A Utility Model for Photo Selection in Mobile Crowdsensing

Tongqing Zhou, Bin Xiao, *Senior Member, IEEE*, Zhiping Cai, *Member, IEEE*, and Ming Xu, *Member, IEEE*

Abstract—Existing mobile photo crowdsensing approaches focus on the participant-to-server photo pre-selection, i.e., reducing the photo redundancy from participants to a server. The server may still receive plenty of photos for a target area. Yet, another important problem is to select a proper photo subset of an area from the server to a requester. This is a challenging problem because the selected subset with a small size should attain both coverage on the PoIs - Points of Interest (i.e., photo coverage of the area) and quality on the views (i.e., view quality). In this paper, we propose a novel and generic server-to-requester photo selection approach even when there are neither photo shooting direction information nor reference photos. A utility model is designed to measure photo merits of coverage and quality by exploiting photos' spatial distribution and visual representativeness. We present two photo selection schemes, basic and PoI number-aware, to maximize the photo selection utility with multiple levels of granularity. Experimental results on real-world datasets show that our basic scheme outperforms the baselines by an average of 33% and 18.7% on photo coverage and view quality, respectively. Our PoI number-aware scheme can yield an additionally 44.8 percent improvement on the photo coverage performance.

Index Terms—mobile crowdsensing, photo selection, photo coverage, view quality, ubiquitous computing

1 Introduction

The prevalence of sensor intensive mobile devices and the demand for pervasive sensing led to the emergence and adoption of a new sensing paradigm, known as mobile crowdsensing (MCS) [1], [2]. Instead of relying on pre-deployed specialized sensors, MCS exploits the mobile devices of individuals to sense and collect real-time environment data [3].

Photo mobile crowdsensing is a predominant technique in the MCS paradigm due to the clarity of the visual sensing services it provides [4] and has fueled a large amount of applications [5]. A typical application scenario for photo mobile crowdsensing is to facilitate online city views (e.g., street view). Industries usually use a war-driving way or leverage passive collection [6], [7] to provide coarse-grained and infrequently updated views for a target. While with MCS, we can obtain more detailed, on-demand, and specific views with the wide-spread crowd's (e.g., visitors') eyes [8], [9], [10]. A sketch of the photo mobile crowdsensing process is shown in Fig. 1, where a participant-to-server *photo pre-selection* stage and a server-to-requester *photo selection* stage are conducted before presenting the final visual report to the requester. Existing work usually focuses on the pre-selection stage for the server to eliminate the photo redundancy and to achieve transmission efficiency [11], [12], [13], [14], [15].

Photo selection is performed on behalf of the requester with the goal of providing an informative area view, thus being a new problem. In order to help the requester to understand the target area comprehensively and clearly, the server-to-requester photo selection should attain *photo coverage* by capturing as many Points of Interest (PoIs) in the area and attain *view quality* with clear and accurate views. We name such a goal of maintaining certainty on views (i.e., not blocked, blurred, or wrong direction views) and achieving coverage on PoIs as certain coverage. In fact, plenty of photos are still aggregated to the server from the crowd even with pre-selection. As a result, the photo selection stage for the transmission of limited photos to the requester, if not being carefully performed, would ruin the effectiveness of MCS and degrade requester experience.

Existing pre-selection approaches fail to consider the uncertainty properties (e.g., a blurred or opposite-view) of crowdsensing photos, thus are not helpful to mitigate the challenges in the photo selection stage. Specifically, in order to facilitate good photo coverage, existing approaches usually leverage a variety of geometric data (especially photos' shooting direction information) to formulate exact coverage models based on the collected photos [11], [16]. However, we point out that the required shooting direction information in these approaches is in fact not recorded by the built-in camera applications of mobile devices. Since photos at a location may have different views when taken in different directions, the photo coverage of an area becomes unsure. When referring to view quality, previous data quality estimation approaches in MCS usually assume that there exists a quality criterion for the sensory data such that they could

- *T. Zhou, Z. Cai, and M. Xu are with the College of Computer, National University of Defense Technology, Changsha 410073, China. E-mail: {zhoutongqing, zpcai, xuming}@nudt.edu.cn.*
- *B. Xiao is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: csbxiao@comp.polyu.edu.hk.*

Manuscript received 22 Nov. 2018; revised 13 Aug. 2019; accepted 12 Sept. 2019. Date of publication 17 Sept. 2019; date of current version 3 Dec. 2020. (Corresponding authors: Zhiping Cai and Ming Xu.)
Digital Object Identifier no. 10.1109/TMC.2019.2941927

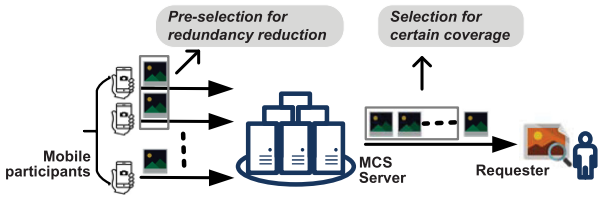


Fig. 1. An illustration of the participant-to-server photo pre-selection stage and the server-to-requester photo selection stage in a photo MCS task. These two stages are performed with different goals, photo redundancy reduction, and certain coverage of an area, respectively.

use data analysis techniques (e.g., EM [17], [18], clustering [19]) or crowdsourcing [20] to qualify the photo collection. However, a photo's view quality is related to the potential PoIs in the target area. It is hard to make certain quality criteria with no reference photos from an uncharted area. Hence, the photo selection for certain coverage is still challenging when considering the photo uncertainties.

In this paper, we propose a novel server-to-requester photo selection approach for MCS. Our approach aims at finding a representative photo subset from the geo-tagged crowdsensing photo collection. The geo-tagged photos have simple location information but without any additional metadata. Our approach can also guarantee area's view quality without any reference photos. Compared with [21], this paper make improvement by exploiting valuable context information (i.e., the PoI number information) to facilitate representative selection with better coverage.

Different photo subsets provide distinct views of an area. Some may miss important information to illustrate the target. Thus, we design a *utility model* to assess the merit of a photo subset for the requester to view an area. The model leverages two factors. The first factor is the spatial diversity, to quantify the area coverage based on the distribution entropy of photo locations. The underlying idea is that two photos located in different subareas capture either two different PoIs or different aspects of one PoI. Hence, we could improve the coverage of PoIs and their aspects by finding a uniformly distributed photo set using this factor. The second factor is the content influence, to quantify the view quality of photos. Note that useful photos of one PoI share some similarities on visual while it is opposite for useless photos. We measure the content influence of a photo subset by calculating the visual correlation between it and its complementary set. The larger the content influence is, the better the view quality that a selected photo subset has.

Our photo selection approach consists of two schemes operating on different levels of granularity based on our utility model. The basic selection scheme (BPS) attempts to find a photo subset with the maximum utility at the entire photo selection granularity, which is proven to be NP-hard with a reduction from the maximum cut problem with given sizes of parts. A greedy strategy that iteratively picks out the photo with the best utility gain is then adopted in BPS in order to generate the approximately best photo selection. Meanwhile, we consider cases where the number of PoIs is known and design a PoI number-aware photo selection scheme (PAPS) accordingly. PAPS represents photos in form of an undirected graph based on a novel graph similarity model. Then it performs BPS at the photo cluster granularity through grouping photos with the spectral clustering technique. Since photos' utility is carefully considered during selection, we can attain with BPS and PAPS good performance on certain coverage.

We highlight that our schemes only need the generalized location metadata and the visual content during selection, which makes them immune to the photo uncertainties. Extensive experiments are conducted on real-world datasets to demonstrate the effectiveness and performance of our approach. The main contributions of this work include:

1) We analyze the certain coverage challenges caused by photo uncertainties in server-to-requester photo selection for MCS. As far as we are aware, this is the first attempt to attain good photo coverage of a target area with uncertain crowdsensing photos.
2) We design a novel utility model to measure the photo coverage and view quality performance of a photo set.
3) We propose two schemes for photo selection with multiple levels of granularity. The BPS scheme selects the maximum utility photo set. The PAPS scheme exploits a novel graph similarity model for fine-grained photo selection. Finding the maximum utility in BPS is proven to be NP-hard and a $(1-1/e)$ approximation greedy algorithm is proposed accordingly.
4) We evaluate the schemes with three real-world photo datasets. Experimental results show that our BPS scheme outperforms the existing clustering-based approach [12] and the random selection approach by at least 23 and 16 percent on photo coverage and at least 12 percent and 6 percent on view quality, respectively. Meanwhile, our PAPS scheme yields an average of 44.8 percent better photo coverage than BPS in the PoI number-aware cases.

The rest of the paper is organized as follows. Related work of photo selection in MCS is discussed in Section 2. In Section 3, we introduce the photo selection problem in MCS tasks and analyze the challenges in achieving certain coverage during photo selection. In Section 4, we present our utility model for measuring the certain coverage performance of a photo subset. We introduce the design of our basic photo selection scheme and PoI number-aware photo selection scheme in Section 5. Then we evaluate the proposal and explain the results in Section 6. Finally, we discuss on the limitations and implications of this work and conclude this paper with Sections 7 and 8.

2 Related Work

We review the relevant work in two main areas: MCS photo selection and data quality estimation in MCS. We also discuss the differences of our work.

2.1 MCS Photo Selection

Existing work on MCS photo selection mainly focuses on the participants-to-server photo selection stage, which attempts to discard redundant reports and to upload valuable photos with limited resources. The corresponding approaches can be classified into two categories based on the different strategies they adopt, namely, coverage maximization or redundancy minimization.

For coverage maximization, SmartPhoto [22] quantifies a photo's coverage as the PoI aspects it captures and selects photos that captures the most aspects. Among its follow-up

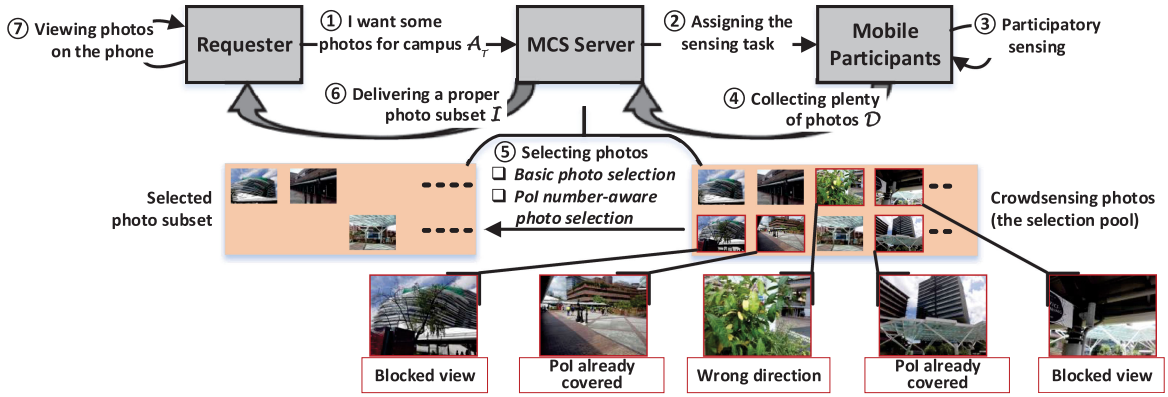


Fig. 2. A typical photo MCS task example in which the photo collection and selection process for a campus scenario is presented.

work, delivery probability of photos is considered in [23] in order to maximize the expectation of PoI aspects coverage with the selected photo subset based on delay tolerant networks (DTNs). In [11], a photo's utility is represented as a circular sector (i.e., area coverage) and the selection problem transforms into maximizing the size of the polygon that covered by the selected photos. However, the required shooting direction information in these approaches is in fact not available as the built-in camera applications on mobile devices will not record it. We point out that it will discourage the participation of the mobile users by requiring them to install additional applications for information collection.

Another strategy for participants-to-server photo selection is to minimize the redundancy in the crowdsensing photos. PhotoNet [13] reduces redundancy among delivered photos by prioritizing the transmission of photos that present diversity on time, color, and space. PhotoNet+ [24] further regards photos within certain distance as redundant and proposes to group photos into clusters based on their logical distances during photo selection. CARE [14] and Smarteye [15] use image similarity detection algorithms to eliminate similar-looking photos during data delivery. In [12], a PTree model is developed to enable photo selection based on clustering, where one photo is selected from each cluster with the rest photos in that cluster regarded as redundant reports. Focusing on redundancy elimination or the diversity properties during selection may accidently introduce outlier, namely, the irrelevant and inaccurate photos may be favored for delivery.

Different from the above work, we investigate the photo selection problem in the server-to-requester stage of the MCS tasks. Such a selection process cares more about finding a photo subset to satisfy the requester's expectation in viewing and understanding the target area than eliminating redundancy. Namely, we try to find representative photos for the requester. A closely related research field is scene summation [25] that chooses views from an online photo set for highlights. Photos that span the visual feature space [6], [25],[26] or time domain [27] are picked out during summation. In comparison, we focus on the photo coverage and view quality of a photo selection as they are the essentials for a crowdsensing task to be informative and satisfactory.

2.2 Data Quality Estimation in MCS

Lots of work has been done for estimating data quality in MCS [17], [18], [19]. However, schemes belonging to this category are usually tailored for the structured data (i.e., decimal measurements or binary observations), thus would fail to handle the unstructured photos. For example, we can pick out the high quality ingredient from the decimal data through Expectation Maximization (EM), while calculating the expectation of several photos is meaningless.

In respect of photo quality estimation, a crowdsourcing-based photo quality verification method that sends out photos with unknown quality together with reference photos is adopted in [20]. However, reference photos are not available for either automatically or manually verification of the uncharted target areas in MCS applications. In [16], [28], [29], some no-reference photo techniques are proposed by exploring the context information, such as photos' light intensity and motion blur, for quality assessment. In this work, we pay more attention to avoiding photos with blocked views and wrong shooting directions as they are not helpful for depicting the target area.

Existing approaches on MCS photo selection and data quality estimation cannot be simply combined or directly applied to our scenarios due to the uncertainty properties of crowdsensing photos. We attempt to find a high quality photo subset from the uncertain photos (i.e., photos with merely generalized information) for the requester in this paper. Finally, this work significantly extends our preliminary work in [21]. We adopt a more realistic photo coverage model by considering photos' coverage performance on PoIs' aspects. Rather than assuming absolutely unknown context information, the selection approach proposed in this paper operates on two levels of granularity by referring to the PoI number information, which yields better coverage performance. We introduce the clustering-based selection approach in [12] to evaluate the performance of the proposal and present more detailed comparison between the two proposed schemes based on three new photo datasets we collected.

3 Problem Statement

3.1 The Process of Photo Selection

As an essential stage of photo MCS, photo selection is performed by the server to obtain a photo set for the concern of the requester under the offloading constraint. We can depict the selection process of a typical photo MCS task with Fig. 2. Among all the steps, those pointing at the mobile participants facilitate visual sensing for the target and the other steps retrieve a representative photo summation to the

requester based on the photo selection strategy. Specifically, in order to campaign a photo MCS task, the requester first releases a query for visuals of a target area via the server. Mobile crowd within certain spatial boundary are recruited as participants (participatory sensing with or without monetary incentives [1]), and the built-in camera applications on their own mobile devices are used to collect photos for PoIs in the area. In fact, photos can be either just captured or records stored in the albums. Each photo p_i has a location tag L_i, which denotes where it is captured and can be easily extracted from the photo's EXIF header.[1]

A large amount of photos can be collected with even only one photo from each participant. Offloading all these photos is undoubtedly not efficient for the requester considering the huge transmission overhead and the inconvenience of viewing photos on the small screens of mobile devices. Hence, a photo selection process is conducted in the 5th step in Fig. 2 to distill a representative subset for the requester. As shown in the 4th step in Fig. 2, the crowd-contributed photos together capture a group of PoIs with varied quality on the views (e.g., clear view, blocked view). To this end, in order to satisfy the requester's expectation in understanding the area, the selection process should attain both photo coverage to cover as many PoIs and view quality to provide useful summations (i.e., certain coverage). The server can access the photos and in some cases, the PoI number information, during photo selection, which can facilitate both a basic selection strategy and a PoI number-aware selection strategy (we will discuss about this in Section. 5).

3.2 Certain Coverage and Its Challenges

We formulate the certain coverage-oriented photo selection problem and describe its 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 subset (i.e., photo selection). Within the target area A_T, a group of uniformly distributed objects (i.e., PoIs) are captured by the photos in $\mathcal{D}$. Specifically, the photos of the ith PoI constitute subset $\mathcal{S}^i_{poi}\subset\mathcal{D}$, with which we have $\bigcup_{i=1}^{n_{poi}}\mathcal{S}^i_{poi}\subseteq\mathcal{D}$ (n_{poi} is the number of PoIs captured by $\mathcal{D}$).[2] Since $\mathcal{S}^i_{poi}$ consists of photos capturing different PoI aspects (i.e., views from different directions), we further denote the photo subset that belongs to the jth aspect of the ith PoI as $\mathcal{S}^{ij}_{asp}$. Obviously, we can have $\bigcup_{j=1}^{n^i_{asp}}\mathcal{S}^{ij}_{asp}=\mathcal{S}^i_{poi}$ (n^i_{asp} is the number of different aspects for the ith PoI).

Definition 1 (Photo Coverage). *Photo coverage quantifies how well the PoIs are geographically captured by the photos in selection $\mathcal{I}$ and is formulated as,*

$$C(\mathcal{I})=-\sum_{i=1}^{n_{poi}}\sum_{j=1}^{n^i_{asp}}g(\mathcal{I}\cap\mathcal{S}^{ij}_{asp})\cdot\sum_{i=1}^{n_{poi}}(R_i\cdot\log_2 R_i). \quad (1)$$

The first term of Eq. (1) denotes the number of covered PoI aspects with $g(\mathcal{A})=\begin{cases}0, & if\ \mathcal{A}=\emptyset\\ 1, & otherwise\end{cases}$. *The second term is the entropy of photo distribution for each PoI, where* $R_i=\frac{|\mathcal{I}\cap\mathcal{S}^i_{poi}|}{|\mathcal{I}|}$ *and* $|\mathcal{A}|$ *is the number of photos in set* $\mathcal{A}$.

A photo set that can cover as many aspects of the PoIs uniformly has a large photo coverage and is preferred during selection.

Definition 2 (View Quality). *View quality quantifies the view acceptance level of the photos in selection $\mathcal{I}$. Generally, a photo is determined to have a 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 8th photo in Fig. 2). Then the view quality of a selection $\mathcal{I}$ can be defined as,*

$$Q(\mathcal{I})=1-\frac{n^{\mathcal{I}}_{in}}{|\mathcal{I}|}, \quad (2)$$

where $N^{\mathcal{I}}_{in}$ is the number of unexpected photos with low quality in $\mathcal{I}$.

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

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

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

However, it is hard (if not impossible) to find the photo selection with the best certain coverage from the raw collection. The challenges 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. Meanwhile, the distribution of potential PoIs (i.e., PoIs' locations) 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 usually no reference photos for the uncharted area in a photo MCS task. Hence, we cannot simply assess $Q(\mathcal{I})$ by comparing photos in $\mathcal{I}$ with a predefined visual criterion and calculating the number $N^{\mathcal{I}}_{in}$.

Facing these challenges, existing approaches require the participants to provide additional metadata information to estimate photo coverage [11], [23] and to verify view quality manually [20]. In order to provide a generic and scalable approach, this work designs a novel utility measure based on merely the generalized location metadata and the visual content, and uses it to quantify the certain coverage level of a photo subset during photo selection. To ease the presentation, we summarize some important notations used in the paper in Table 1.

1. 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 [11], [16], [27] cannot be obtained with the camera applications.

2. There may be some inaccurate photos in $\mathcal{D}$ which capture no PoI.

TABLE 1
Frequently Used Notations

$\mathcal{D}$	a MCS photo collection
$\mathcal{I}$	a selected photo set (or photo selection) for offloading
K	the maximum number of photos that can be offloaded
$F_s(\mathcal{A})$	spatial diversity factor of photo set $\mathcal{A}$
$F_c(\mathcal{A})$	content influence factor of photo set $\mathcal{A}$
$U(\mathcal{A})$	the utility (merit) of a photo set

4 Photo Set Utility Model

In this section, we design a novel photo set utility model for certain coverage estimation of the uncertain crowdsensing photos. There are two factors in the model: 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 one subset for the other ones. Better content representativeness basically indicates better performance on view quality.

An example of the spatial distribution and the visual correlations of five photos is illustrated in Fig. 3. As shown, spatial diversity and content influence are estimated in the spatial layer and the content layer separately. We will present the calculation process for these two factors following their formulation next.

4.1 The Spatial Diversity Factor

Intuitively, a photo set that captures as many potential PoIs from different directions is favored by the requester. However, we cannot directly estimate how many PoIs and aspects are included as the shooting direction information is not available. During a photo MCS task, visual descriptions of a range of objects (i.e., PoIs) are collected. We highlight that for two photos located differently in A_T, they could either be visual summations of two PoIs or views from different aspects of one PoI. To this end, a photo subset with photos' location uniformly distributed shall be a good selection as photos of the same aspect are avoided and photos of different PoIs are very likely selected.

Based on the discussion above, we model the spatial diversity of a photo set as the location distribution entropy of the photos in it. The rationale behind this model is that entropy has the following properties: if the locations of a photo set are 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. Namely, the entropy is suitable for qualifying the uniform degree of the selected photos' locations. Without loss of generality, we regard A_T as a rectangle area and define the ratio of its height and width as r. Then we propose to partition A_T into $m_1=\lceil(|\mathcal{I}|/r)^{1/2}\rceil$ columns horizontally and $m_2=\lfloor|\mathcal{I}|/n_w+0.5\rfloor$ rows vertically. In this way, the partition granularity will be adaptively tuned according to the size of the target area and the size of the selection. After generating the grids, we can calculate a photo set's spatial diversity with,

$$F_s(\mathcal{I})=-\sum_{i=1}^{m_1\cdot m_2}\frac{n_i}{|\mathcal{I}|}\cdot\log_2(\frac{n_i}{|\mathcal{I}|}),\ n_i\neq 0 \tag{4}$$

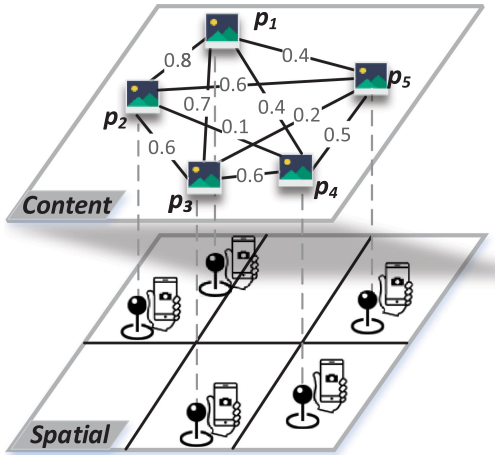


Fig. 3. An example of spatial distribution and content relations of a crowdsensing photo collection. Each photo captures a unique aspect of an object in content and is located in a grid in space. As a result, photos are correlated with each other via their visual similarities (i.e., weight on the edge)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

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

We further use the case in Fig. 3 to illustratively show the estimation of the spatial diversity factor. Assuming the offloading constraint is $K=3$ photos, then the area would be partitioned into 6 grids. Given two possible photo selection $\mathcal{I}_1=\{p_1,p_2,p_5\}$ and $\mathcal{I}_2=\{p_1,p_4,p_5\}$, their spatial diversity are around 0.9 and 1.6 according to Eq. (4). From the aspect of photo coverage, we will select $\mathcal{I}_2$ to enlarge the coverage. Note that this toy example only shows the preference for more covered grids, while spatial diversity also cares about the uniformness when having more photos in each grid.

Finally, we leverage the Campus dataset (see Section 6.1) to examine the effectiveness of our spatial diversity model. Four constraint conditions are considered with the offloading constraint ranging from 15 to 21 percent of the size of the raw collection. For each case, we first enumerate all the possible selections,[3] and calculate their prior spatial diversity using Eq. (4) and posterior photo coverage using Eq. (1). Then we sort each selection in the ascending order of its spatial diversity and calculate the mean photo coverage of selections under each diversity level. As shown in Fig. 4, a photo set with large spatial diversity (normalized to range [0,1]) presents high photo coverage. This validates our proposal of qualifying photo coverage with location distribution entropy.

4.2 The Content Influence Factor

The consideration of the spatial diversity factor can provide a series of selections with proper photo coverage. In order to guarantee certainty on the coverage, we adopt the content influence factor to measure the view quality of a photo set.

A crowdsensing photo's content captures a potential PoI from an aspect. A pair of useful photos of one PoI are likely to show high similarity on content due to the intersection of their covered aspects, while photos with low view quality

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

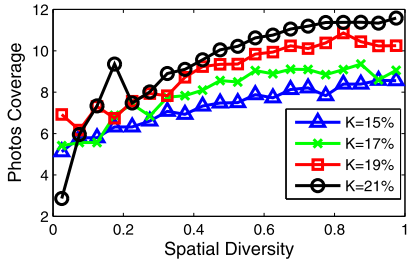


Fig. 4. The relations between the spatial diversity factor and photo coverage for photos from the Campus dataset. $K{=}x\%$ means that the maximum number of offloading photos is $x\% \cdot |\mathcal{D}|$.

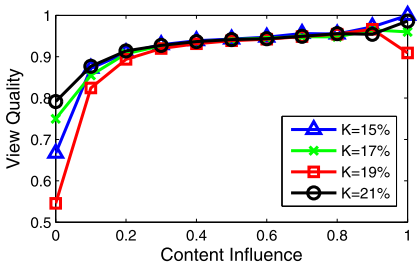


Fig. 5. The relations between the content influence factor and view quality for photos from the Campus dataset.

(i.e., blocked, blurred, or wrong direction) are supposed to have their own defects, thus distinct on their content. We denote the visual similarity between a pair of photos as their content influence on each other. Thus, a photo with high content influence to the other photos is expected to present good view quality. In respect of a photo selection, we regard photos in this selection as 'candidates' for offloading and the rest photos in the raw collection as 'voters', and then denote the comprehensive visual similarity level between the 'candidates' and the 'voters' as the content influence of this selection. In this way, a photo subset with high content influence is deemed to have good performance on view quality as it earns many votes from its complementary set (i.e., the unselected photos believe in the representativeness of the selection on the visual dimension).

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

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

where $sim(p_i, p_j)$ is the visual similarity function and reflects the visual 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. In this work, we adopt the SURF feature detection method because it shows good feature matching performance in most situations (e.g., illumination changes, affine transformation), and most importantly, it is fast in image processing [30]. The similarity between two photos can be calculated as,

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

which equals the number of matched features of the photos over all the features extracted from them. Wherein, F_i denotes the SURF features of photo p_i. An effective way to find a correct match for a feature in F_i is to compare the distance to its second-closest neighbor to that of its closest neighbor in F_j. If the ratio of the two distances is larger than a predefined threshold (usually 1.5), the closest neighbor is considered to be a match. Then we can calculate how many correct matches exist in F_j for features in F_i [31], denoted as $N(F_i, F_j)$. Apparently, having more matches indicates sharing a larger similarity. Since features from two different photos are not the same, the matching process between two photos is asymmetric (i.e., $N(F_i, F_j) \neq N(F_i, F_j)$). Hence, we adopt function $M(F_i, F_j){=}\frac{1}{2} \cdot (N(F_i, F_j){+}N(F_j, F_i))$ to facilitate a symmetric similarity measure (i.e., to yield $sim(p_i, p_j){=}sim(p_j, p_i)$).

The above influence model is for one photo. For a photo selection $\mathcal{I}$, its content influence factor is defined as,

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

where $F_c(\mathcal{I})$ measures the visual correlation between subset $\mathcal{I}$ and subset $\mathcal{D}{-}\mathcal{I}$. For the photos left behind in $\mathcal{D}{-}\mathcal{I}$, a strong correlation means that the selected photo set $\mathcal{I}$ can effectively represent them on content, which in turn indicates the high view quality of $\mathcal{I}$ as photos in $\mathcal{D}{-}\mathcal{I}$ give massive support to it. It is straightforward to see that $F_c(\mathcal{I}){=}F_c(\mathcal{D}{-}\mathcal{I})$. Also, given two subsets $\mathcal{I}_1, \mathcal{I}_2 \subset \mathcal{D}$, we can have the following properties:

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

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

We can use photo subset $\{p_1, p_2\}$ in the content layer of Fig. 3 as an example of a selection. The sum of visual similarities between $\{p_1, p_2\}$ and its complementary set $\{p_3, p_4, p_5\}$ is 2.3, which denotes the content influence of $\{p_1, p_2\}$.

Finally, we evaluate the relations between the content influence factor and the view quality of a photo set. We first enumerate all the possible photo selections in the considered conditions and calculate their prior content influence using Eq. (7) as well as posterior view quality using Eq. (2). Then the photo selections are sorted in the ascending order of their content influence. We calculate the mean view quality for photo selections under each content influence level for the analysis of the relations. As shown in Fig. 5, a photo selection with high content influence (normalized) shows well view quality. Such positive correlation validates our content influence model as a metric for view quality.[4]

4.3 Photo Set Utility Measure

In order to qualify the merit of a photo subset in terms of photo coverage and view quality, we jointly consider the spatial factor F_s and the content factor F_c in our utility

4. Note that even the impact of content influence on view quality is weaker after it passes 0.3, view quality still increases nearly linearly with it, making it an effective indicator of view quality.

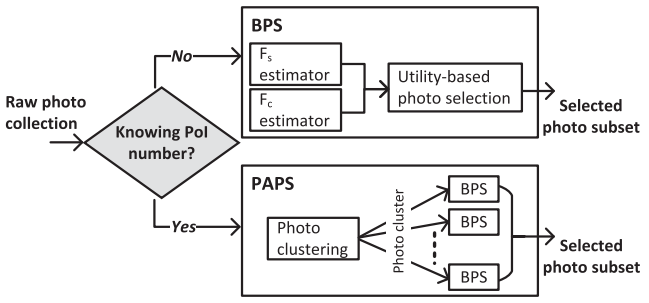


Fig. 6. The framework of our photo selection approach.

measure. Compared to a selection with fewer photos, more photos (within the offloading constraints) can have no doubt provide more information about the target. However, F_c, as a view quality metric, is not monotone, so merely adopting F_s and F_c for utility estimation may lead to a small selection. Hence, the set size $|\mathcal{I}|$ is also introduced in the utility to indicate the preference for a large selection.[5] Finally, we formulate the utility of a photo selection as,

$$U(\mathcal{I})=(1-\alpha)\cdot\|F_s(\mathcal{I})\|+\alpha\cdot\|F_c(\mathcal{I})\|+|\mathcal{I}|, \tag{10}$$

where factors F_s and F_c are normalized to the range of [0,1] and α is the weight characterizing and balancing the importance of photo coverage and view quality. We will evaluate the impact of weight α on the performance of photo selection in Section. 6.2. As an alternative for the certain coverage model in Eq. (3), $U(\mathcal{I})$ shares a similar form as $V(\mathcal{I})$ with the spatial and visual characteristics considered.

5 The Photo Selection Approach

The photo selection objective is to maximize the photo coverage and view quality of the offloading photos for the requester. Considering the difficulty in directly estimating photos' certain coverage, our photo selection approach alternatively looks for the best utility photo subset. As shown in Fig. 6, our approach consists of two schemes according to whether the PoI number information is available.

The basic photo selection scheme takes the crowdsensing photos as inputs and jointly exploits the spatial and content characteristics of a photo subset to facilitate a utility measure for the contribution of the subset. Then the subset with the maximum utility while satisfying the offloading constraint is selected and delivered to the requester. In some scenarios, the number of PoIs (i.e., n_{poi}) is provided by the requester or is known as common knowledge. For example, a requester is likely to specify the names of the PoIs in one target area where s/he has been to while don't have a visual record. These information will help to collect useful photos as the requests are further clarified. The PoI number-aware photo selection scheme uses the PoI number information to group photos into clusters first. By regarding each cluster as a summation of some PoIs, PAPS then performs BPS on the clusters to select photos for each PoI independently.

5. The cardinality term ensures the size of the selection to reach the upper limitation, while the spatial and content factors are for picking out the subset with well certain coverage from all the possible photo combinations with such a size.

5.1 BPS: Basic Photo Selection

BPS attempts to find the photo subset with the best utility from the scratch (i.e., only photos and their locations are provided). Based on the utility model and the constraint, the selection process of BPS can be formulated into the following problem.

Definition 4 (Utility-Based Photo Selection Problem). *Given a 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.*

Proof. We prove that the utility-based photo selection problem is NP-hard by reducing the well-known *maximum cut problem with given sizes of parts* (*MCGS* for short) [32] to a special case of the problem (i.e., α=1). This case finds a subset $\mathcal{I}$ ($|\mathcal{I}|\leq K$) that maximizes $\|F_c(\mathcal{I})\|+|\mathcal{I}|$. Since $\|F_c(\mathcal{I})\|\leq 1$, the special case can be further transformed into finding a subset $\mathcal{I}$ ($|\mathcal{I}|$=K) that maximizes $\|F_c(\mathcal{I})\|$.

An instance of the *MCGS* problem involves a graph G=$(\mathcal{V},\mathcal{E})$, in which each edge (u,v) has a nonnegative weight ω_{uv}. Given a positive integer $\mathrm{q}\leq|\mathcal{V}|/2$, the objective of such a problem is to find a cut $(\mathcal{S},\mathcal{V}-\mathcal{S})$ that can maximize $\sum_{u\in\mathcal{S},v\in\mathcal{V}-\mathcal{S},(u,v)\in\mathcal{E}}\omega_{uv}$ with $|\mathcal{S}|$=q.

Our reduction considers $\mathcal{V}$ as $\mathcal{D}$, where each vertex of G becomes a photo in $\mathcal{D}$ accordingly. We denote the visual similarity level between each pair of photos as the weight of each edge in $\mathcal{E}$ (0 if no edge exists between two vertex). Then we consider $\mathcal{S}$ as $\mathcal{I}$ and set K=q. In this way, finding the maximum cut $(\mathcal{S},\mathcal{V}-\mathcal{S})$ with $|\mathcal{S}|$=q is equivalent to selecting K photos from $\mathcal{D}$ so that the content influence of the selection $\mathcal{I}$ to the rest photos is maximized. Since this reduction process can be done in polynomial time, we prove that maximizing $\|F_c(\mathcal{I})\|+|\mathcal{I}|$ under the constraint is NP-hard.

Finally, if we assume that the original problem can be solved in polynomial time, then we can also find a polynomial solution for the special case by setting α=1. This contradicts the fact that the maximization problem of the special case is NP-hard. Therefore, our assumption is wrong, and we conclude that the utility-based photo selection problem is NP-hard. □

Designing algorithms to find the best offloading photo subset $\mathcal{I}$ is challenging since it belongs to NP-hard. To relieve this pitfall, the BPS scheme leverages the greedy strategy to obtain an approximate solution. The greedy-based algorithm *BasicSelection* is listed in Algorithm 1.

The basic idea of *BasicSelection* is to iteratively find the best photo which yields the maximum increase on the utility (ties are broken arbitrary) and add it into the selection (Lines 9-10). In each iteration, the spatial diversity gain $\Delta^i_{F_s}$ of each photo p_i and its content influence to the other photos are calculated in order to estimate its utility gain (Lines 5-8). The algorithm stops when the number constraint is active or all photos have been selected (Line 12).

We use the example in Fig. 3 to explain the algorithm. Here, we suppose that only two photos can be selected and delivered. Initially, the spatial gain for each photo is the

same, so *BasicSelection* looks for the most influential photo. The sum of similarities between photo p_1 and photos in $\mathcal{D}-p_1$ (the complementary set of p_1) is the largest among all combinations of $(p_i, \mathcal{D}-p_i)$ (i.e., 2.3), so p_1 is selected in the first round. Next, p_3 is picked out in the second round as $\Delta U(p_3)$ is the largest among the utilities of photos in subset $\mathcal{D}-p_1$. Thus, *BasicSelection* provides a solution $\mathcal{I}_1=\{p_1, p_3\}$, which has a spatial diversity of 0.5 and a content influence of 3. Note that this is not the optimal solution, in which subset $\mathcal{I}_2=\{p_2, p_4\}$ is selected. $\mathcal{I}_2$ yields the same spatial diversity as $\mathcal{I}_1$ while presenting a bigger content influence of 3.5 than that of $\mathcal{I}_1$. The approximation ratio of *BasicSelection* is analyzed as follows.

Algorithm 1. *BasicSelection*$(\mathcal{D},\ K,\ \alpha)$

```
1  begin
2    I ← ∅;
3    Compute sim(p_i, p_j) for each photo pair in D;
4    repeat
5        for each photo p_i in D−I do
6          Δ^i_{F_s} ← F_s(I ∪ {p_i}) − F_s(I);
7          INF_i^{D−I−p_i} ← Σ_{p_j∈D−I−p_i} sim(p_i, p_j);
8          ΔU(p_i) ← (1−α) · ‖Δ^i_{F_s}‖ + α · ‖INF_i^{D−I−p_i}‖;
9        end
10       p̂ ← argmax_{p_i}{ΔU(p_i)};
11       I ← I ∪ {p̂};
12   until |I| > K or D−I=∅
13   return I;
14 end
```

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

Proof. We first define the monotonicity and submodularity properties of 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 $F_s(\mathcal{I})$ and $F_c(\mathcal{I})+|\mathcal{I}|$ in the utility function separately. For simplicity, we ignore the weights and normalization function in Eq. (10) during the proof.

Remember that we denote $F_s(\mathcal{I})$ as the one dimension entropy of the distributed photo locations. Since the Shannon entropy is known to be monotone submodular [33], we claim that $F_s(\mathcal{I})$ is a monotone submodular function.

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

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

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

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

Then we substitute the term with the intersection and union of the two subsets and give the following proof,

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

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

Theorem 2. *The* BasicSelection *algorithm in BPS provides an approximation ratio of* $(1-1/e)$*, where* e *is the base of the natural logarithm.*

Proof. According to [34], the greedy algorithm can provides a $(1-1/e)$ approximation ratio for the optimization problem of a monotone submodular function. We have proved in *Lemma 1* that the utility function in Eq. (10) is monotone and submodular. □

Discussion on BPS. One may argue that we can simply filter those photos with low visual similarities to the others and then select photos merely based on spatial diversity. However, a photo showing weak visual correlations with the others may be the only one piece of description for a PoI or a description from a unique aspect. Hence, in order to obtain a proper photo selection in terms of certain coverage, the content factor must be considered jointly with the spatial factor.

5.2 PAPS: PoI Number-Aware Photo Selection

We refer to BPS when only limited information is available for selection. In cases where the number of PoIs (i.e., n_{poi}) is known, PAPS is believed to generate photo selection with better certain coverage than BPS by grouping photos into clusters before performing BPS on each cluster.

By introducing a photo clustering stage, PAPS tries to group photos capturing the same PoI into one cluster. We

point out that mean-based clustering techniques (e.g., k-means), which are widely used when the number of clusters is given, are not adequate for photo clustering as calculating the mean of several visual features is meaningless. Also, this category of algorithms is extremely sensitive to the initial configuration [25]. Fortunately, the similarity between two photos can be easily measured from the spatial dimension and the content (or visual) dimension. As stated in [35], if we cannot attain more information than similarities between data points, a proper way of representing the data is in form of the similarity graph. By representing photos in a similarity graph, the photo clustering process is then reformulated into a graph partition problem, which can be well handled by spectral clustering.

In order to construct a graph for spectral clustering, we first formally define the graph similarity measure used in PAPS.[6] Generally, photos of the same PoI usually present smaller geographical distances and higher matching levels on visual than photos of different PoIs. Hence, we novelly combine spatial similarity and visual similarity together and formulate the graph similarity between photos p_i and p_j ($i \neq j$) as,

$$\omega_{ij}=\beta \cdot \|sim(p_i,p_j)\|+(1-\beta)\cdot \|e^{(-\frac{dis(p_i,p_j)}{d_{lim}})}\|, \tag{14}$$

where $sim(p_i,p_j)$ is the visual similarity in Eq. (6) and the second term denotes spatial similarity. Parameter β is the weight that balances these two similarity factors. In the second term, $dis(p_i,p_j)$ is the euclidean distance between photo p_i and p_j and parameter $d_{lim}=\frac{1}{|D|^2-|D|}\cdot\sum_{i\neq j} dis(p_i,p_j)$ is introduced in order to narrow down the range of spatial similarity. Note that if we merely leverage the geographical relations for graph similarity (i.e., β=0), we would fail to distinguish neighboring photos of different PoIs. On the other hand, if we only consider the visual correlations (i.e., β=1), photos for different aspects of one PoI may be grouped into different clusters.

Based on the graph similarity model, the photo selection process of PAPS is designed as Algorithm 2. As shown, the graph similarity between each pair of photos is first calculated (Line 2). Then we construct a graph in which each node represents a photo and each edge is tagged with a weight indicating the degree of graph similarity between these two photos (Line 3). Given this fully connected graph, normalized spectral clustering in [36] can be used to partition the nodes into clusters based on the n_{poi} smallest eigenvectors of the graph's Laplacian matrix (Line 4). This graph similarity-based clustering stage is named gs-sym (postfix 'sym' means that a symmetric Laplacian matrix is used). Next, we select photos from each cluster proportional to the size of the cluster based on BPS (Lines 5-8). Finally, the selected photos from different clusters are combined into one set and returned to the requester (Line 9).

Discussion on PAPS. The locations of the PoIs in a target area are not supposed to be known in prior as it is hard and not flexible for a requester to tag the locations. In fact, the PoIs in photo crowdsensing scenarios can be objects with irregular shapes and relatively huge sizes, making it infeasible to describe their accurate locations.

Algorithm 2. $AdvancedSelection(\mathcal{D},\ K,\ \alpha,\ \beta,\ n_{poi})$

1 **begin**
2 *Compute ω_{ij} between photo p_i and p_j ($i \neq j$) using Eq. (14);*
3 *Construct similarity graph G based on $\mathcal{D}$ and $\{\omega_{ij}\}$;*
4 $\{Cls_i\} \leftarrow SpectralClustering(G, n_{poi})$;
5 **for** *each cluster Cls_i* **do**
6 $K_i \leftarrow \lfloor K \cdot (|Cls_i|/|\mathcal{D}|) + 0.5 \rfloor$;
7 $\mathcal{I}_i \leftarrow BasicSelection(Cls_i,\ K_i,\ \alpha)$;
8 **end**
9 $\mathcal{I} \leftarrow \bigcup_{i=1}^{n_{poi}} \mathcal{I}_i$;
10 **return** $\mathcal{I}$;
11 **end**

Two different similarity measures are designed in Eqs. (6) and (14). This is because content influence models peer-to-community similarity which calculates the sum of one photo's similarities to the others. Visual similarities between one photo and the other photos are additive, while adding up the spatial similarities is meaningless (we will not select a photo just because it is adjacent to the other photos). Therefore, we use merely visual similarity when formulating content influence. On the other hand, similarity graph considers peer-to-peer similarity, so both visual and spatial similarity can be helpful in constructing and processing this graph.

6 Experiments

This section evaluates the performance of our approach to crowdsensing photo selection. First, we examine the effectiveness of our photo set utility model by varying the weight α. Second, we compare the performance of our BPS scheme with that of the clustering-based and the random photo selection scheme. Finally, we evaluate the photo clustering performance and the photo selection performance of our PAPS scheme.

6.1 Experimental Setup

We conduct extensive experiments on three real-world photo datasets collected by ourselves.[7] During the photography, we have two independent researchers take photos around three target 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 considered as the participants-to-server data collection phase of mobile crowdsensing tasks.

We present the statistics of these three photo datasets in Table 2 and show the distribution of PoI and photos in Fig. 7. The Campus dataset consists of photos collected from the campus of PolyU. As shown in Fig. 7a, totally 57

6. We use graph similarity in order to distinguish it with the visual similarity in Eq. (6). We will explain why we introduce two similarity measures in the discussion.

7. 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 2
Statistics of the Photo Datasets

Dataset	# PoIs	# aspects	# photos	# inaccurate photos
Campus	3	21	57	7
TPark	4	23	40	3
RPark	4	16	42	4

photos about 3 PoIs are taken in this target area and 7 of them are of low quality. The TPark dataset collects a group of 40 photos in a technology park at NUDT, where 4 PoIs are captured and 3 photos are with low quality as illustrated in Fig. 15b. We collect the RPark dataset in a park near the Xiangjiang River. As we can see from Fig. 7c, 42 photos of 4 PoIs are included and 4 of them are inaccurate. In each dataset, PoIs are uniformly distributed and captured from several aspects with photos at different locations. Totally 21, 23, and 16 PoI aspects are captured in datasets Campus, TPark, and RPark, respectively. During photo selection, all the collected photos were resized to 240×180 (preserving the aspect ratios of the raw photos) in order to accelerate the process of visual feature extraction.

We use photo coverage in Eq. (1), view quality in Eq. (2), and certain coverage in Eq. (3) as the performance metrics during our evaluation. Notice that merely having advantages in either photo coverage or view quality is not good enough for the requester. A proper photo selection should yield good overall performance (i.e., certain coverage). We asked students who are familiar with the sensing areas to exploit the ground truth knowledge of the photos.[8] Specifically, they are first asked to determine the PoI and the aspect that each photo captures to attain $\mathcal{S}^i_{poi}$ and $\mathcal{S}^{ij}_{asp}$ in Eq. (1). Then they label the accuracy of each photo by verifying whether its view is blocked, blurred, or wrongly recorded, based on which we can calculate $n^{\mathcal{I}}_{in}$ (Eq. (2)) for each selection $\mathcal{I}$.

We implement our photo selection approach and the baselines using Matlab. Our experiments are conducted on a workstation with 4 GB memory and a 2.94 GHz Intel(R) CPU.

6.2 Evaluation on the Utility Model

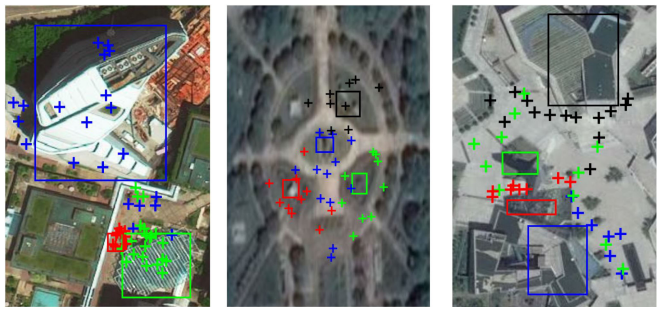
We first evaluate the effectiveness of our photo set utility model with varied weight α on the three datasets. α is used to reconcile spatial diversity (i.e., α=0) and content influence (i.e., α=1) in order to facilitate the utility estimation. Specifically, we vary α from 0 to 1 and test the overall performance[9] of the selected photo set that the *BasicSelection* algorithm generates. The test is performed under several different budgets (i.e., the percentages of photos allowed for offloading).

As shown in Fig. 8, for every dataset, depending only on either spatial diversity (α=0) or content influence (α=1) usually cannot provide satisfactory result. We notice that there are some special cases where α=1 can also yield the best performance (e.g., K=12% in Campus). This happens when the representative photos selected based on content influence cover the PoIs and aspects evenly, namely, the locations of the K-most influential photos happen to be one of the combinations that can yield the biggest distribution entropy. We suppose that the selection process for datasets from relatively small target area (e.g., the Campus dataset) tends to have such results as photos' locations would be easily distributed into different grids, but this is definitely not the general case. On the other hand, since the location distribution of photos is not relevant to their quality, the view quality of the selection cannot be guaranteed when only using spatial diversity for the utility. Hence, α=0 provides the worst overall performance in almost all the tested situations.

As we can tell from Fig. 8, jointly considering both factors ($0 < \alpha < 1$) is generally the best choice to attain a photo selection with well certain coverage. We also observe that α=0.4 provides the best overall performance under different budgets in three datasets. Thus, in the rest of the experiments, we set 0.4 as the default value of α. One can adjust α according to different application requirements (e.g., choosing a bigger value helps to yield better view quality).

8. These students are not part of the research team, making them independent consultants.

9. We have already examined the effectiveness of spatial diversity on attaining photo coverage in Fig. 4 and content influence on improving view quality in Fig. 5. Thus, we only test the level of certain coverage (i.e., overall performance) that can be achieved with varied α.



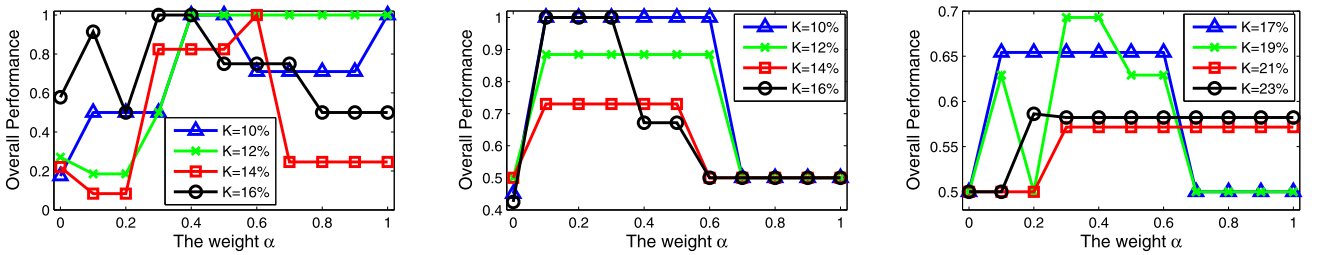
(a) Dataset Campus. (b) Dataset TPark. (c) Dataset RPark.

Fig. 7. The distribution of PoIs and collected photos on the satellite images of the target areas in three datasets. Wherein, a rectangle encloses a PoI and symbol '+' tags the location of a photo (A photo and the PoI it captures are painted with the same color.) Inaccurate photos are not presented as they capture no PoI.

6.3 Performance Evaluation for BPS

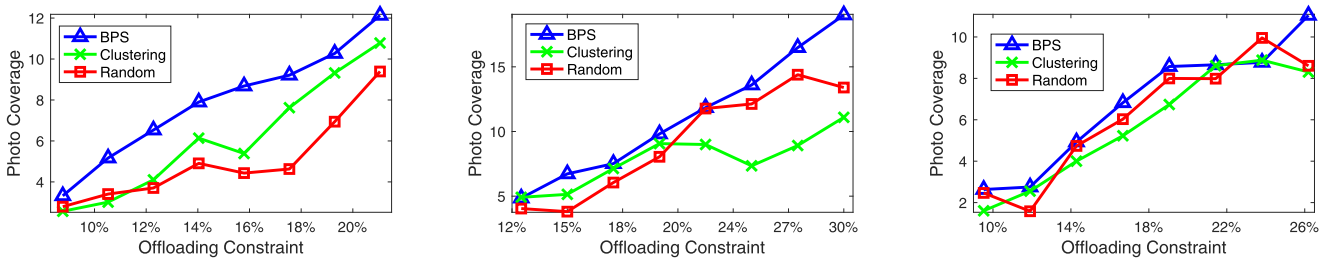
Next, we evaluate the performance of our BPS scheme by comparing it with the following two typical photo selection schemes.

- *Clustering-based selection*. In this scheme, the raw photo collection is clustered into groups based on the pyramid tree (i.e., PTree-based clustering) [12]. 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 (Eq. (6)) and spatial distance. The photo is added to the first cluster that presents a visual similarity larger than a visual threshold and a euclidean distance smaller than a spatial threshold. Otherwise, a new cluster is formed with that photo as the centroid. Instead of setting up a specific threshold, we use the mean value of the similarities and the inter-distances between all pairs of photos in the raw collection as the visual and spatial threshold, respectively. Based on the clustering results, we select one photo from each cluster (clusters with larger size first) until the offloading constraint is active.



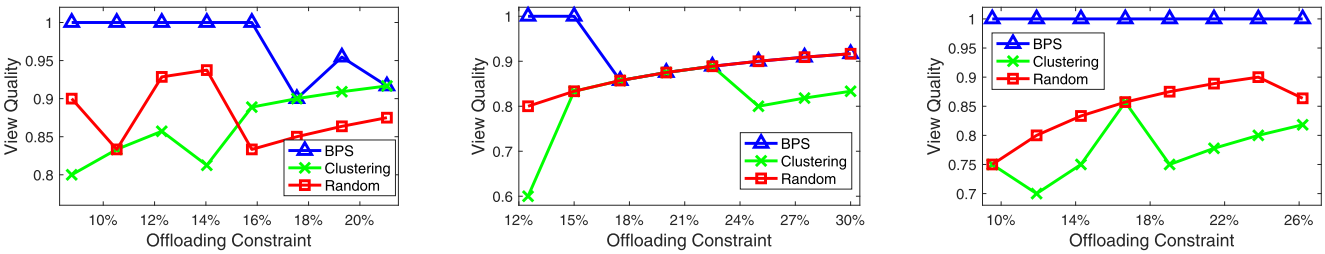
(a) Impact analysis for the Campus dataset. (b) Impact analysis for the TPARK dataset. (c) Impact analysis for the RPARK dataset.

Fig. 8. Impact of the weight α on the overall performance of the photo selection.



(a) Performance on the Campus dataset. (b) Performance on the TPark dataset. (c) Performance on the RPark dataset.

Fig. 9. Evaluation of the photo coverage performance on three datasets.



(a) Performance on the Campus dataset. (b) Performance on the TPark dataset. (c) Performance on the RPark dataset.

Fig. 10. Evaluation of the view quality performance on three datasets.

- *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 8 constraint levels and examine the photo coverage and view quality performance of BPS, clustering-based selection, and random selection on them.

First, from Fig. 9, we can see that an informative photo subset can be selected by BPS. As expected, the performance on posterior photo coverage of BPS's photo selection keeps increasing linearly with the budget as more PoIs and more aspects are included. We also observe an increasing photo coverage for the other two schemes, but their performance is generally worse than that of BPS. Specifically, in terms of photo coverage, BPS outperforms clustering-based selection by an average of 37, 39, and 23 percent and outperforms random selection by an average of 60, 26, and 16 percent for the three datasets, respectively. Compared with the results in Fig. 14a and 14b, the performance advantages of BPS in RPark are less obvious (Fig. 14c). This is because that RPark has less PoI aspects than the other two datasets, making it easily for the two baselines to cover more aspects with the same budget.

As shown in Fig. 10, BPS can formally avoid the ingredients with low view quality during the selection process in the three scenarios, while the other two schemes have no guarantee on this aspect. The average performance improvement of our scheme on view quality are 12, 15, and 29 percent for clustering-based selection, and 10, 6, and 18 percent for random selection. We also observe that the view quality of BPS's photo selection will degrade slightly when the budget exceeds a certain level in Campus and TPark. This is because that some inaccurate photos will be unfortunately mixed in during selection as the proportion of useless ingredients in the unselected photos increases. Though, BPS can still yield a better view quality than the two baselines.

We give an example to illustrate the selected photos of the three schemes in Fig. 11. In this example, BPS provides a selection which covers 3 PoIs, while the other two schemes have only 2 PoIs covered and the selections contain photos with low quality. Overall, BPS can extract more valuable photos from the collection than the other two schemes under different budget constraints. We owe the performance advantages of BPS to the jointly consideration of spatial and content factor during selection. In contrast, clustering-based selection 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 strategy is not steady, but its view quality performance is usually better than the clustering scheme. The reason is that the clustering scheme may accidently

Fig. 11. Examples of photos selected by BPS, clustering-based selection, and random selection on the Campus dataset with offloading constraint K=12%. The first two images in the second line and the 5th image in the third line are with blocked views. These are the results of one round of selection. We calculate the average performance of the selected photos from ten rounds of selection during our evaluation.

classify inaccurate photos into one single cluster, which gives the inaccurate ones a relatively high probability of being selected.

6.4 Performance Evaluation for PAPS

In this section, we evaluate our PAPS scheme on crowdsensing photo selection by investigating its photo clustering (i.e., gs-sym) performance, and the selection's performance on photo coverage and view quality.

We first evaluate the effectiveness of the graph similarity model (i.e., Eq. (14)) for photo clustering with the weight β varying from 0 to 1. β is introduced in the graph similarity in order to facilitate an integrated measure based on visual and spatial similarity. As shown in Fig. 12, merely adopting visual similarity (i.e., β=1) for clustering, as proposed in [37], yields the worst accuracy in all the three datasets. Using spatial similarity (i.e., β=0) for clustering presents better accuracy than using visual similarity as crowdsensing photos tend to be taken around their corresponding PoIs that uniformly distributed in the area. However, as we can see from Fig. 7, the geographic boundary for the locations of each PoI's photos is non-convex, namely, photos of different PoIs may geographically mix with each other in some regions. By introducing visual similarity, we can distinguish photos of different PoIs that have small inter-distance. This explains the clustering accuracy advantages of graph similarity ($0 < \beta < 1$) in terms of simply based on visual or spatial similarity during clustering in Fig. 12. Particularly, we point out that the performance differences between using graph similarity and the other two metrics are more significantly in TPark and RPark than in Campus. This is because more photos are geographically mixed in TPark and RPark. We set β=0.5 in the rest experiments as graph similarity with β=0.5 provides a high clustering accuracy in the tests.

In order to evaluate the performance of gs-sym, we compare it with the following two clustering methods. 1) k-means performs traditional k-means clustering on the locations of photos (considering a photo's GPS location as its two-dimensional feature). 2) gs-rm (postfix 'rm' represents random walk) also leverages our graph similarity model, but uses the spectral clustering technique proposed in [38]. gs-sym is not compared with the PTree-based clustering because the latter is not a fixed-width method (generating 21, 16, and 18 clusters during the tests on Campus, TPark, and RPark) and its clustering accuracy cannot be estimated. As shown in Fig. 13, gs-sym presents the best accuracy among the three methods for all three datasets. Particularly, it outperforms the k-means baseline by 20 percent on clustering accuracy in Campus. The facts that gs-sym outperforms gs-rm in all the scenarios evaluates the use of a symmetric Laplacian matrix in spectral clustering.

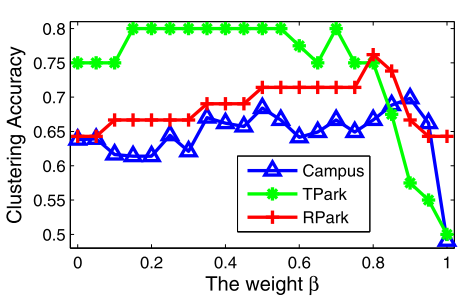


Fig. 12. The impact of weight β on the clustering accuracy of gs-sym.

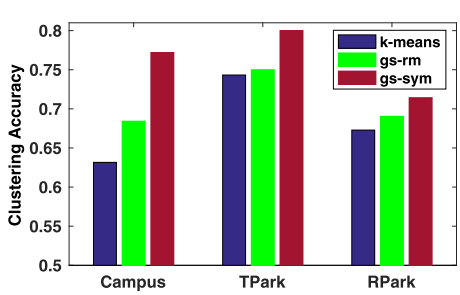
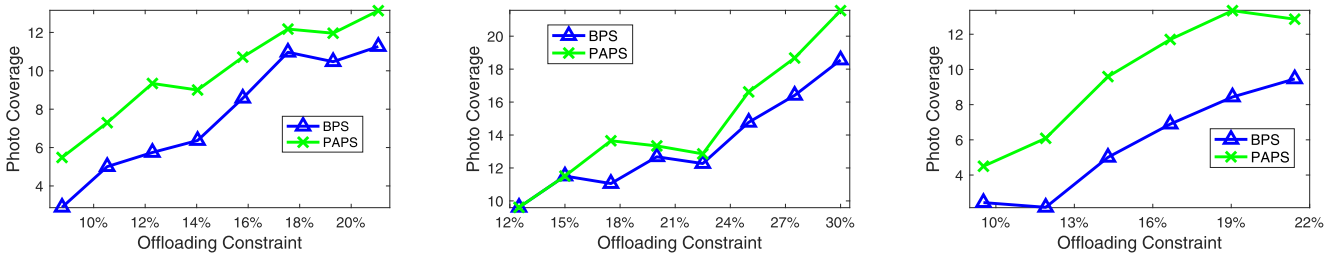


Fig. 13. The accuracy performance of k-means, gs-rm, and gs-sym.



(a) Performance on the Campus dataset. (b) Performance on the TPark dataset. (c) Performance on the RPark dataset.

Fig. 14. Comparison on the photo coverage performance of PPAS and BPS.

We evaluate the photo selection performance of PAPS by comparing it with BPS. For the performance on photo coverage, evaluation results is presented in Fig. 14. As expected, PAPS can provide better photo coverage than BPS. The advantage owes to the photo clustering stage that transforms the selection process into per-PoI granularity, in which way more PoI aspects can be selected than simply selecting from the scratch (i.e., BPS). Specifically, PAPS outperforms BPS on photo coverage by an average of 38, 9.5, and 87 percent for Campus, TPark, and RPark, respectively. On the other hand, since PAPS and BPS both adopts the content influence model for view quality assurance, the performance of these two schemes on view quality should be of no obvious difference. As illustrated in Fig. 15, the photo selections of PAPS and BPS yield almost the same view quality (the largest improvement of PAPS from BPS is 6.8 percent in Campus).

Finally, we analyze the execution cost of our approach. The execution process can be roughly divided into a photo preprocessing stage (i.e., SURF feature extraction and visual similarities calculation) and a photo selection stage. In our tests, the preprocessing times for the collected photos are $59s$, $27s$, and $25s$ for Campus, TPark, and RPark, respectively. Meanwhile, the time cost of photo selection is quite small, only taking an average of $0.005s$ and $0.04s$ for BPS and PAPS based on the three datasets. Note that PAPS requires more execution time than BPS due to the additional clustering stage. We emphasize that photo preprocessing can be performed during photo collection to reduce this part of time cost and it only has to be performed once for each collected photo.

7 Limitations and Implications

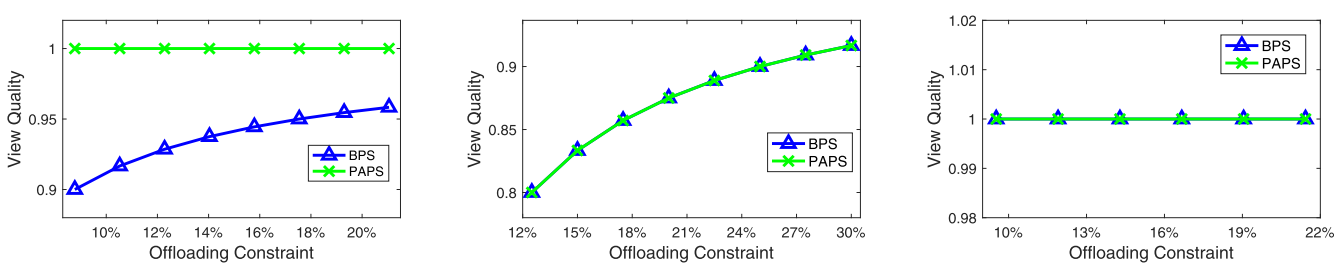
Limitations. The proposed schemes in this paper are for server-to-requester photo selection, so it cannot save the uploading cost of the participants. One can reduce the uploading cost by requiring the participants to extract the SURF features locally and upload the features to the MCS server for selection. Then only those photos corresponding to the selected features need to be uploaded and transmitted to the requester. However, such an interaction process introduces additional latency and requires more participants involvement, which is against our intention to encourage user participation.

We also state that the effectiveness of the proposed schemes could be destroyed by replication threats, in which the same photo is reported several times. Such behaviors will introduce useless computation and will mislead the schemes to select the replicated photos even they are of low quality. One possible solution for such threats is to set up an empirical threshold to filter out photos that present abnormal similarity to each other.

Implications. We find it hard to estimate the quality of a piece of data without any references or criteria, especially when the data is with uncertain properties. However, our efforts in this work indicate that it is possible to measure the merit of a set of uncertain data by exploiting the statistical characteristics of the set and the relations among the data points. We consider this as an application of the crowd wisdom.

8 Conclusion

In the photo MCS applications, plenty of crowdsensing photos aggregate to the server, but only a representative photo subset are expected to be delivered to the requester for an online view. In this paper, we first study the server-to-requester photo selection problem and analyze the certain coverage challenges in terms of crowdsensing photo uncertainty. We design a utility measure, which integrates a spatial factor and a content factor, to quantify the contribution of a photo set on certain coverage. Then we propose the BPS scheme that leverages a greedy-based strategy to select the photo set with the approximately maximum



(a) Performance on the Campus dataset. (b) Performance on the TPark dataset. (c) Performance on the RPark dataset.

Fig. 15. Comparison on the view quality performance of PPAS and BPS.

utility. An advanced selection scheme (i.e., PAPS) is also proposed by introducing a novel photo clustering stage before performing BPS under PoI number-aware situations. With the help of BPS and PAPS, we can formally bridge the gap between the uncertain crowdsensing photos and the certain coverage expectation of a target area. We conduct extensive experiments on three real-world photo datasets. Experimental results demonstrate the effectiveness of our proposal and show the performance advantages of both BPS and PAPS over the clustering-based scheme and the random selection scheme. In the future, we are interested in studying the detection of anomalies (e.g., riots) with a selection of passive crowdsensing photos collected through opportunistic sensing.

Acknowledgments

The work is supported by the National Key Research and Development Program of China under Grant Nos. 2018YFB1800202, 2016YFB1000302 and the National Natural Science Foundation of China under Grant Nos. 61872372, 61772446, 61672195.

References

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

[2] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'Hamed, "Sparse mobile crowdsensing: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 161–167, Jul. 2016.

[3] 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 *Proc. ACM Conf. Comput. Supported Cooperative Work Social Comput.*, 2017, pp. 1139–1151.

[4] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, 2014, Art. no. 55.

[5] 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 Comput. Surveys*, vol. 48, no. 1, 2015, Art. no. 7.

[6] 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 *Proc. ACM Int. Conf. Multimedia*, 2007, pp. 631–640.

[7] S. Kisilevich, D. Keim, N. Andrienko, and G. Andrienko, *Towards Acquisition of Semantics of Places and Events by Multi-perspective Analysis of Geotagged Photo Collections*. Berlin, Germany: Springer, 2013.

[8] K. Tuite, N. Snavely, D.-y. Hsiao, N. Tabing, and Z. Popovic, "Photocity: Training experts at large-scale image acquisition through a competitive game," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 1383–1392.

[9] D. Quercia, N. K. O'Hare, and H. Cramer, "Aesthetic capital: What makes london look beautiful, quiet, and happy?" in *Proc. ACM Conf. Comput. Supported Cooperative Work Social Comput.*, 2014, pp. 945–955.

[10] 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 Trans. Veh. Technol.*, vol. 66, no. 4, pp. 2902–2914, Apr. 2016.

[11] Y. Wu, Y. Wang, and G. Cao, "Photo crowdsourcing for area coverage in resource constrained environments," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2017, pp. 1–9.

[12] H. Chen, B. Guo, Z. Yu, L. Chen, and X. Ma, "A generic framework for constraint-driven data selection in mobile crowd photographing," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 284–296, Feb. 2017.

[13] 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 *Proc. IEEE Real-Time Syst. Symp.*, 2011, pp. 317–326.

[14] U. Weinsberg, Q. Li, N. Taft, A. Balachandran, V. Sekar, G. Iannaccone, and S. Seshan, "Care: Content aware redundancy elimination for challenged networks," in *Proc. ACM Workshop Hot Top. Netw.*, 2012, pp. 127–132.

[15] Y. Hua, W. He, X. Liu, and D. Feng, "Smarteye: real-time and efficient cloud image sharing for disaster environments," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2015, pp. 1616–1624.

[16] 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 Trans. Mobile Comput.*, vol. 14, no. 10, pp. 2020–2033, Oct. 2015.

[17] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng, "Truth discovery on crowd sensing of correlated entities," in *Proc. ACM Int. Conf. Embedded Netw. Sensor Syst.*, 2015, pp. 169–182.

[18] 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 *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2016, pp. 1–12.

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

[20] T. Yan, V. Kumar, and D. Ganesan, "Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones," in *Proc. ACM Int. Conf. Mobile Syst. Appl. Serv.*, 2010, pp. 77–90.

[21] T. Zhou, B. Xiao, Z. Cai, M. Xu, and X. Liu, "From uncertain photos to certain coverage: a novel photo selection approach to mobile crowdsensing," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2018, pp. 1979–1987.

[22] Y. Wu, Y. Wang, W. Hu, and G. Cao, "Smartphoto: A resource-aware crowdsourcing approach for image sensing with smartphones," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1249–1263, May 2016.

[23] Y. Wu, Y. Wang, W. Hu, X. Zhang, and G. Cao, "Resource-aware photo crowdsourcing through disruption tolerant networks," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2016, pp. 374–383.

[24] M. Y. S. Uddin, M. T. A. Amin, T. Abdelzaher, A. Iyengar, and R. Govindan, "PhotoNet+: Outlier-resilient coverage maximization in visual sensing applications," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2012, pp. 143–144.

[25] I. Simon, N. Snavely, and S. M. Seitz, "Scene summarization for online image collections," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[26] Y. Jiang, X. Xu, P. Terlecky, T. Abdelzaher, A. Bar-Noy, and R. Govindan, "Mediascope: Selective on-demand media retrieval from mobile devices," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2013, pp. 289–300.

[27] H. Chen, B. Guo, Z. Yu, and Q. Han, "Toward real-time and cooperative mobile visual sensing and sharing," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.

[28] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *Proc. ACM Int. Conf. Ubiquitous Comput.*, 2010, pp. 33–36.

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

[30] J. Luo and G. Oubong, "A comparison of sift, pca-sift and surf," *Int. J. Image Process.*, vol. 4, no. 3, pp. 143–152, 2009.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[32] A. A. Ageev and M. I. Sviridenko, "Approximation algorithms for maximum coverage and max cut with given sizes of parts," in *Proc. Int. Conf. Integer Program. Combinatorial Optim.*, 1999, pp. 17–30.

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

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

[35] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[36] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," *Proc. 14th Int. Conf. Neural Inf. Process. Syst.*, 2002, vol. 14, pp. 849–856.

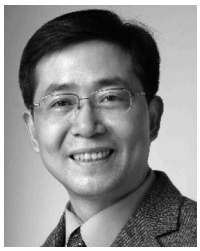
[37] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the world's photos," in *Proc. Int. Conf. World Wide Web*, 2009, pp. 761–770.

[38] J. Shi and J. Malik, *Normalized Cuts and Image Segmentation*. Los Alamitos, CA, USA: IEEE Comput. Soc. Press, 2000.

Tongqing Zhou received the BS, MSc, and PhD degrees in computer science and technology from the National University of Defense Technology (NUDT), Changsha, in 2012, 2014, and 2018, respectively. He was a visiting PhD student with the Hong Kong Polytechnic University under the supervision of Dr. Bin Xiao in 2017. His main research interests include mobile sensing, edge computing, and data security.

Bin Xiao received the BSc and MSc degrees in electronics engineering from Fudan University, China, in 1997 and 2000, respectively, and the PhD degree in computer science from the University of Texas at Dallas, in 2003. After the PhD graduation, he joined Hong Kong Polytechnic University as an assistant professor. Currently, he is an associate professor with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. His research interests include mobile cloud computing, data management, network security, and RFID systems. He is an associate editor for the *International Journal of Parallel, Emergent and Distributed Systems*. He is a recipient of the Best Paper Award from IEEE/IFIP EUC 2011. He is a senior member of the IEEE and a member of the ACM.

Zhiping Cai received the BS, MSc, and PhD degrees in computer science and technology with honor from the National University of Defense Technology (NUDT), in 1996, 2002, and 2005, respectively. Currently, he is a professor with the Department of Network Engineering, College of Computer, NUDT at Changsha, China. His current research interests include big data, network security, and network virtualization. His doctoral dissertation has been rewarded with the Outstanding Dissertation Award of the Chinese PLA. He is a member of the ACM and IEEE.



Ming Xu received the BSc and MSc degrees from Wuhan University, China, in 1984 and 1987, respectively, and the PhD degree from the National University of Defense Technology (NUDT), China, in 1995. He is a professor with the Department of Network Engineering, College of Computer, National University of Defense Technology, China. His research interests include ad-hoc networks, vehicular networks, and wireless mesh networks. He has published more than 130 academic papers in journals and conferences, co-authored of three books. He is an editor of the *Journal of Communications* and the *International Journal of Pervasive Computing*, and has been program committee members for more than 30 international conferences or workshops. He is a member of the IEEE.