}$, we also have the following properties:

$$\sum_{p_i\in\mathcal{I}_1} INF_i^{\mathcal{A}} < \sum_{p_i\in\mathcal{I}_2} INF_i^{\mathcal{A}}\ if\ \mathcal{I}_1 \subset \mathcal{I}_2 \tag{8}$$

$$\sum_{p_i\in\mathcal{I}_1\cup\mathcal{I}_2} INF_i^{\mathcal{A}}=\sum_{p_i\in\mathcal{I}_1} INF_i^{\mathcal{A}}+\sum_{p_i\in\mathcal{I}_2} INF_i^{\mathcal{A}}-\sum_{p_i\in I'} INF_i^{\mathcal{A}}$$
$$if\ \mathcal{I}_1\cap\mathcal{I}_2=\mathcal{I}' \tag{9}$$

We examine the relation between our content influence factor and view quality of a photo set. We first enumerate all the possible photo selections in the considered conditions, and calculate their prior content influence (Eq. 7) as well as posterior view quality (Eq. 2). Then the selections are sorted in the ascending order of their content influence, and we calculate the mean view quality for photo selections under each influence level for relation analysis. As shown in Fig. 5, a photo selection with higher content influence (normalized) shows better view quality. Such positive correlation validates our content influence model for view quality improving.

Taking photo set $\{p_4, p_5\}$ in the content layer of Fig. 3 as an example, it has a content influence of 2.9, which equals to the sum of similarities between set $\{p_4, p_5\}$ and set $\{p_1, p_2, p_3\}$.

3) Photo Set Utility Measure: In order to attain photo coverage and view quality simultaneously, we jointly consider spatial factor SD and content factor CI in the utility measure. Besides, we would select as many photos to cover more PoIs and their aspects if the constraint allows, so the set size is also introduced to assure the priority of a larger photo set. To this end, the utility of a photo selection is formulated as,

$$U(\mathcal{I})=(1\text{-}\alpha)\cdot\|SD(\mathcal{I})\|+\alpha\cdot\|CI(\mathcal{I})\|+|\mathcal{I}| \tag{10}$$

where SD and CI are normalized to the range of $[0, 1]$, and α is the weight characterizing and balancing the importance of the two factors.

Algorithm 1 The Greedy-based Photo Selection Algorithm

1: **procedure** GREEDY_PS($\mathcal{D}$, K, α)
2: $\mathcal{I} \leftarrow \emptyset$;
3: Compute $S(p_i, p_j)$ for each photo pair in $\mathcal{D}$;
4: **repeat**
5: **for** each photo p_i in $\mathcal{D}$-$\mathcal{I}$ **do**
6: $\Delta_{SD}^i \leftarrow SD(\mathcal{I}\cup\{p_i\})$-$SD(\mathcal{I})$;
7: $INF_i^{\mathcal{D}\text{-}\mathcal{I}\text{-}p_i} \leftarrow \sum_{p_j\in\mathcal{D}\text{-}\mathcal{I}\text{-}p_i} S(p_i, p_j)$;
8: $\Delta U(p_i) \leftarrow (1\text{-}\alpha)\cdot\|\Delta_{SD}^i\|+\alpha\cdot\|INF_i^{\mathcal{D}\text{-}\mathcal{I}\text{-}p_i}\|$;
9: **end for**
10: $\hat{p} \leftarrow \underset{p_i}{argmax}\{\Delta U(p_i)\}$;
11: $\mathcal{I} \leftarrow \mathcal{I}\cup\{\hat{p}\}$;
12: **until** $|\mathcal{I}| > K$ or $\mathcal{D}$-$\mathcal{I}$=$\emptyset$
13: return $\mathcal{I}$;
14: **end procedure**

B. Utility-based Photo Selection

Recall that the selection objective is to maximize the photo coverage and view quality of the offloading photos to the requester. Based on the constraint and utility model, we can formulate our photo selection problem as follows.

Definition 4: **(Utility-based Photo Selection Problem)** Given a mobile crowdsensing photo set $\mathcal{D}$ and an offloading budget K, the selection problem finds a subset $\mathcal{I} \subseteq \mathcal{D}$ so that utility of $\mathcal{I}$ in Eq. 10 is maximized under constraint $|\mathcal{I}| \leq K$.

Theorem 1: The utility-based photo selection problem is NP-hard.

We prove the NP-hardness of this problem by a reduction from the *Max Cut* problem. The detailed proof is provided in the appendix of this paper. Designing algorithms to find the best offloading photo subset $\mathcal{I}$ is challenging since it belongs to NP-hard. To relieve this pitfall, we propose to leverage the greedy strategy to obtain an approximate solution. Our greedy-based algorithm Greedy_PS is listed in Algorithm 1.

The basic idea of Greedy_PS is to iteratively find the best photo which yields the maximum increase on the utility (ties are broken arbitrary), and add it to the selection (Line 9-10). In each iteration, the spatial diversity gain Δ_{SR}^i and content influence of each photo p_i are calculated to estimate the utility gain (Line 5-8). The algorithm stops when either constraint is active (Line 12). The time complexity of Greedy_PS is bounded by $O(K\cdot|D|\cdot\log_2|D|)$ when merge sort is used.

Further, we use the example in Fig. 3 to explain the algorithm. We constrain that only two photos can be delivered. Initially, the spatial gain for each photo is the same, so Greedy_PS looks for the most influential photo. The sum of similarities between p_2 and $\mathcal{D} - p_2$ (the unselected subset) is the largest (i.e., 2.4), so p_2 is selected in the first round. Similarly, p_4 is picked out in the second round as $\Delta U(p_4)$ is the largest in subset $\mathcal{D} - p_2$. Thus, Greedy_PS provides a solution $\{p_2, p_4\}$, which has a spatial diversity of 0.5 and a content influence of 3.3. Note that this is not the optimal solution, in which $\{p_3, p_5\}$ is selected with the same

spatial diversity while a bigger content influence of 3.5. The approximation ratio of Greedy_PS is analyzed next.

Lemma 1: The utility of a photo set given by Eq. 10 is monotone and submodular.

Proof of this lemma can be found in the appendix.

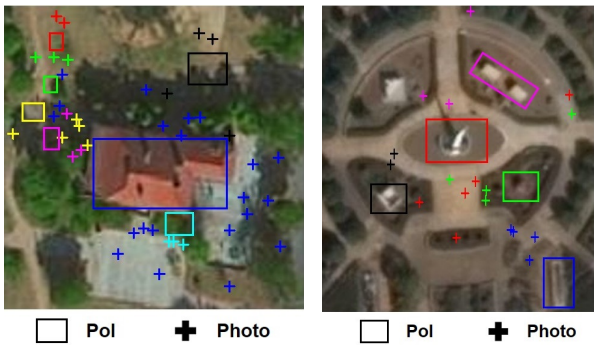
Theorem 2: The proposed greedy-based algorithm provides an approximation ratio of $(1\text{-}1/e)$, where e is the based of the natural logarithm.

Proof: According to [16], the greedy algorithm can provides a $(1\text{-}1/e)$ approximation ratio for the optimization problem of a monotone submodular function. We have proved in *Lemma 1* that the utility function in Eq. 10 posses such properties. ■

Discussion: One may argue that we can simply filter out some photos with low visual similarities to the rest, and then select photos merely based on spatial diversity. However, a photo showing weak visual correlations with others may be the only description for a PoI or a description from a unique aspect. Hence, the content factor must be considered jointly with the spatial factor to obtain a proper photo selection.

IV. Evaluation

In this section, we first examine the effectiveness of our photo selection approach, and then evaluate its performance by comparing it with two other photo selection approaches.



(a) Sensing area in a resort. (b) Sensing area in a park.

Fig. 6. PoIs distribution and collected photos' locations on the satellite images of the target areas.

A. Experimental Setup

We conduct our experiments on two real-world photo datasets collected by ourselves[4]. During the photography, we have two independent researchers take photos around the interested spots. We use the built-in camera applications of two Huawei Mate 7 phones with Android 6.0 to take photos and stored them in the albums. The localization service is turned on for the application to automatically record the location where each photo is taken. In fact, this photo collection process can be compared to the participants-to-server data collection phase of mobile crowdsensing tasks.

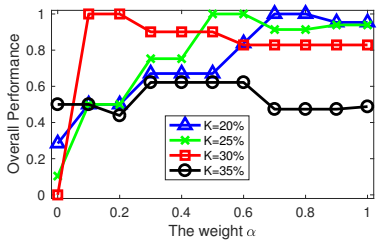
[4]As the first work devoted to photo selection for certain coverage, there lack photo datasets with location tags and view quality labels in MCS. Manually collecting photos is labor intensive, which limits the scale of our datasets.

TABLE II
Statistics of photo datasets.

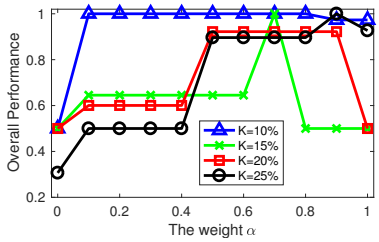
Dataset	# PoIs	# photos	Ratio of inaccurate photos
RESORT	7	43	19%
PARK	5	20	35%

Table. II shows the statistics of these two photo datasets. The RESORT dataset consists of photos collected from an islet. The distribution of PoIs and photos are depicted (photos with wrong directions are ignored) in Fig. 6(a). Totally 43 photos are taken within the interested area, and 19% of them are with awful quality. For the PARK dataset, a group of 20 photos are collected regarding 5 PoIs in a technology park as shown in Fig. 6(b). Wherein, 7 of them are inaccurate reports with either wrong shooting direction or blocked view. Typically, PoIs are uniformly distributed and each PoI is captured from several aspects with photos at different locations.

We use the photo coverage in Eq. 1, view quality in Eq. 2, and certain coverage in Eq. 3 as performance metrics during evaluation. We emphasize that merely having advantage in either $C(\mathcal{I})$ or $Q(\mathcal{I})$ is not enough, a proper selection should yield better $V(\mathcal{I})$. In order to obtain the ground truth about the coverage and view quality of each photo, we ask students who are familiar with the sensing areas to tag the photos[5].



(a) Impact analysis for the RESORT dataset.



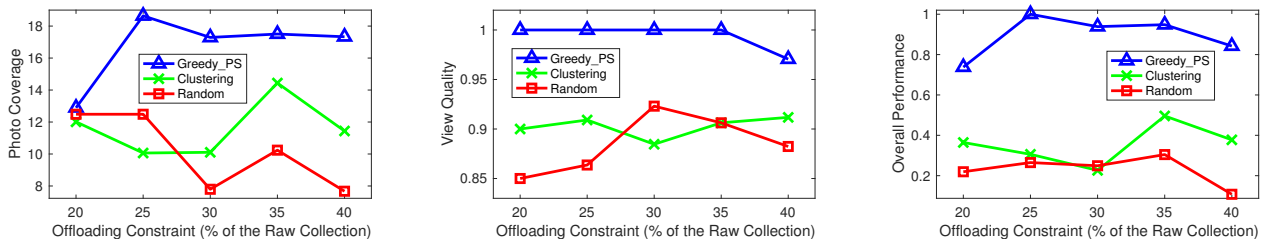
(b) Impact analysis for the PARK dataset.

Fig. 7. Impact of the weight α on the overall performance of the selection.

B. The Impact of the weights

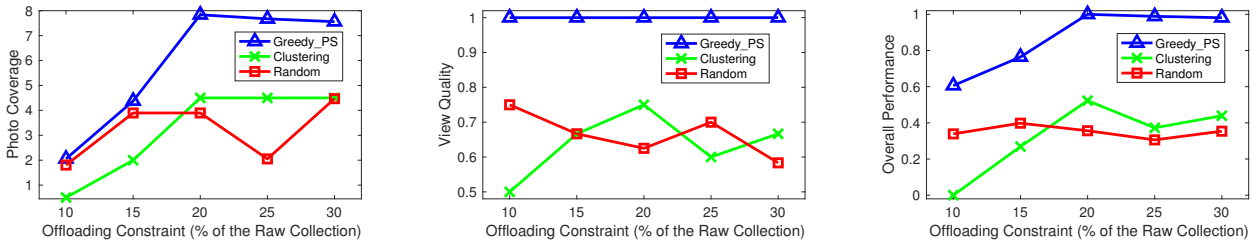
We first study the impact of the weights for the spatial diversity factor (α in Eq. 10) and content influence factor ($1\text{-}\alpha$). The weights are used to reconcile SD and CI to facilitate a utility estimation. We vary α from 0 to 1, and test the overall performance of the selected photo set our

[5]These students are not part of the research team, making them independent consultants.



(a) Photo coverage of the selection. (b) View quality of the selection. (c) Overall performance of the selection.

Fig. 8. Evaluation results for the RESORT dataset.



(a) Photo coverage of the selection. (b) View quality of the selection. (c) Overall performance of the selection.

Fig. 9. Evaluation results for the PARK dataset.

Greedy_PS algorithm generates[6]. The test is performed under different budgets (i.e., the percentages of photos allowed for offloading) for both photo datasets. Compared with the RESORT dataset, smaller budgets are assigned to the PARK dataset as it has fewer photos.

As shown in Fig. 7, for both datasets, depending only on either spatial diversity (α=0) or content influence (α=1) cannot provide satisfactory result. In contrast, jointly considering both factors ($0 < \alpha < 1$) generates a selection with better overall performance. Without loss of generality, in the rest of the experiments, we set 0.5 as the default value of α. One can easily adjust α to satisfy different application requirements (e.g., assigning a bigger value would yield better view quality).

C. Performance Evaluation

Next, we evaluate the performance of our approach on mobile crowdsensing photo selection by comparing it with the following selection approaches.

- **Clustering-based selection**. The raw photo collection is clustered into groups by using visual similarity in Eq. 6 as the distance metric (i.e., a high similarity between two photos indicates a small distance). Specifically, the first photo of each cluster is set to be the centroid. For each unclassified photo, we compare it with every cluster's centroid by calculating their visual similarity, and add it to the first cluster that presents a similarity larger than a predefined threshold. Otherwise, a new cluster is formed with that photo as the centroid [6]. Instead of setting up a specific threshold, we use the mean value of the similarities between each pair of photos in the raw collection as the threshold. Based on the clustering results, we select one photo from each cluster (clusters with more photos first) until the offloading constraint is active.
- **Random selection**. A set of photos with a size equal to the offloading amount constraint are randomly selected from the raw photo collection.

For each dataset, we choose 5 constraint levels (smaller budgets for the PARK dataset) and examine the photo coverage, view quality, and overall performance of the photo sets selected by these three schemes. Fig. 8 and Fig. 9 summarize the evaluation results.

From Fig. 8(a) and Fig. 9(a), we see that a more informative photo subset can be selected by Greedy_PS. When fewer photos are allowed for offloading, the performance on photo coverage of three approaches is similar as the number of covered PoIs is limited for them all. While with a looser constraint, Greedy_PS would more likely select photos for those uncovered PoIs. With Greedy_PS, the posterior photo coverage of its selected photo set keeps increasing linearly with the budget until nearly all PoIs are covered (i.e., K=25% for the RESORT and K=20% for the PARK). Specifically, our approach outperforms the clustering-based approach and the random one by an average of 87% and 84% on photo coverage.

As shown in Fig. 8(b) and Fig. 9(b), Greedy_PS can formally avoid the low view quality ingredients during the selection for both test scenarios, while the other two approaches have no guarantee on this metric. The improvement of our approach on view quality are 35% and 32% on average regarding the other two. Note that the selected set's view quality of Greedy_PS would degrade slightly if the budget exceeds a certain level, because it would be unavoidable to accept some noise when they become the majority of the remaining photos. Though,

[6]We have already examined the effectiveness of spatial diversity on attaining photo coverage, and content influence on improving view quality. Thus, we only test the level of certain coverage that can be achieved with varied α.

Greedy_PS would still yield a better view quality.

Fig. 8(c) and Fig. 9(c) depict the selected photo sets' overall performance of three approaches. As shown, our algorithm can provide more valuable photos to the requester than the other two approaches under different budgets. We owe the performance advantages of Greedy_PS to the jointly consideration of spatial and content factor during the selection. The clustering-based approach fails to group photos for different aspects of each PoI, and ignores the view quality expectation of the requester. Meanwhile, we notice that the performance of the random approach is not steady, and generally worse than the clustering-based one.

Finally, the execution time of our approach is mainly occupied by the photo analysis (i.e., feature extraction) process, while the photo selection time is quite small due to the greedy-based strategy. Specifically, in our two test cases, the photo analysis time are $71s$ and $30s$, while both taking a selection time of less than $0.1s$. We emphasize that photo analysis can be performed during photo collection to reduce this part of time cost, and it only has to be performed once for each task.

V. Related Work

We summarize relevant work of photo selection for photo coverage and view quality separately, and discuss the differences from our work.

Existing efforts on photo coverage during photo selection can be classified into two categories: context-based approaches and clustering-based approaches. The context-based approaches refer to the photo metadata to obtain a representative photo set [7][9]. In [7], the value of photos is quantified by their expected spatial coverage, and the selection process attempts to find photos that yield the maximum coverage. Guo et al. [9] use multi-dimension metadata (i.e., light, shooting angle) to extract the most representative flier photo set for repost. However, the required geometric information in these models is in fact difficult to obtain as the built-in camera applications on mobile devices would only record photos' locations. It would undoubtedly discourage the participation of the mobile crowd by requiring them to install additional applications for information collection. The clustering-based approaches propose to group (i.e., k-means) a given photo collection into clusters based on either content or location [6][17][18]. Then first arrival (i.e., selecting the 1st photo) [6], semantic tag-based [17], or centroid-photo similarity-based [18] strategy is used to select certain number of photos from each cluster. We point out that it is not scalable to setup some thresholds (e.g., number of clusters) for clustering in every crowdsensing case.

There exists many work for data quality estimation in MCS [10][11][12]. These schemes focus on the analysis of structured data (i.e., decimal measurements, or binary observations), which are not applicable to the unstructured photos. For example, we can fit decimal data into a normal distribution to filter out the outliers, while we cannot analyze the photos with a statistical distribution. Crowdsourcing-based photo quality verification is adopted in [15] to assure that an accurate result is returned during the photo searching service. However, no reference photo is available for either automatically or manually verification in our server-to-requester selection. A photo quality assessment technique is proposed in [19]. This technique focuses on picking out the photos with distortion caused by light intensity or motion, while we pay more attention to avoid photos with blocked views and wrong shooting directions.

Different from the previous work, for the first time, we consider the server-to-requester photo selection problem, and attempt to attain photo coverage and view quality based on uncertain photos. We argue that existing approaches cannot be simply combined or directly applied to our scenarios due to the uncertainty properties of crowdsensing photos. That is, we manage to solve a more generic problem without requiring either photos' shooting directions or photo references.

VI. Conclusion

In photo mobile crowdsensing applications, the server collects tons of photos of a target area, and delivers a representative photo set to the requester to view the area online. In this paper, we first study the server-to-requester photo selection problem and analyze the certain coverage challenges in terms of photo uncertainty. In view of these challenges, we design a utility measure to quantify the contribution of a photo set in terms of photo coverage and view quality based on a spatial factor and a content factor. Then we propose a greedy-based algorithm to select the photo set with the approximately maximum utility. In this way, our approach can formally bridge the gap between uncertain crowdsensing photos and certain coverage of a target area. We conduct extensive experiments on two real-world photo datasets. Experimental results demonstrate the effectiveness of our proposal, and show the performance advantages over a clustering-based approach and a random selection approach. In the future, we will investigate the effective offloading of crowdsensing short videos.

Appendix

Proof of Theorem 1:

We prove the NP-hardness of the utility-based photo selection problem by reducing the *Max Cut* problem to it.

An instance of *Max Cut* problem involves a graph $G = (\mathcal{V}, \mathcal{E})$, in which each edge (u, v) has a weight ω_{uv}. The optimization objective is to find a cut $(\mathcal{S}, \mathcal{V}\text{-}\mathcal{S})$ that maximizes $\sum_{u \in \mathcal{S}, v \in \mathcal{V}\text{-}\mathcal{S}, (u,v) \in \mathcal{E}} \omega_{uv}$.

Our reduction regards $\mathcal{V}$ as $\mathcal{D}$, where each vertex of G becomes a photo in $\mathcal{D}$. We construct the similarities between each pair of photos with the weight of each edge in $\mathcal{E}$ (0 if no edge exists between two vertex), and set $|\mathcal{I}|$=$|\mathcal{S}|$. In this way, finding the maximum cut $(\mathcal{S}, \mathcal{V}\text{-}\mathcal{S})$ is equivalent to selecting $|\mathcal{S}|$ photos so that the content influence to the rest photos is maximized. Since this process can be done in polynomial time, we prove that maximizing content influence under the constraint, which is a special case of the selection problem (i.e., α=1), is NP-hard. Therefore, we conclude that the utility-based photo selection problem is NP-hard. ■

Proof of Lemma 1:

We define the two properties, namely, monotonicity and submodularity regarding set function as follows.

Definition 5: **(Monotonicity)** A function $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ is monotone if for every $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, $f(\mathcal{A}) \leq f(\mathcal{B})$.

Definition 6: **(Submodularity)** A function $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ is submodular if for every $\mathcal{A},\ \mathcal{B} \subseteq \mathcal{V}$, $f(\mathcal{A})+f(\mathcal{B}) \geq f(\mathcal{A} \cup \mathcal{B})+f(\mathcal{A} \cap \mathcal{B})$.

Since the linear combination of monotone submodular functions is still monotone and submodular, we prove the monotonicity and submodularity of the two terms $SD(\mathcal{I})$ and $CI(\mathcal{I})+|\mathcal{I}|$ in the utility function separately. For simplicity, we ignore the weights and normalization symbol of SD and CI during the proof.

Recall that we formulate $SD(\mathcal{I})$ as the one dimension entropy of the distributed locations in the sensing area. The Shannon entropy is shown to be monotone submodular in [20]. Hence, $SD(\mathcal{I})$ is a monotone submodular function.

As to $CI(\mathcal{I})+|\mathcal{I}|$, we first prove the monotonicity. Given two subsets $\mathcal{I}_1 \subseteq \mathcal{I}_2$ (we assume $\exists \mathcal{I}' \subset \mathcal{D},\ \mathcal{I}_1 \cup \mathcal{I}'=\mathcal{I}_2$) of $\mathcal{D}$, we have,

$$\begin{aligned} & CI(\mathcal{I}_2)+|\mathcal{I}_2|-(CI(\mathcal{I}_1)+|\mathcal{I}_1|) \\ =& CI(\mathcal{I}_2)-CI(\mathcal{I}_1)+|\mathcal{I}'| \\ \geq& 0 \end{aligned} \tag{11}$$

where $|\mathcal{I}'| \geq 1$ and $CI(\mathcal{A}) \leq 1$ is a normalized value. Hence, function $CI(\mathcal{I})+|\mathcal{I}|$ is monotone.

Next, we prove the submodularity. Given two subsets $\mathcal{I}_1,\ \mathcal{I}_2 \subseteq \mathcal{I}$, and $\mathcal{I}'=\mathcal{I}_1 \cap \mathcal{I}_2$, we first have,

$$\begin{aligned} \mathbf{L}=& CI(\mathcal{I}_1)+|\mathcal{I}_1|+CI(\mathcal{I}_2)+|\mathcal{I}_2| \\ =& \sum_{p_i \in \mathcal{I}_1} INF_i^{\mathcal{D}-\mathcal{I}_1}+|\mathcal{I}_1|+\sum_{p_i \in I_2} INF_i^{\mathcal{D}-\mathcal{I}_2}+|\mathcal{I}_2| \end{aligned} \tag{12}$$

Then with Eq. 8 and Eq. 9, we could have,

$$\begin{aligned} \mathbf{R}=& CI(\mathcal{I}_1 \cup \mathcal{I}_2)+|\mathcal{I}_1 \cup \mathcal{I}_2|+CI(\mathcal{I}_1 \cap \mathcal{I}_2)+|\mathcal{I}_1 \cap \mathcal{I}_2| \\ =& \sum_{p_i \in \mathcal{I}_1} INF_i^{\mathcal{D}-(\mathcal{I}_1 \cup \mathcal{I}_2)}+\sum_{p_i \in \mathcal{I}_2} INF_i^{\mathcal{D}-(\mathcal{I}_1 \cup \mathcal{I}_2)} \\ & -\sum_{p_i \in \mathcal{I}'} INF_i^{\mathcal{D}-(\mathcal{I}_1 \cup \mathcal{I}_2)}+\sum_{p_i \in \mathcal{I}'} INF_i^{\mathcal{D}-\mathcal{I}'}+|\mathcal{I}_1|+|\mathcal{I}_2|-|\mathcal{I}'|+|\mathcal{I}'| \\ =& \sum_{p_i \in \mathcal{I}_1} INF_i^{\mathcal{D}-\mathcal{I}_1}-\sum_{p_i \in \mathcal{I}_1} INF_i^{\mathcal{I}_2-\mathcal{I}'}+\sum_{p_i \in \mathcal{I}_2} INF_i^{\mathcal{D}-\mathcal{I}_2} \\ & -\sum_{p_i \in \mathcal{I}_2} INF_i^{\mathcal{I}_1-\mathcal{I}'}+\sum_{p_i \in \mathcal{I}'} INF_i^{(\mathcal{I}_1 \cup \mathcal{I}_2)-\mathcal{I}'}+|\mathcal{I}_1|+|\mathcal{I}_2| \\ =& \mathbf{L}-\sum_{p_i \in \mathcal{I}_2-\mathcal{I}'} INF_i^{\mathcal{I}_1}-\sum_{p_i \in \mathcal{I}_1-\mathcal{I}'} INF_i^{\mathcal{I}_2}+\sum_{p_i \in (\mathcal{I}_1 \cup \mathcal{I}_2)-\mathcal{I}'} INF_i^{\mathcal{I}'} \\ =& \mathbf{L}+\sum_{p_i \in \mathcal{I}_1-\mathcal{I}'} INF_i^{\mathcal{I}'}-\sum_{p_i \in \mathcal{I}_1-\mathcal{I}'} INF_i^{\mathcal{I}_2}+\sum_{p_i \in \mathcal{I}_2-I'} INF_i^{\mathcal{I}'} \\ & -\sum_{p_i \in \mathcal{I}_2-\mathcal{I}'} INF_i^{\mathcal{I}_1} < \mathbf{L} \end{aligned} \tag{13}$$

where the two inequations are deduced from the photo sets' relations mentioned in Eq. 8 and Eq. 9. Hence, function $CI(\mathcal{I})+|\mathcal{I}|$ is also submodular. Since both two terms in $U(\mathcal{I})$ are monotone and submodular, we prove the *Lemma*. ■

ACKNOWLEDGMENT

This work is partially supported by NSFC grant Nos. 61379144, 61772446, 61379145, 61672195, 61402513, 61602167, and HK PolyU 4-ZZFF, and Hunan Provincial NSFC grant No. 2017JJ3037. Zhiping Cai, and Ming Xu are the corresponding authors.

REFERENCES

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] J. Wang, Y. Wang, D. Zhang, F. Wang, Y. He, and L. Ma, "PSAllocator: Multi-task allocation for participatory sensing with sensing capability constraints," in *CSCW*. ACM, 2017, pp. 1139–1151.

[3] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Computing Surveys (CSUR)*, vol. 48, no. 1, p. 7, 2015.

[4] "Geograph: geographically representative photographs of the whole Great Britain and Ireland," http://www.geograph.org.uk/.

[5] X. Wang, L. Ding, Q. Wang, J. Xie, T. Wang, X. Tian, Y. Guan, and X. Wang, "A picture is worth a thousand words: Share your real-time view on the road," *IEEE Transactions on Vehicular Technology*, 2016.

[6] B. Guo, H. Chen, Z. Yu, X. Xie, and D. Zhang, "PicPick: a generic data selection framework for mobile crowd photography," *Personal and Ubiquitous Computing*, vol. 20, no. 3, pp. 325–335, 2016.

[7] Y. Wu, Y. Wang, and G. Cao, "Photo crowdsourcing for area coverage in resource constrained environments," in *INFOCOM*. IEEE, 2017.

[8] M. Y. S. Uddin, H. Wang, F. Saremi, G.-J. Qi, T. Abdelzaher, and T. Huang, "Photonet: a similarity-aware picture delivery service for situation awareness," in *RTSS*. IEEE, 2011, pp. 317–326.

[9] B. Guo, H. Chen, Z. Yu, X. Xie, S. Huangfu, and D. Zhang, "FlierMeet: a mobile crowdsensing system for cross-space public information reposting, tagging, and sharing," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2020–2033, 2015.

[10] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng, "Truth discovery on crowd sensing of correlated entities," in *SenSys*. ACM, 2015, pp. 169–182.

[11] S. Yao, M. T. Amin, L. Su, S. Hu, S. Li, S. Wang, Y. Zhao, T. Abdelzaher, L. Kaplan, C. Aggarwal *et al.*, "Recursive ground truth estimator for social data streams," in *IPSN*. IEEE, 2016, pp. 1–12.

[12] T. Zhou, Z. Cai, K. Wu, Y. Chen, and M. Xu, "FIDC: a framework for improving data credibility in mobile crowdsensing," *Computer Networks*, vol. 120, pp. 157–169, 2017.

[13] H. Chen, B. Guo, Z. Yu, and Q. Han, "Toward real-time and cooperative mobile visual sensing and sharing," in *INFOCOM*. IEEE, 2016, pp. 1–9.

[14] Y. Wu, Y. Wang, W. Hu, X. Zhang, and G. Cao, "Resource-aware photo crowdsourcing through disruption tolerant networks," in *ICDCS*. IEEE, 2016, pp. 374–383.

[15] T. Yan, V. Kumar, and D. Ganesan, "Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones," in *MobiSys*. ACM, 2010, pp. 77–90.

[16] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions- I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[17] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: Context and content in community-contributed media collections," in *MM*. ACM, 2007, pp. 631–640.

[18] Y. Jiang, X. Xu, P. Terlecky, T. Abdelzaher, A. Bar-Noy, and R. Govindan, "Mediascope: Selective on-demand media retrieval from mobile devices," in *IPSN*. IEEE, 2013, pp. 289–300.

[19] D. Lee and K. N. Plataniotis, "Toward a no-reference image quality assessment using statistics of perceptual color descriptors," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3875–3889, 2016.

[20] A. Krause and D. Golovin, "Submodular function maximization," *Tractability: Practical Approaches to Hard Problems*, vol. 3, no. 19, p. 8, 2012.