Exploiting Social-Mobile Information for Location Visualization

JITAO SANG, QUAN FANG, and CHANGSHENG XU, Chinese Academy of Sciences

With a smart phone at hand, it becomes easy now to snap pictures and publish them online with few lines of texts. The GPS coordinates and User-Generated Content (UGC) data embedded in the shared photos provide opportunities to exploit important knowledge to tackle interesting tasks like geographically organizing photos and location visualization. In this work, we propose to organize photos both geographically and semantically, and investigate the problem of location visualization from multiple semantic themes. The novel visualization scheme provides a rich display landscape for geographical exploration from versatile views. A two-level solution is presented, where we first identify the highly photographed places of interest (POI) and discover their focused themes, and then aggregate the lower-level POI themes to generate the higher-level city themes for location visualization. We have conducted experiments on crawled Flickr and Instagram data and exhibited the visualization for the cities of Singapore and Sydney. The experimental results have validated the proposed method and demonstrated the potentials of location visualization from multiple themes.

Categories and Subject Descriptors: H.4 [**Information Systems Applications**]: Miscellaneous

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Geotagged photo, point of interest, location visualization

ACM Reference Format:
Jitao Sang, Quan Fang, and Changsheng Xu. 2017. Exploiting social-mobile information for location visualization. ACM Trans. Intell. Syst. Technol. 8, 3, Article 39 (January 2017), 19 pages.
DOI: http://dx.doi.org/10.1145/3001594

1. INTRODUCTION

By watching movies, TVs, news, or hearing from others who have been to Singapore, we may know that Singapore is a popular tourist city with its great landscape, intersecting culture, delicious food, and excellent entertainment. The landscape, entertainment, food, and culture are called *themes*. For new visitors to Singapore and for those who want to explore the city thoroughly, they will be also interested to explore the *Place Of Interest* (*POI*) that is featured in each theme, for example, Singapore has wonderful landscape and various entertainment choices on Sentosa Island, the best food in Singapore can be found in China Town, etc. This work addresses the problem of extracting the representative POIs and themes for a location from user-generated social-mobile [Del Bimbo et al. 2014] information for location visualization.

This work is supported by the National Natural Science Foundation of China (No. 61432019, 61225009, 61303176, 61672518, 61272256, 61332016).
Authors' addresses: J. Sang, Q. Fang, C. Xu (corresponding author), National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China; emails: {jtsang, qfang, csxu}@nlpr.ia.ac.cn.

DOI: http://dx.doi.org/10.1145/3001594

Recent years have witnessed the prosperity of various social media websites in the mobile platforms. According to the Nielsen report,[1] about 59% of web users in Asia interact with social media services directly from a mobile device. Along with the trend to share user-generated content (UGC) via mobile devices, most photo sharing websites have released their mobile apps or focused on the mobile platforms. For example, the mobile photo sharing service, Instagram, has approached 80 million registered users with 4 billion photos in less than two years.[2] Equipped with the mobile photo sharing apps, it is now easy for users to share photos anywhere and anytime, which has generated two types of social-mobile information: (1) the GPS information associated with the mobile devices records important geographical context of the photo being captured, which confidently reflects user's daily activities and mobility patterns; (2) the UGC metadata of title, tag, and comment provides valuable descriptions for the shared photos and the locations where the photos are taken, from which we can explore rich semantics. Recent work on utilizing the geotagged photos from photo sharing websites have either been devoted to exploiting the GPS information for location estimation [Liu et al. 2012b; Fang et al. 2013a] and landmark recognition [Zheng et al. 2009; Ji et al. 2011, 2012], or focusing on the UGC metadata for semantic knowledge mining like POI annotation [Rattenbury and Naaman 2009] and travel route summarization [Hao et al. 2010]. In this work, we propose to exploit the geotagged photos both semantically and geographically, and address a novel problem of *Location Visualization from Multiple Themes* (LVMT).

In the context of the LVMT problem, now we formally introduce the two important concepts of *POI* and *theme*. POI is a highly photographed place which is marked by its centroid. Since most of the geotagged photos are captured and shared through mobile devices, the highly photographed place tends to attract abundant user activities and is thus a potentially interesting one. POI is the basic unit in LVMT, around which photos are geographically organized and the low-level semantics are extracted. Theme indicates certain interesting topics or representative patterns within a location. For example, Singapore is expected to be visualized from its finance, shopping, natural scene, landmark, culture, and so on. In this work, themes are first extracted at the POI level, where interesting topics within a POI are first identified. Themes at the higher city level are then aggregated from the discovered POI themes and facilitate the task of city visualization. The proposed visualization solution is expected to have two levels: (1) POI visualization - the identified POIs inside a city, the POI themes with the exemplary photos; (2) Location visualization - the summarized city themes, the representative POIs, and exemplary photos for each city theme. Figure 1 shows the visualization scheme for Singapore.

The challenges are four-fold: (1) UGC-based POI identification is not trivial. Users tend to take photos from arbitrary angles and views, which leads to large visual variance among photos of the same POI. (2) POI themes are not readily available. In addition to the large visual variance, the associated textual metadata are rather noisy for understanding the underlying semantics within each POI. (3) Lower POI-level themes need to be aggregated to generate city-level themes. There exist both overlap and divergence between different POI themes. (4) The discovered themes are originally described by a group of tags. For ease of visualization, more compact theme labels are desired. To address these issues, we propose the LVMT solution with three stages: (1) *POI Identification*, to identify the POIs from geotagged photos and estimate the belonging POIs for the non-geotagged photos; (2) *POI Theme Discovery*, to discover

[1]http://thenextweb.com/asia/2012/11/16/report-half-of-worlds-social-media-users-go-mobile-as-us-and-europe-lag-asia/.
[2]http://blog.instagram.com/post/28067043504/the-instagram-community-hits-80-million-users.

Fig. 1. Problem illustration.

the underlying semantic themes and representative photos for each POI by a novel incremental learning scheme; (3) *City Theme Aggregation and Labeling*, to generate city themes by clustering similar POI themes and assigning theme labels by matching with the corresponding Wikipedia pages. With the discovered POI and city themes, we can easily visualize both POI and the city from multiple themes and in a coarse-to-fine fashion. We have conducted experiments on a crawled Flickr and Instagram dataset of the cities of Singpaore and Sydney. Objective, as well as subjective evaluations have validated the effectiveness of the proposed solution, and demonstrated the great potentials of LVMT.

A preliminary conference version of this work was introduced in Fang et al. [2013b]. In this extension, we (1) address the problem of city theme labeling to facilitate the visualization task; (2) add a separate section to review state-of-the-arts in related problems and highlight the differences of this work; (3) provide more implementation details for POI identification and POI theme discovery; (4) examine more baselines for experimental comparison, enrich the original Flickr-based dataset with mobile-oriented Instagram photos, and conduct experiments in both Singapore and Sydney; and (5) implement and introduce a demo system to demonstrate the potentials of LVMT.

The rest of the article is organized as follows. We review and discuss the related work in Section 2. From Section 3 to Section 5, we introduce the proposed solution for POI identification, POI theme discovery, city theme aggregation, and labeling, respectively. Section 6 shows the experimental results and analysis. Finally, in Section 7, we conclude the article with some open research problems.

2. RELATED WORK

Along with the popularity of positioning device and social media, geotagged multimedia resources have grown explosively and opened great opportunities for multimedia research and applications. Recently, many researchers have leveraged the geotagged photos toward both traditional and novel multimedia problems [Zheng et al. 2011; Luo et al. 2011]. According to the different application scenarios, we roughly categorize the related work into location estimation, landmark modeling, geographical photo organization, and semantic knowledge mining.

Geographical location estimation has drawn extensive interest from multimedia and computer vision communities. The early work of Hays and Efros [2008] studied geographical location prediction using a pure scene matching method and showed several important observations. Yu et al. [2011] developed an active query sensing system to suggest the best view for mobile location estimation and showed its effectiveness. Recently, Liu et al. [2012b] proposed to combine the techniques in large-scale image retrieval, 3D model reconstruction, and localization by 2D-to-3D matching toward robust location estimation in the urban areas. In Fang et al. [2013a], we have proposed to extract geo-informative attributes at the city level for both photo location estimation and geographical exploration.

Landmark modeling is another important research line. Quack et al. [2008] and Kennedy and Naaman [2008] employed the community photo collections by analyzing the geometric, visual, geographical, and textual cues to extract a clean set of landmark images. Zheng et al. [2009] built a landmark recognition engine to recognize landmarks at world-scale. Ji et al. [2011] proposed to mine representative landmarks from geotagged Flickr photo collections by modeling the reconstruction sparsity.

Geographically organizing and summarizing multimedia sources has received considerable research attentions. PhotoCamps by Naaman et al. [2004] is one of the earliest works to employ geographical information for image organization via a location and event hierarchy. Jaffe et al. [2006] developed a system to automatically select representative and relevant photos for a particular spatial region. Crandall et al. [2009] proposed to geographically organize a global collection of images by exploiting the visual features and associated context information. In Gao et al. [2010], the authors presented a travel guidance system W2Go (Where to Go) to recognize and rank the landmarks for travellers. More work has been devoted to positioning photos on the world map by exploiting textual metadata and social streams [Serdyukov et al. 2009; Hauff and Houben 2012a, 2012b].

The large volume of available geotagged multimedia documents with associated textual metadata provides opportunities to extract important semantic and social knowledge of the world. Kennedy et al. [2007] first attempted to generate a summary of important locations and events by mining large-scale geotagged photos. Rattenbury and Naaman [2009] addressed the problem of extracting place tags from the associated metadata of geotagged photos in Flickr. Papadopoulos et al. [2010] developed an online city exploration application to help users identify interesting spots within a city by providing photo clusters corresponding to landmarks or events. Many recent works [Liu et al. 2012a] have combined the geotagged photos with check-in records to discover regiosn of interest in a city, that is, the tourist attractions and popular venues among the locals.

Our work can be seen as the combination of geographical photo organization and semantic knowledge mining. In contrast to the existing studies, we investigate the problem of organizing photos considering both geographical information and the underlying semantic themes. The goal is to automatically generate a location visualization scheme consisting of POI level and city level. In terms of methodology, at the POI level, different from most clustering-based semantic mining methods, we propose an incremental learning scheme for POI theme discovery to address the noisy user-generated tags and photos. At city level, auxiliary Wikipedia pages are leveraged to match the aggregated city theme representation with compact labels.

3. POI IDENTIFICATION

The input of the proposed LVMT problem is a set of photos within a location: $\mathcal{P} = \{p_i\}_{i=1}^{N_{photo}^{all}}$. Each photo p is a tuple $(\mathbf{x}_p, \mathbf{t}_p, g_p, u_p)$ recording the visual feature vector $\mathbf{x}_p$, the user-generated text $\mathbf{t}_p$ including title, tags and description, the capture GPS coordinate g_p (if available), and the ID of the uploader u_p. The goal of POI identification is to find the highly photographed areas which are the potential places of interests.

Since in the crawled dataset, the GPS coordinates for some photos are not available, POI detection is first conducted by clustering the geotagged photos, and non-geo-tagged photos are then assigned to the detected POIs by POI estimation.

3.1. POI Detection

According to the definition of POI, we conduct clustering on the geotagged photos to find the popular places where people frequently take photos. Specifically, since the number of POIs is unknown in advance, MeanShift-based clustering is applied on the GPS coordinate of the geotagged photos. Each derived cluster is expected to correspond to one POI. To remove the unpopular POI, we measure the popularity of a cluster by (1) how many people have uploaded photos, and (2) how many photos have been uploaded in this place. Formally, the popularity score for POI l_k is calculated as:

$$F(l_k) = \sum_{i=1}^{N_{user}} \log(N_{photo}(u_i) + 1), \tag{1}$$

where $N_{photo}(u_i)$ is the number of photos uploaded by user u_i within l_k. The K POIs with the highest popularity scores are selected to construct the final POI set.

To roughly describe each POI in the POI set, we then extract the discriminative tags for each POI to construct a POI vocabulary $V_k = \{t_1, \ldots, t_N\}$, t_n is the n^{th} discriminative tag for the k^{th} POI. Inspired by the TagMaps method [Rattenbury and Naaman 2009], the tags are ranked considering their popularity, document frequency, and user frequency. A TF-IDF–like formulation is utilized to calculate a score for each tag t for POI l as:

$$Score(l, t) = \mathrm{tf}(l, t) \cdot \mathrm{idf}(t) \cdot \mathrm{uf}(l, t), \tag{2}$$

where $tf(l, t) = |N_{l,t}|$ is tag frequency, $idf(t) = |N|/|N_t|$ is the inverse document frequency in geo-tagged photos of POI l, and $uf(l, t) = |U_{l,t}|/|U_l|$ is the user frequency to remove the tags used by only a small portion of users. We keep the top 10 tags with highest $Score(l, t)$ as the discriminative tags for each POI. In this case, the tags more frequently appeared in this POI and widely used by many users are selected into the POI vocabulary.

3.2. POI Estimation

In our crawled dataset, the percentages of photos containing geo-tags in Flickr and Instagram are 23.5% and 77.8%, respectively. The large number of non-geotagged photos contain very important information for the POI theme discovery. In this subsection, with the detected POI and POI vocabulary, we introduce how to estimate the belonging POI of the non-geotagged photos.

Two POI estimation methods are examined to exploit the visual content and associated text, respectively. For each non-geotagged photo p, we first find its 10 visual nearest neighbors within the geotagged photos according to Euclidean distance $d(\mathbf{x}_p, \mathbf{x}_{p_i})$. The nearest neighbors constitute a photo set $\mathcal{N}_p$. Neighbor voting is utilized for POI estimation. The most probable POI for the non-geotagged photo p is obtained by:

$$l^* = \arg\max_l \sum_{p_i \in \mathcal{N}_p} \mathtt{I}_v(p_i, l) w_v(p_i, p), \tag{3}$$

where $\mathtt{I}_v(p_i, l)$ equals 1 if p_i belongs to POI l, and otherwise 0; $w_v(p_i, p)$ is the weight of photo p_i, which is in inverse proportion to its distance with p: $w_v(p_i, p) = \frac{(d(\mathbf{x}_p, \mathbf{x}_{p_i}))^{-1}}{\sum_{p_i \in \mathcal{N}_p} (d(\mathbf{x}_p, \mathbf{x}_{p_i}))^{-1}}$.

ALGORITHM 1: Incremental Learning-based POI Theme Discovery

Input: a set of photos $\mathcal{P}_l$ for POI l.
Output: multiple themes $\mathcal{A}_l = \{a_i\}_{i=1}^{K_l}$.

initialize: select one photo $p_s \in \mathcal{P}_l$ to initialize the photo set for theme 1: $\mathcal{P}_l^{a_1} = p_s$;
the number of themes $K_l \leftarrow 1$.

repeat
Salient Tag Extraction: extract/update the salient tags $\mathcal{T}_l^{a_i}$ that best describe the photo set $\mathcal{P}_l^{a_i}, i = 1, \dots, K_l$;
Photo Theme Assignment: add new photo p_n into one of the existing themes $a_1, \dots, a_{K^l}$, or initialize a new theme with $\mathcal{P}_l^{a_{K_l+1}} = p_n$; $\quad K_l \leftarrow K_l + 1$;
until photos $p_n \in \mathcal{P}_l$ exhausted

A similar method is utilized to exploit the associated text for POI estimation. For a given photo p, the most probable POI is obtained by:

$$l^* = \arg\max_l \sum_{t \in V_l} \mathtt{I}_t(t, \mathbf{t}_p) w_t(t, l), \tag{4}$$

where $\mathtt{I}_t(t, \mathbf{t}_p)$ equals 1 if the associated text $\mathbf{t}_p$ contains the POI tag t, and otherwise 0; $w_t(t, l)$ is the weight of tag t in POI vocabulary V_l, which is calculated by $w_t(t, l) = \frac{Score(t,l)}{\sum_{t \in V_l} Score(t,l)}$. To consider both the visual content and associated text, we can simply combine Equations (3) and (4) to obtain the belong POI as follows:

$$l^* = \arg\max_l \left(\sum_{p_i \in \mathcal{N}_p} \mathtt{I}_v(p_i, l) w_v(p_i, p) + \sum_{t \in V_l} \mathtt{I}_t(t, \mathbf{t}_p) w_t(t, l) \right) \tag{5}$$

After POI detection and POI estimation, all the geotagged and non-geotagged photos belonging to POI l constitute the photo set $\mathcal{P}_l$.

4. POI THEME DISCOVERY

Now we consider the problem of theme discovery for each POI l. The goal is to discover the multiple popular themes within the POI and allocate the photos within $\mathcal{P}_l$ to the respective themes. Each theme is expected to be represented by a set of salient tags and exemplary photos. As aforementioned, the challenges for POI theme discovery include the large visual variance and high tag noise. To address these issues, we propose an incremental learning-based method. The basic workflow is as follows: (1) for a given POI, one photo with abundant text information is selected to initialize theme 1; (2) salient tags are extracted from the theme photo set to describe the existing themes; (3) according to the extracted salient tags and theme photos, new photos are added into existing themes or initialize new themes. This process iterates, until all the photos are assigned. The algorithm for theme discovery for one POI is summarized in Algorithm 1. In the following, we introduce the two basic components of *salient tag extraction* and *photo theme assignment*, respectively.

4.1. Salient Tags Extraction

Given photos $\mathcal{P}_l$ within the POI l and photos $\mathcal{P}_l^a$ of one theme a, we aim to extract the salient tags $\mathcal{T}_l^a$ that well-describe the visual and semantic content in this theme. Two characteristics are assumed for these salient tags:

- **Discriminative:** the salient tag t should be observed more frequently in this theme than in other themes. This is similar to the requirement on the tags in POI vocabulary in Section 3.1.
- **Consistent:** the photos associated to the salient tag t should be visually similar to each other. This is to encourage the extracted salient tags to describe a concrete concept that is visually consistent, for example, an object or a scene.

Regarding the discriminative assumption, we measure the tag observing probability by examining the tag co-occurrence. Specifically, the observing probability of tag t_i in photo set $\mathcal{P}_l^a$ is calculated as:

$$p(t_i|\mathcal{P}_l^a) = \frac{N(t_i \cap \mathcal{P}_l^a)}{N(\mathcal{P}_l^a)}, \tag{6}$$

where $N(\mathcal{P}_l^a)$, $N(t_i \cap \mathcal{P}_l^a)$ denote the number of photos in theme a and photos associated with tag t_i in theme a, respectively. Similarly, the observing probability of tag t_i in photo set $\mathcal{P}_l$ is calculated as: $p(t_i|\mathcal{P}_l) = \frac{N(t_i \cap \mathcal{P}_l)}{N(\mathcal{P}_l)}$. We define the discriminative score of tag t_i with respect to theme a in POI l as:

$$\mathcal{R}(t_i, \mathcal{P}_l^a) = f(p(t_i|\mathcal{P}_l^a) - p(t_i|\mathcal{P}_l)), \tag{7}$$

where $f(\cdot)$ is the standard sigmoid function $f(x) = \frac{1}{1+e^{-x}}$, which is monotonically increasing.

Regarding the consistent assumption, pairwise similarity is utilized to measure the consistency of photos associated with the tag t_i. Denoting $\mathcal{P}_l^a(t_i)$ as the set of photos containing tag t_i in $\mathcal{P}_l^a$, we use Gaussian kernel to calculate the visual pairwise similarity between photos in $\mathcal{P}_l^a(t_i)$ as:

$$Sim(\mathcal{P}_l^a(t_i)) = \frac{1}{(N(\mathcal{P}_l^a(t_i)))^2} \sum_{u,v \in \mathcal{P}_l^a(t_i)} K_\sigma(x_u - x_v), \tag{8}$$

where $K_\sigma(\cdot)$ is the Gaussian kernel function with radius parameter σ, and σ is assigned as the median value of all pairwise Euclidean distances between photos. The consistent score of tag t_i with respect to theme a in POI l is defined as:

$$\mathcal{C}(t_i, \mathcal{P}_l^a) = 1 - f(Sim(\mathcal{P}_l^a(t_i))) \tag{9}$$

With the derived discriminative and consistent scores of tag t_i, we now present the formulation for extracting the salient tag set $\mathcal{T}_l^a$ as follows:

$$(\mathcal{T}_l^a)^* = \arg\max_{\mathcal{T}_l^a} \left\{ \frac{1}{N(\mathcal{T}_l^a)} \sum_{t_i \in \mathcal{T}_l^a} \lambda \mathcal{R}(t_i, \mathcal{P}_l^a) + (1-\lambda)\mathcal{C}(t_i, \mathcal{P}_l^a) \right\}, \tag{10}$$

where $N(\mathcal{T}_l^a)$ is the number of the tags in the salient tag set, and $\lambda \in [0, 1]$ is the weighting parameter. $\phi(t_i, \mathcal{P}_l^a) \triangleq \lambda \mathcal{R}(t_i, \mathcal{P}_l^a) + (1-\lambda)\mathcal{C}(t_i, \mathcal{P}_l^a)$ is defined as the saliency score of tag t_i. Equation (10) is a non-linear integer programming problem which is computationally intractable [Boyd and Vandenberghe 2004]. Regarding one theme for a POI, only a small set of tags are potentially salient. In implementation, we first filter out the tags with a small value of $\mathcal{R}(t_i, \mathcal{P}_l^a)$ to construct a tag candidate set $\mathcal{T}$, which will significantly reduce the number of tags to be examined. Then, we resort to a greedy strategy to sequentially select tag t_i with the highest saliency score into the salient tag set $(\mathcal{T}_l^a)^*$, by solving $\max_{t_i} \lambda \mathcal{R}(t_i, \mathcal{P}_l^a) + (1-\lambda)\mathcal{C}(t_i, \mathcal{P}_l^a)$ from the difference set $\mathcal{T} \setminus (\mathcal{T}_l^a)^*$. The selection process ends when $N((\mathcal{T}_l^a)^*)$ reaches a pre-fixed number or $\mathcal{T} \setminus (\mathcal{T}_l^a)^* = \emptyset$.

Note that in the incremental learning process, the salient tag set for the existing themes need to be updated when a new photo is assigned. In one update of a certain theme, it is not necessary to re-process all the tags. We only need to examine the saliency score $\phi(t_i, \mathcal{P}_l^a)$ of the tags associated with the new photo, and update the salient tag set by comparing them with the salient scores of previous salient tags. This can largely reduce the computational complexity for POI theme discovery.

4.2. Photo Theme Assignment

Given the salient tags and assigned photos for each of the existing themes within a POI, for each new photo, we need to assign a theme index or initialize a new theme. Both the text metadata and visual feature of the new photo are leveraged: (1) Since the extracted salient tags semantically represent the corresponding themes, we first examine the relevance between the text metadata of new photos with the salient tags, and filter out themes with low relevances to generate a candidate theme set. (2) Different from the photos within one POI, we expect the photos assigned to the same theme to have low visual variance, for example, landmark, a specific view of the POI. Therefore, we model theme assignment as a cost minimization problem on reconstructing the visual feature of the new photo on the candidate theme photo sets. We now elaborate the two steps as follows.

Given a new photo p and its text information $\mathbf{t}_p$, a similar formulation to Equation (4) in POI estimation is utilized to measure the semantic relevance between p and the salient tags $\mathcal{T}_l^a$ for theme a in POI l:

$$SR(\mathbf{t}_p, \mathcal{T}_l^a) = \sum_{t \in \mathcal{T}_l^a} \mathtt{I}_t(t, \mathbf{t}_p) w_t(t, \mathcal{T}_l^a), \tag{11}$$

where $\mathtt{I}_t(t, \mathbf{t}_p)$ outputs 1 if there exists a salient tag $t \in \mathcal{T}_l^a$ in $\mathbf{t}_p$, and otherwise 0; $w_t(t, \mathcal{T}_l^a) = \frac{\phi(t, \mathcal{T}_l^a)}{\sum \phi(t_i, \mathcal{T}_l^a)}$ is the weight of salient tag t in $\mathcal{T}_l^a$. Based on the semantic relevance score, we select the correlated themes $A_{correlated}$ for new photo p as follows:

$$A_{correlated} = \{a | SR(\mathbf{t}_p, \mathcal{T}_l^a) > thre_A\}, \tag{12}$$

where $thre_A$ is a threshold value. The themes with SR score higher than $thre_A$ are selected into the candidate theme set and used to determine the theme of p in later sparse representation-based classification. If a new photo p is not semantically correlated to any of the existing themes, all the existing themes are remaining to construct the candidate theme set.

Given the candidate theme set, we then consider the visual information of the new photo to assign the belonging theme. As mentioned earlier, we expect existence of similar or duplicate visual features coming from the same theme within a POI. Inspired by the success of sparse representation in face recognition [Wright et al. 2009] and other visual classification tasks, we reconstruct the new photo's visual feature using an over-complete dictionary constituted by all photos from existing themes and classify the photo to the theme with the lowest reconstruction error. Specifically, we define a dictionary matrix B as the concatenation of assigned photos from all the existing $N(A_{correlated})$ candidate themes

$$B \triangleq [B_1, B_2, \ldots, B_{A_{correlated}}] \in \mathbb{R}^{m \times \sum_{a_i} N(\mathcal{P}_l^{a_i})},$$

where $B_i = [\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,N(\mathcal{P}_l^{a_i})}]$, and $N(\mathcal{P}_l^{a_i})$ is the number of photos in theme a_i within POI l. Then, given a new photo p and its visual feature vector $\mathbf{x}_p \in \mathbb{R}^m$, we reconstruct

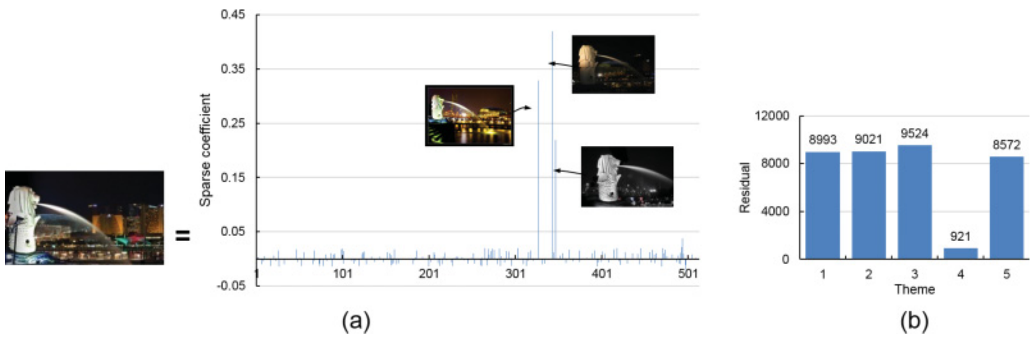

Fig. 2. The sparse representation for a new photo: (a) sparse coefficients on all the theme photos; (b) residuals of different themes.

$\mathbf{x}_p$ using B as the coding dictionary as follows:

$$\mathbf{x}_p = \sum_{i=1}^{A_{correlated}} B_i \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}, \tag{13}$$

where $\boldsymbol{\alpha}_i = [\alpha_{i,1}, \ldots, \alpha_{i,N(\mathcal{P}_l^{a_i})}]$ is the reconstruction coefficients, and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is a noise vector. We learn the reconstruction coefficients by solving the least square error with ℓ_1-norm regularization, which is formulated as:

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \|B\boldsymbol{\alpha} - \mathbf{x}\|_F^2 + \tau\|\boldsymbol{\alpha}\|_1, \tag{14}$$

where τ is the regularization parameter. L1-Homotopy [Asif and Romberg 2009] is utilized to solve the aforementioned optimization problem.

With the derived reconstruction coefficients, we now classify p based on how well the photos from each candidate theme reproduce p. For each candidate theme a_i, we calculate the residual $r_i(\mathbf{x}_p) = \|\mathbf{x}_p - B_i\boldsymbol{\alpha}_i\|_2$, where $i = 1, \ldots, A_{correlated}$. We then classify p to the theme with the smallest residual:

$$a(p) = \arg\min_i r_i(\mathbf{x}_p) \tag{15}$$

Figure 2 illustrates an example for sparse reconstruction-based theme assignment. Figure 2(a) shows the coefficients of 513 photos from 5 existing candidate themes for a new photo presented on the left. The photos corresponding to the three largest coefficients are also presented. The three photos are all from theme 4. Figure 2(b) shows the residuals with respect to the 5 themes. It is shown the residual for theme 4 is much smaller than those of other themes, with the ratio between the two smallest residuals of theme 5 and theme 4 as $8,572 : 921 = 9.3$. The new photo is confidently assigned to the existing theme 4.

When a new photo cannot be confidently assigned to any of the existing themes, we need to initialize a new theme. With the derived residuals for each existing theme, if the ratio between the two smallest residuals is smaller than 2, we think that the new photo cannot be well represented with low error by any of the existing theme photos. In this case, a new theme will be initialized with this photo.

Through the incremental learning-based POI theme discovery, we can obtain multiple POI themes $\mathcal{A}_l = \{a_i\}_{i=1}^{K_l}$, each of which is associated with the salient tags $\mathcal{T}_l^{a_i}$ and assigned photos $\mathcal{P}_l^{a_i}$. For one POI l, we rank its multiple POI themes considering both the number of contributing users and photos, which is the same as Equation (1).

To visually represent each POI theme, affinity propagation [Frey and Dueck 2007] is conducted over $\mathcal{P}_l^{a_i}$ to obtain the representative photos.

5. CITY THEME AGGREGATION AND LABELING

5.1. City Theme Aggregation

After obtaining POIs with multiple POI themes, we aim to discover themes at the city level by aggregating similar POI themes. After POI theme discovery, we have obtained K_l POI themes for each POI l. We first represent each POI theme on the same semantic space. All the salient tags of POI themes are fused to construct a unique tag vocabulary for the city, which is denoted as $V^{city} = \{t_i^{city}\}_{i=1}^{D}$, constituting of D distinct tags. Each theme a_i from POI l is mapped onto this vocabulary based on its salient tags $\mathcal{T}_l^{a_i}$ and represented as a D dimensional vector:

$$\begin{aligned} &\boldsymbol{\mu}_l^{a_i} = [\mu_1, \ldots, \mu_D]; \quad \mu_i = 1 \text{ if } t_i^{city} \in \mathcal{T}_l^{a_i}, \text{otherwise } \mu_i = 0; \\ &\quad l = 1, \ldots, K; a_i = 1, \ldots, K^l, \end{aligned} \tag{16}$$

where K is the number of POIs and K^l is the number of themes within POI l. K-means is conducted to cluster the D-dimensional theme vectors $\boldsymbol{\mu}_l^{a_i}$. The number of clusters is empirically predefined. Each of the derived clusters represents a high-level city theme c_i, with the frequent salient tags of the POI themes in this cluster as the semantic representation $\mathcal{T}^{c_i}$, and the representative photos of the POI themes in this cluster as the visual representation $\mathcal{P}^{c_i}$.

5.2. Theme Labeling

After city theme aggregation, the semantic representation for the obtained city theme is a tag set, fore example, $\mathcal{T}^{c_i} = [garden, reserve, flower, nature, park, \ldots]$. The task of theme labeling is to assign a concise and more readable label to each of the obtained city themes, for example, *natural scene*.

Theme/topic representation is an important problem in topic modeling. The standard way of representing a topic is to use the top-N terms according to the associated marginal probabilities. This entails a significant cognitive load in interpretation [Lau et al. 2011]. It is necessary to develop solutions for topic labeling to transfer a list of terms to a single label. Various methods have been proposed to achieve this goal via advanced random processes and graphical models [Chang et al. 2009; Ramage et al. 2011]. One critical step in topic labeling is to generate label candidates, which can well cover the semantics described in the topics to be interpreted. The task of city theme labeling in this work can be viewed as a special case of topic labeling in the city domain. Wikipedia provides a comprehensive introduction of cities in the subcategory "Geography by city." For example, Singapore is introduced in its Wikipedia page along with directories of history, geography, climate, economy, religion, culture, etc.[3] These directory phrases provide high-quality label candidates. In this work, we introduce a simple solution to label the discovered city themes by exploiting the Wikipedia pages of the corresponding cities.

Candidate Label Generation: With Wikipedia Application Programming Interface (API), we first crawl the directory phrases in both the root level and leaf level to construct the candidate label set $\mathcal{L}$. Since the Wikipedia directory phrases may not cover all the UGC-based semantics, to guarantee that each discovered theme will have a label correspondence and allow the possibility of labeling a theme using its own

[3]https://en.wikipedia.org/wiki/Singapore.

salient tag, for each city theme c_i, we also add the top-5 frequent salient tags from $\mathcal{T}^{c_i}$ as the candidate labels to construct the candidate label set $\mathcal{L}_{c_i}$.

Label Matching: A good label is expected to fully capture the theme semantics represented by the salient tag set $\mathcal{T}^{c_i}$. Since the candidate label set includes both Wikipedia directory phrases and top-5 salient tags, we introduce the methods to measure their relevances, respectively.

To measure the relevance between a directory phrase and $\mathcal{T}^{c_i}$, we conduct topic modeling on the Wikipedia text corresponding to the directory phrases to obtain the phrase representation. Specifically, with the corresponding text as document, the words appeared in the Wikipedia page constructing the vocabulary,[4] and the standard LDA is utilized to obtain the document-topic distribution $p(\boldsymbol{\theta}|l)$ for each directory phrase l. The semantic relevance between a directory phrase l and the salient tag set of city theme c_i is defined as:

$$Rel_1(l, c_i) = \log p(\mathcal{T}^{c_i}|l) = \sum_{t \in \mathcal{T}^{c_i}} \sum_{\theta_k} \log p(t|\theta_k) p(\theta_k|l), \tag{17}$$

where $p(t|\theta_k)$ is the topic-word distribution: $p(t|\theta_k) = 0$ when the salient tag t is not included in the topic modeling vocabulary. The phrases with high relevance scores are expected to be the confident labels for the city themes.

When the relevance score $Rel_1(l, c_i)$ for all the candidate phrases are below a predefined threshold, we turn to examine the representative capability of single salient tag $t_i \in \mathcal{T}^{c_i}$. The most representative salient tag is expected to have the highest average conditional probability given each of the other salient tags. The conditional probability of two salient tags is measured by their co-occurrence in the UGC tags:

$$p(t_i, t_j) = \frac{N(t_i \bigcap t_j)}{N(t_i)},$$

where $N(t_i \bigcap t_j)$ denotes the number of photos containing both tag t_i and t_j. The semantic relevance between a salient tag t and the salient tag set of city theme c_i is then calculated as:

$$Rel_2(t, c_i) = \frac{1}{N(\mathcal{T}^{c_i})} \sum_{t_j \in \mathcal{T}^{c_i}; t_j \neq t} p(t_i, t_j) \tag{18}$$

The salient tag with the highest $Rel_2(t, c_i)$ is selected as the label for theme c_i. Note that the salient tag only serves as complementary label candidates when no confident directory phrases are available. Generally, the lower the semantic consistency among the theme salient tags, the more difficult one single salient tag describes the theme.

5.3. Location Visualization

After city theme aggregation and labeling, each city will be involved with multiple city themes, with each city theme c_i associated with a theme label l_{c_i}, a salient tag set $\mathcal{T}^{c_i}$, and representative photos $\mathcal{P}^{c_i}$. At the city level of location visualization, we provide the multiple city themes as shown in Figure 4. At the POI level, the POIs clustered to the city themes are shown under the corresponding city themes. The respective POI themes, their salient tags, and representative photos will be further presented when choosing certain POIs. This will be illustrated with the introduction of a demo system in the experiments.

[4]The references that linked to the text are also crawled and included for topic modeling. The number of topics is set as the same with the number of directory phrases.

Table I. Dataset Statistics

	Flickr	Instagram	Total
Singapore	110,846 (26,623)	183,229 (142,593)	294,075 (169,216)
Sydney	89,775 (20,553)	202,993 (157,995)	292,768 (178,548)

Note: The numeric in the bracket indicates the number of geotagged photos.

6. EXPERIMENTS

6.1. Dataset and Experimental Setting

Our dataset was constructed by crawling photos using Flickr and Instagram APIs. In Flickr, we use "Singapore" and "Sydney" as the query word and collected all the returned photos with the associated information, including title, tags, description, and GPS coordinates (if available). The initial Flickr dataset consisted of 263,953 photos for Singapore and 195,228 photos for Sydney. A pre-filtering process is performed to remove the duplicate photos and photos with GPS coordinates outside of Singapore and Sydney, respectively. We also remove photos with incomplete associated textual metadata of title, tags, and description. In this way, 110,846 photos for Singapore and 89,775 photos for Sydney remained. In Instagram, we use the central coordinates of Singapore and Sydney as the query locations to search the photos uploaded in the given areas. Since Instagram API only returns photos uploaded in the past 7 days, we continuously crawled the data from March 1, 2013 to February 1, 2014. After removing the photos with no text metadata, this resulted in 183,229 photos for Singapore and 202,993 photos for Sydney. Table I summarizes the basic statistics for the crawled data.

For photo visual content, we extract five types of visual features to form a 809-dimension vector $\mathbf{x}_p$, including 81-dimension color moment, 37-dimension edge histogram, 120-dimension wavelet texture feature, 59-dimension LBP feature, and 512-dimension GIST feature. For semantic tags, we removed stop words, time and numeric related words, camera configuration words, and some general frequent tags. The resultant tag vocabulary consists of 10,000 tags for POI theme discovery. The weighting parameter λ in Equation (10) is empirically set to 0.7 to emphasize on the discriminative assumption. The threshold parameter $thre_A$ in Equation (12) is chosen as 0.005 through qualitative cross-validation. For city theme aggregation, we extract the top 50 salient tags of each POI theme to construct the vocabulary for city theme aggregation. The number of clusters in K-means for city theme aggregation is set as 20 for both Singapore and Sydney.

6.2. Evaluation of POI Identification

We use all the geotagged photos from Flickr and Instagram to detect POIs. The bandwidth for meanshift is set as 0.023 according to the scale of photo distribution. The detected POIs are ranked based on their popularity score calculated by Equation (1). The top 20 POIs with the highest popularity score remained for Singapore and Sydney, respectively. Table II shows the top 7 detected POI of Singapore city with the associated POI vocabularies. The POI name is obtained by issuing the GPS-coordinates of the POI cluster center to the Flickr API. It is shown that the detected POIs cover very popular places of interest in Singapore. For each POI, 10 tags are extracted to construct the POI vocabulary. We can see that the tags in the POI vocabulary generally well describe the corresponding POI. For example, for the POI of "Downtown, Central Singapore," vocabulary tags such as *merlion, marina, sands, skyline* characterize the most popular landmarks in the downtown area.

For POI estimation, we compare between the three settings of *visual-based*, *textual-based*, and *combined*, according to Equations (3), (4), and (5), respectively. For each POI, we use half of the geotagged photos for training and the other half for testing.

Table II. POI Vocabulary of the Top 7 POIs in Singapore

POI	POI Vocabulary
Downtown, Central Singapore	bay marina night merlion skyline sands river asia esplanade marinabay
Central Catchment Reserve	reserve catchment macritchie turf whirlwind asia night beautiful people skyline
Sentosa, South West	sentosa island universal studios resortsworld resorts vivocity shrek capella palawan
Jurong Town, South West	jurong jurongtown chinesegarden garden singaporechinesegarden juronglake lanternfestival pagoda shoebill lake
Singapore Changi Airport	changi airport airlines singaporeairlines terminal airbus departure plane flight airplane
Jalan Kayu, Central Singapore	jalan kayu punggol punggolbeach beach lowerseletarreservoir waterway sengkang reservoir
Queenstown, Central Singapore	queenstown hawparvilla buona portsdownroad dover holland haw westcoast hollandvillage tigerbalm

Table III. mAP of POI Estimation for Singapore and Sydney

	Singapore			Sydney		
Setting	Flickr	Instagram	all	Flickr	Instagram	all
Visual-based	0.3218	0.3496	0.3588	0.3335	0.3502	0.3617
Textual-based	0.4205	0.3817	0.4192	0.4177	0.3692	0.4052
Combined	0.5233	0.5064	**0.5492**	0.5295	0.4982	**0.5527**

Mean Average Precision (mAP) is utilized as the evaluation metric. To examine the POI estimation performance in different photo sharing websites, we calculate the mAP in the Flickr and Instagram dataset, respectively, and show the results in Table III. We have the following observations: (1) *Textual-based* setting generally outperforms *Visual-based*. This is possibly due to the large visual variance within the same POI. Actually, the Average Precision (AP) varies significantly from POI to POI. We find that *Visual-based* AP for the POI "Singapore Changi Airport" is as high as 0.5652 due to its consistent visual appearances. (2) For *Visual-based* setting, the photos in Instagram have a higher mAP than that in Flickr. Instagram has more geotagged photos, which helps detect the POIs and provides more confident visual neighbors. (3) For *Textual-based* setting, the photos in Instagram obtain a much lower mAP than those in Flickr. Developed for the mobile platform, Instagram attracts less textual annotation from users than Flickr, which limits the potentials for *Textual-based* POI estimation. In our collected dataset, the average number of tags for each photo in Instagram and Flickr is 3.92 and 6.01, respectively. (4) *Combined* improves the mAP significantly over *Textual-based*. This demonstrates that the textual metadata and visual content play complementary roles in POI estimation. By combining the two types of features, we can enrich the photo set for each POI and facilitate the following POI theme discovery task.

6.3. Evaluation of POI Theme Discovery

Qualitative Evaluation: We exploit all the photos from Flickr and Instagram to conduct POI theme discovery. We qualitatively evaluate the discovered POI themes for three of the detected POIs in Singapore. In Figure 3, we show the top-3 POI themes for each of the POIs, represented by the salient tags and representative photos. It is shown that within the same POI theme, the extracted photos and tags are generally consistent and jointly describe one aspect of the POI. For example, for the POI of

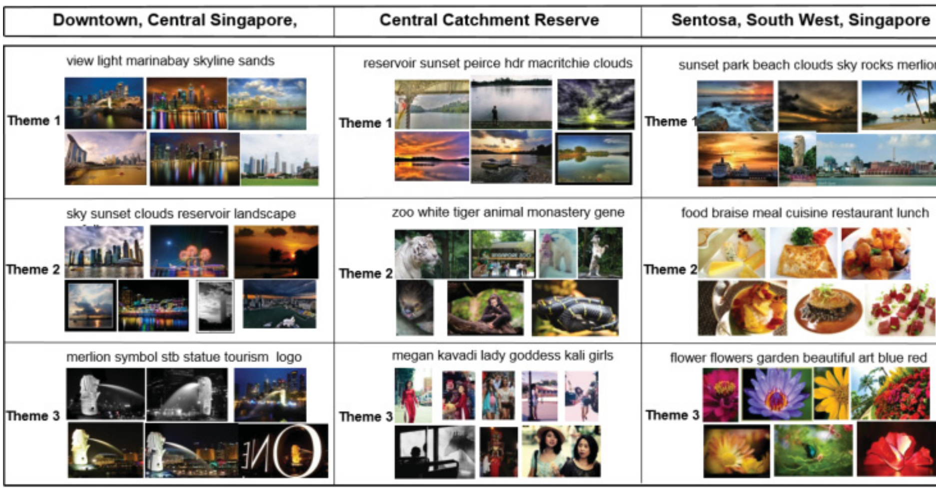

Fig. 3. POI themes of three POIs in Singapore.

"Sentosa, South West Singapore," the three discovered POI themes address topics of natural scene, food, and vegetation, respectively.

Quantitative Evaluation: For quantitative evaluation, since no ground-truth POI themes are available, we manually construct the POI theme sets for the top-7 POIs of Singapore. The manual labeling process is: given the photo set $\mathcal{P}_l$ for each POI l, according to the visual content and associated textual information, we group the photos describing common patterns such as objects, scenes, and events into one POI theme. For each POI l, $\{\hat{\mathcal{P}}_l^{a_1}, \cdots, \hat{\mathcal{P}}_l^{K_m}\}$ denotes the photo sets for the manually labeled POI themes. We use the traditional clustering metric *purity* [Zhao and Karypis 2001] to evaluate the performance of the discovered POI themes, which is defined as

$$\text{Purity}(l) = \frac{1}{N(\mathcal{P}_l)} \sum_{i=1}^{K_l} \max_{j=1:K_m} N(\mathcal{P}_l^{a_i} \cap \hat{\mathcal{P}}_l^{a_j}), \tag{19}$$

where $N(\mathcal{P}_l^{a_i} \cap \hat{\mathcal{P}}_l^{a_j})$ is the number of photos in the intersection set between the discovered POI theme a_i and manually labeled POI theme a_j. A higher purity score indicates the better accordance with the human understanding, and thus, a possibly satisfied POI theme discovery performance.

The essence of POI theme discovery is photo clustering, considering the visual and semantic information. Therefore, we compare the proposed incremental learning-based solution with two classical clustering methods:

—***K*-Means**, which considers the distance from instances to the K cluster centers. We apply K-Means on the POI photos represented by the combined feature vector of visual content and textual metadata. Visual content is represented by the 809-D visual feature as introduced earlier. Textual metadata is represented in the 10,000 tag vocabulary.
—**Spectral Clustering**, which exploits pairwise similarities between the data instances. We exploit the textual metadata and visual content of each POI photo to compute a similarity matrix and use the weighted sum to construct the final similarity matrix for spectral clustering. The weight is empirically set as 0.5.

Table IV. Purity of POI Theme Discovery for Seven POIs in Singapore

POI	*K*-Means	Spectral Clustering	Incremental Learning
Downtown, Central Singapore	0.4992	0.4898	**0.711**
Central Catchment Reserve	0.2531	0.2947	**0.7995**
Sentosa, South West	0.3907	0.4327	**0.816**
Jurong Town, South West	0.6573	0.6808	**0.8482**
Singapore Changi Airport	0.5157	0.6493	**0.8283**
Jalan Kayu, Central Singapore	0.6396	0.6493	**0.8181**
Queenstown, Central Singapore	0.6065	0.6316	**0.7669**
Avg.	0.5091	0.5468	**0.7975**

Fig. 4. Visualizing Singapore from multiple city themes.

K-Means and Spectral Clustering need to specify the number of clusters in advance. We set the number of POI themes for each POI as 10. Table IV presents the evaluation results. We can see that *Spectral Clustering* performs slightly better than *K-Means*. The proposed *Incremental Learning*-based solution significantly outperforms the two baselines in all the examined POIs. The average improvement over *Spectral Clustering* and *K-Means* is 45.8% and 56.7%, respectively. The requirements of the semantic representation on discriminative and consistent effectively alleviate the issue of noisy tags. The sparse reconstruction-based theme assignment enjoys the advantage of locating the most visually similar photos. Without special treatment, *Spectral Clustering* and *K-Means* are vulnerable to the noisy tags and photos with large visual variance. Though *Spectral Clustering* and *K-Means* are not specially modified to solve the POI theme discovery task and the relevant parameters are not optimized, we believe this significant improvement and high purity score of 0.7975 have demonstrated the effectiveness of an *Incremental Learning*-based solution for POI theme discovery.

6.4. Evaluation of City Theme Generation and Location Visualization

In Figure 4, we show the Singapore visualization map from the discovered top-7 city themes. For each city theme, the mined theme label, salient tags, and the belonging POIs are provided. It is shown that some city themes are labeled by the Wikipedia

Fig. 5. LVMT demo system: city-level visualization.

Table V. Evaluation of City Theme Labels

	Wikipedia	Salient tag	Combined
Singapore	3.14	3.17	**3.86**
Sydney	3.06	2.98	**3.91**

directory phrases (e.g., "Geography," "Economy," "Culture"), and some are labeled by one of its most representative salient tags (e.g., "Aeroplane," "People," "Animal"). With the association with the belonging POIs, we can have easy access to the specific places of interest under certain themes, which enables a natural and comprehensive way for Singapore exploration.

We also implemented a demo system based on the proposed LVMT solution. Figure 5 provides a snapshot of the demo system for Singapore. In the user interface, two levels of location visualization can be switched on the left. The POIs with the highest popularity scores are shown. Each POI is marked by a red icon on the map, where the number in the icon indicates its rank. At the city level, as illustrated in Figure 5, the aggregated city themes are shown with the theme labels, representative photos, and salient tags. The belonging POIs of this city theme are also highlighted on the map (POI-5 and POI-8 for the city theme of "Aeroplane"). When one POI is selected, a window will pop up near the POI icon to show the POI themes and representative theme photos of this POI.

Quantitative Evaluation: To quantitatively evaluate the matched city theme labels, we invite six graduate students from Singapore and Sydney as annotators to conduct the evaluation task. For each city theme c_i, the salient tags and representative photos are shown, followed by the candidate labels $\mathcal{L}_{c_i}$. The annotators were asked to rate each candidate label based on the descriptive capability of the corresponding themes. A rating of $1-5$ is allowed, with 1 indicating "incomplete to represent, or unrelated to the theme" and 5 indicating "well represent the theme." The goal is to examine how the annotators rate the matched theme labels. The averaged rates on all the city themes of Singapore and Sydney are shown in Table V. In addition to the rates for the final matched labels (*Combined*), we also show the results for the two label candidate subsets, that is, the Wikipedia directory phrase with the highest $Rel_1(l, c_i)$ in Equation (17) (*Wikipedia*) and the salient tag with the highest $Rel_2(t, c_i)$ in Equation (18) (*Salient tag*). We observe that only using the Wikipedia directory phrase or the most

Table VI. Evaluation of the Overall LVMT Solution

	Consistency	Relatedness	Satisfaction
Singapore	3.96	4.17	4.13
Sydney	3.59	4.06	4.12

representative salient tags fails to describe the city themes. The highest possible rates of all label candidates, according to the annotator rating, are 4.14 and 4.23, respectively. Achieving the average rates of 3.86 and 3.91, combining the two label candidate subsets, generally finds the labels most consistent to the human understanding.

We conduct another user study to evaluate the user experience on the LVMT demo system. Twenty participants are invited to rate the system from three criteria: (1) *consistency*, the level of consistency within the representation of the city themes, that is, the theme labels, the salient tags, and the representative photos; (2) *relatedness*, the level of relation between the discovered city themes to the belonging POIs; and (3) *satisfaction*, the level of satisfaction that the overall LVMT scheme facilitates the task of city understanding and exploration. For each criterion, a rating of $1-5$ is allowed with 1 indicating the worst and 5 indicating the best. The averaged rates over the 20 participants are summarized in Table VI. It is shown that the participants generally give positive feedback to all three criteria. We can see that while the *consistency* is relatively low for Sydney, the overall *satisfaction* is guaranteed, which demonstrates the advantage of the proposed integrated solution, in spite of inferior performance from certain components.

7. CONCLUSIONS

With user-generated annotation as the weak semantics from social media, and photo geotagging as the analogy to mobile activity log, we have exploited social-mobile information to present a novel location visualization scheme. Both geographical and semantic knowledge are leveraged to organize photos into multiple themes. A two-level POI-city visualization solution framework is proposed, where an incremental learning approach is introduced for POI theme discovery and city theme is then obtained by aggregating similar POI themes. Experiments on a dataset of Singapore and Sydney collected from Flickr and Instagram have shown its advantage in deriving compact city themes and improving user experiences.

This work provides one substantial attempt toward the goal of automatically theme mining and location visualizing. The future work can be extended mainly in two directions. (1) Applications, such as travel recommendation and location-based POI search, will be developed based on the proposed visualization scheme. (2) We will apply our algorithm to more locations and evaluate its generalization ability in a large-scale dataset. Moreover, quantitative evaluation metrics for the visualization tasks will be specially designed and examined.

REFERENCES

Muhammad Salman Asif and Justin K. Romberg. 2009. Dynamic updating for L1 minimization. *CoRR* abs/0903.1443 (2009).

S. P. Boyd and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L. Boyd-Graber, and David M. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in Neural Information Processing Systems*. 288–296.

David J. Crandall, Lars Backstrom, Daniel P. Huttenlocher, and Jon M. Kleinberg. 2009. Mapping the world's photos. In *WWW*. 761–770.

Alberto Del Bimbo, K. Selcuk Candan, Jiebo Luo, Tao Mei, Nicu Sebe, Heng Tao Shen, Cees GM Snoek, Rong Yan, and others. 2014. Guest editorial special section on socio-mobile media analysis and retrieval. *IEEE Trans. Multimedia* 3, 16 (2014), 586–587.

Quan Fang, Jitao Sang, and Changsheng Xu. 2013a. Giant: Geo-informative attributes for location recognition and exploration. In *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 13–22.

Quan Fang, Jitao Sang, Changsheng Xu, and Ke Lu. 2013b. Paint the city colorfully: Location visualization from multiple themes. In *MMM*. 92–105.

Brendan J. Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *Science* 315 (2007), 972–976.

Yue Gao, Jinhui Tang, Richang Hong, Qionghai Dai, Tat-Seng Chua, and Ramesh Jain. 2010. W2Go: A travel guidance system by automatic landmark ranking. In *Proceedings of the 18th ACM International Conference on Multimedia*. ACM, 123–132.

Qiang Hao, Rui Cai, Changhu Wang, Rong Xiao, Jiang-Ming Yang, Yanwei Pang, and Lei Zhang. 2010. Equip tourists with knowledge mined from travelogues. In *WWW*. 401–410.

Claudia Hauff and Geert-Jan Houben. 2012a. Geo-location estimation of Flickr images: Social web based enrichment. In *ECIR*. 85–96.

Claudia Hauff and Geert-Jan Houben. 2012b. Placing images on the world map: A microblog-based enrichment approach. In *SIGIR*. 691–700.

James Hays and Alexei A. Efros. 2008. im2gps: Estimating geographic information from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Alexander Jaffe, Mor Naaman, Tamir Tassa, and Marc Davis. 2006. Generating summaries and visualization for large collections of geo-referenced photographs. In *Multimedia Information Retrieval*. 89–98.

Rongrong Ji, Ling-Yu Duan, Jie Chen, Hongxun Yao, Junsong Yuan, Yong Rui, and Wen Gao. 2012. Location discriminative vocabulary coding for mobile landmark search. *Int. J. Comput. Vision* 96, 3 (2012), 290–314.

Rongrong Ji, Yue Gao, Bineng Zhong, Hongxun Yao, and Qi Tian. 2011. Mining Flickr landmarks by modeling reconstruction sparsity. *TOMCCAP* 7, Supplement (2011), 31.

Lyndon S. Kennedy and Mor Naaman. 2008. Generating diverse and representative image search results for landmarks. In *WWW*. 297–306.

Lyndon S. Kennedy, Mor Naaman, Shane Ahern, Rahul Nair, and Tye Rattenbury. 2007. How Flickr helps us make sense of the world: Context and content in community-contributed media collections. In *ACM Multimedia*. 631–640.

Jey Han Lau, Karl Grieser, David Newman, and Timothy Baldwin. 2011. Automatic labelling of topic models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume* 1. 1536–1545.

Heng Liu, Tao Mei, Jiebo Luo, Houqiang Li, and Shipeng Li. 2012b. Finding perfect rendezvous on the go: Accurate mobile visual localization and its applications to routing. In *ACM Multimedia*. 9–18.

Jiajun Liu, Zi Huang, Lei Chen, Heng Tao Shen, and Zhixian Yan. 2012a. Discovering areas of interest with geo-tagged images and check-ins. In *ACM Multimedia*. 589–598.

Jiebo Luo, Dhiraj Joshi, Jie Yu, and Andrew C. Gallagher. 2011. Geotagging in multimedia and computer vision - A survey. *Multimedia Tools Appl.* 51, 1 (2011), 187–211.

Mor Naaman, Yee Jiun Song, Andreas Paepcke, and Hector Garcia Molina. 2004. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004)* (June 2004). http://ilpubs.stanford.edu:8090/766/.

Symeon Papadopoulos, Christos Zigkolis, Stefanos Kapiris, Yiannis Kompatsiaris, and Athena Vakali. 2010. ClustTour: City exploration by use of hybrid photo clustering. In *ACM Multimedia*. 1617–1620.

Till Quack, Bastian Leibe, and Luc J. Van Gool. 2008. World-scale mining of objects and events from community photo collections. In *CIVR*. 47–56.

Daniel Ramage, Christopher D. Manning, and Susan Dumais. 2011. Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 457–465.

Tye Rattenbury and Mor Naaman. 2009. Methods for extracting place semantics from Flickr tags. *TWEB* 3, 1 (2009), 1.

Pavel Serdyukov, Vanessa Murdock, and Roelof van Zwol. 2009. Placing Flickr photos on a map. In *SIGIR*. 484–491.

John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma. 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 2 (2009), 210–227.

Felix X. Yu, Rongrong Ji, and Shih-Fu Chang. 2011. Active query sensing for mobile location search. In *ACM Multimedia*. 3–12.

Ying Zhao and George Karypis. 2001. Criterion functions for document clustering: Experiments and analysis. *Machine Learning*.

Yan-Tao Zheng, Zheng-Jun Zha, and Tat-Seng Chua. 2011. Research and applications on georeferenced multimedia: A survey. *Multimedia Tools Appl.* 51, 1 (2011), 77–98.

Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig Adam, Ulrich Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua, and Hartmut Neven. June, 2009. Tour the world: Building a web-scale landmark recognition engine. In *Proceedings of International Conference on Computer Vision and Pattern Recognition*. DOI:http://dx.doi.org/10.1145/1291233.1291448

Received June 2015; accepted September 2016