Multimedia Systems
DOI 10.1007/s00530-014-0431-8

SPECIAL ISSUE PAPER

An incremental probabilistic model for temporal theme analysis of landmarks

Weiqing Min · Bing-Kun Bao · Changsheng Xu

Abstract Social media sites (e.g., Flickr) generate a huge amount of landmark photos with temporal information in the real-world, such as the photos describing the events happening near landmarks, and those showing different seasonal sceneries. Analyzing this temporal information of landmarks can benefit various applications, such as landmark timeline construction and tour recommendation. In this paper, we propose a novel Incremental Spatio-Temporal Theme Model (ISTTM), which can incrementally mine temporal themes that characterize the temporal information of landmarks, by differentiating them from the other three kinds of themes, i.e., general themes shared by most of all landmarks, local themes related to certain landmarks and the background theme including non-informative content. ISTTM works in an online way and is capable of selectively processing the updates of the distributions on different types of themes. Based on the proposed ISTTM, we present a framework, namely Temporal Theme Analysis for Landmarks (TTAL), which enables both periodic theme detection from discovered temporal themes and temporal theme visualization by selecting the relevant photos. We have conducted experiments on a large-scale landmark dataset from Flickr. Qualitative and quantitative evaluation results demonstrate the effectiveness of the ISTTM as well as the TTAL framework.

W. Min · B.-K. Bao · C. Xu (✉)
National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
e-mail: csxu@nlpr.ia.ac.cn

W. Min
e-mail: wqmin@nlpr.ia.ac.cn

B.-K. Bao
e-mail: bingkunbao@gmail.com

Keywords Landmark temporal themes · Online theme model · Periodic themes

1 Introduction

Many significant events occurred around landmarks because of the unique cultural and historical background of these landmarks. For example, the political demonstrations and annual conventions often happen at Trafalgar Square in London (Fig. 1a). Besides the events, some landmarks also show charming views at certain moments, such as "cherry blossom in spring" and "red maple leaves in autumn" at Kiyomizu-dera in Japan (Fig. 1b). These temporal information of landmarks is very useful for landmark history browsing and tour recommendation. As hosting the substantial amount of landmark photos and the associated textual contents, Flickr provides us a good source for mining temporal themes of landmarks.

To handle a huge amount of landmark data from Flickr, the theme mining algorithms should be both scalable and efficient. A common strategy is to adopt the online algorithm. Compared with the offline algorithm, which needs to rebuild the whole model from the scratch at each new timepoint, the online algorithm can process the data stream incrementally and thus is more practical. Therefore, in this paper, we focus on designing the online algorithm for discovering temporal themes of landmarks using the landmark photos with associated textual information from Flickr.

Besides the temporal themes, the photos of all the landmarks generally include the other three kinds of themes, namely the background theme [17], the general theme and local theme [24]. The background theme includes non-discriminative and non-informative content. The general theme indicates that the theme extensively occurs at various

Published online: 27 December 2014

Fig. 1 An illustration of temporal themes from Flickr

landmarks, such as transportation and accommodation. The local theme implies one specific theme only existing at a particular landmark, such as their notable appearances and styles. The probabilistic theme models [5, 14] and their online versions [9, 12] have been successfully applied to discover themes from documents for text mining. However, compared with the existing methods, our task faces the following two challenges:

- Differentiation among multiple kinds of landmark themes. One landmark involves four kinds of themes, namely the background theme, global theme, local theme and temporal theme, which means the temporal themes are intertwined with other three kinds of themes. If we directly applied the existing topic model methods, such as by Probabilistic Latent Semantic Analysis (PLSA) [14] and Latent Dirichlet Allocation (LDA) [5] to mine temporal themes, the discovered temporal themes are probably noisy. Therefore, the first problem is to differentiate among four kinds of themes.
- Update on multiple kinds of landmark themes. Since we design the temporal theme mining algorithm in an online way, our algorithm should be capable of adapting the model parameters to incoming documents. However, the temporal patterns for different kinds of themes are different. For example, the general themes and local themes are time invariant and their theme-word distribution remains fixed when the new data streams arrive. In contrast, the temporal themes evolve with time. Some temporal themes disappear while new temporal themes probably emerge. Their theme-word distributions must be updated to accommodate changes. Therefore, we need to treat differently the update on different kinds of themes.

To address these issues, we propose an online probabilistic theme model, called Incremental Spatio-Temporal Theme Modeling (ISTTM), to mine landmark temporal themes from Flickr in an online way. By decomposing the landmark documents[1] into the background theme, general theme, local theme and temporal theme, ISTTM can learn the corresponding four kinds of the low-dimensional theme spaces, respectively. Furthermore, ISTTM can treat differently the parameter adjustments in the word distribution on different kinds of themes. When new data stream arrives, it can keep the word distribution on general themes and local themes fixed by assigning zero probability to these themes for new words. Meanwhile, it re-estimates the word distribution on temporal themes by folding in new words. As for the background theme, its theme-word distribution is considered to be constant at each new data stream [20]. In addition, different kinds of theme-word distributions from the previous model is re-used for the next update, which largely reduces the time cost.

Based on the proposed ISTTM, we develop a framework, namely Temporal Theme Analysis for Landmarks (TTAL), shown in Fig. 2. The framework contains three components: online theme modeling (i.e., ISTTM), periodic theme identification and theme visualization. Online theme modeling is first to mine and update the temporal theme and other three kinds of themes. Periodic themes are then identified from the mined temporal themes. Based on the learned word distribution on temporal themes, the relevant photos are selected through a regularization scheme for theme visualization. An example of the landmark Trafalgar Square is illustrated in the bottom right of Fig. 2.

The rest of this paper is organized as follows: Sect. 2 surveys related work. The elaborative description of the

[1] Here, the tags, title and descriptions from one photo are concatenated as one document.

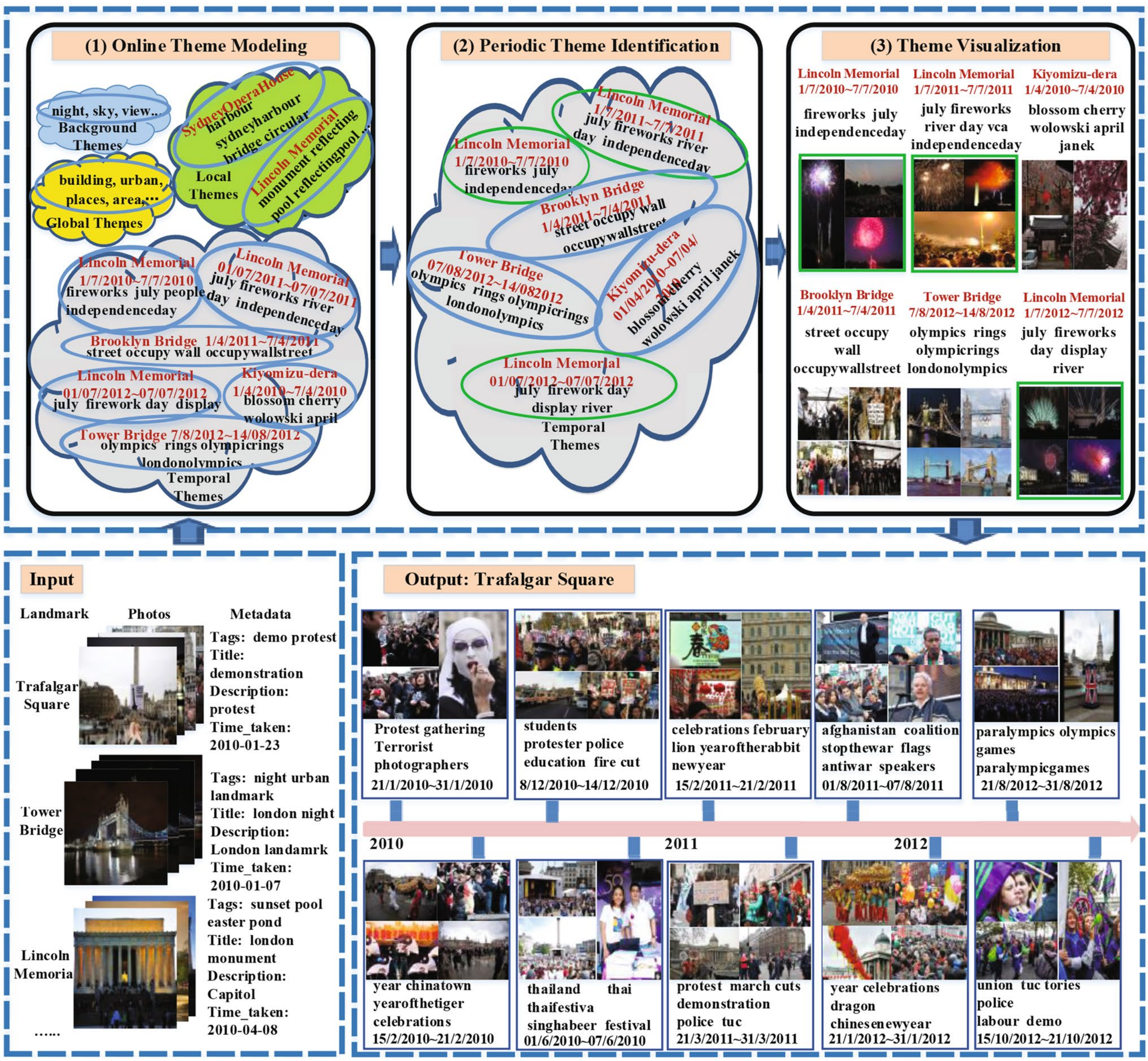

Fig. 2 The framework of ISTTM

framework including online theme modeling, periodic theme identification and theme visualization is presented in Sects. 3, 4 and 5, respectively. Section 6 demonstrates the experimental results while Sect. 7 gives our conclusions and the future work.

2 Related work

Our work is closely related to two research areas: landmark mining and online topic modeling.

2.1 Landmark mining

A lot of work has been proposed to mine landmarks from photo-sharing websites (e.g., Flickr and Panoramio), such as landmark summary [16–18, 26], landmark recognition [19, 28, 29] and landmark retrieval [2, 11, 15, 23]. Kennedy et al. [18] used both context and content information, including tags, geo-tags and visual features to summarize representative views of landmarks from Flickr. Similar to [18], Rudinac et al. [26] also exploited visual features and metadata information for visual summary of geographic areas. Different from the methods in [18, 26], Ji et al. [16] proposed a sparse reconstruction scheme to mine the representative landmark photos. Besides the landmark summary, Li et al. [19] automatically mined landmark photos and then used these relevant photos to train one classifier for landmark classification, while Zheng et al. [29] were dedicated to landmark recognition in large-scale collections. Avrithis et al. [2] proposed a photo clustering scheme to depict different views of one location for further photo retrieval. However, these work only focuses on mining landmarks based on their local characteristics, but ignoring their temporal characteristics.

As discussed before, the landmark involves not only local themes corresponding to the local characteristics, but also temporal themes, such as significant events or distinctive views at certain moments. Little work is dedicated to mining this kind of information. Papadopoulos et al. [25] introduced a clustering method to detect landmarks and events simultaneously. However, the detected events are not relevant to landmarks. One more relevant work [21] employed the bursty detection method to mine the landmark events offline. Different from [21], we resort to the online probabilistic topic model to mine temporal themes in real-time.

2.2 Online topic modeling

The probabilistic topic model, such as PLSA [14] and LDA [5], is to represent a document as a mixture of topics, where each topic is a probability distribution over words. It has been successfully applied to many domains, such as the text [5], the image [4] and the music [13]. A more comprehensive survey of the probabilistic topic model is provided in [3]. In order to apply the topic model to the large-scale data, the online methods are proposed, which are divided into two categories: PLSA-like methods and LDA-like methods. For the PLSA-like online method, the simplest method is just to fold in new documents [14] to the model. The problem of this method is the inability to add new words to the model without retraining. Chou et al. [9] moved one step further and derived the model for the current data stream by adapting the model of the previous data stream for both new documents and new words. This method first folds in new documents into the trained model and then new words. Similar methods are adopted by Gohr et al. [10]. Compared with [9, 10], Brants et al. [6] presented a unified model by simultaneously folding in new documents and new words. These methods all use Maximum likelihood (ML) for parameter estimation of models. In contrast, Chien and Wu [8] proposed another PLSA incremental learning algorithm, which updated the model parameters using the Maximum A Posterior (MAP). Our method is also PLSA-like; however, the difference from the above mentioned methods is that we consider the update problem of multiple kinds of topics instead of single kind. Besides the PLSA-like online methods, the LDA-based online topic methods have also been widely studied, such as [1, 7, 12]. AlSumait et al. [1] proposed an online LDA method, which extended the Gibbs sampling method and utilized it to derive hyper-parameters of the topic-word distribution at the next time slice. The model can capture the emerging topics by adding new words from the new document stream to the vocabulary. Hoffman et al. [12] developed an online variation Bayes algorithm for LDA. However, they assumed that the distribution over words is modeled as a multinomial, which is drawn from a finite Dirichlet distribution, and thus precludes additional words being added over time. Zhai et al. [7] solved this problem by drawing topics from a Dirichlet process rather than the finite Dirichlet distribution. Different from their work, we focus on the online topic modeling based on PLSA rather than LDA.

3 Online theme modeling

In this section, we first describe the Spatio-Temporal Theme Modeling (STTM) [22], which mines multiple kinds of landmark themes in an offline way and then introduce the proposed incremental STTM.

3.1 Spatio-Temporal theme modeling

As discussed in Sect. 1, there are four kinds of themes in the Flickr documents: (1) the background theme, which mainly includes non-discriminative and non-informative content; (2) the general theme, which corresponds to the common ones shared by various landmarks; (3) the local theme, which is related to one certain landmark and characterizes its distinctive features; and (4) the temporal theme, which characterizes some events or distinctive scenic views at specific moments for one certain landmark. The aim of STTM is to mine these four kinds of themes from the Flickr documents.

Given a Flickr document set D, which covers a set of words W, a set of landmarks L and a set of time intervals T, each document d is associated with one landmark $l_d \in L$ and one time interval $t_d \in T$. Based on a set of background themes Z^{bg}, a set of global themes Z^{gl}, a set of local themes Z^{loc} and a set of temporal themes Z^{tl}, each word in document d is generated from one of the aforementioned four kinds of themes (I) a background theme $z \in Z^{bg}$ chosen according to η_D, namely a document set-specific distribution over background themes; (II) a general theme $z \in Z^{gl}$ chosen according to θ_d, namely a document-specific distribution over global themes; (III) a local theme $z \in Z^{loc}$ chosen according to ψ_{l_d}, namely a landmark-specific distribution over local themes or (IV) a temporal theme $z \in Z^{tl}$ chosen according to $\phi_{(l_d,t_d)}$, namely a landmark and time-specific distribution over temporal themes. The variable $x \in \{bg, gl, loc, tl\}$ is sampled from a document-specific distribution over theme types π_d and decides to which type of themes will be sampled. Similar to other topic models, all the distributions are multinomial distributions. The details of the generative process for each document d in D are as follows:

For each word $w_{d,n} \in \{w_{d,1}, w_{d,2}, ..., w_{d,N_d}\}$ (where N_d is the total number of words in d),

1. Draw $x_{d,n} \sim Multi(\pi_d)$
2. If $x_{d,n} = bg$, then draw a background theme $z_{d,n} \sim Multi(\eta_D)$
3. If $x_{d,n} = gl$, then draw a general theme $z_{d,n} \sim Multi(\theta_d)$
4. If $x_{d,n} = loc$, then draw a local theme $z_{d,n} \sim Multi(\psi_{l_d})$
5. If $x_{d,n} = tl$, then draw a temporal theme $z_{d,n} \sim Multi(\phi_{(l_d,t_d)})$
6. Draw $w_{d,n} \sim Multi(\varphi^{x_{d,n}}_{z_{d,n}})$,

where $Multi(\cdot)$ denotes the multinomial distribution. $\varphi^{x_{d,n}}_{z_{d,n}}$ is a multinomial distribution over words specific to the background theme $z_{d,n}(x_{d,n} = bg)$, the general theme $z_{d,n}$ $(x_{d,n} = gl)$, the local theme $z_{d,n}(x_{d,n} = loc)$ or the temporal theme $z_{d,n}(x_{d,n} = tl)$. $\pi_d \triangleq \{p(x|d)\}_{x\in\{bg,gl,loc,tl\}}$ is a multinomial distribution over theme types in document d. $\eta_D \triangleq \{\eta_{D,z}\}_{z\in Z^{bg}}$ is a multinomial distribution over background themes. $\theta_d \triangleq \{\theta_{d,z}\}_{z\in Z^{gl}}$ is a multinomial distribution over global themes. $\psi_{l_d} \triangleq \{\psi_{l_d,z}\}_{z\in Z^{loc}}$ is a multinomial distribution over local themes. $\phi_{(l_d,t_d)} \triangleq \{\phi_{(l_d,t_d),z}\}_{z\in Z^{tl}}$ is a multinomial distribution over temporal themes.

According to the generative process of the document set D, the log-likelihood of D can be written as

$$L(D) = \sum_{d\in D}\sum_{w\in W} n(d,w) \times log\Bigg[p(x=bg|d)\sum_{z\in Z^{bg}}\eta_{D,z}\varphi^{bg}_{z,w} + p(x=gl|d)\sum_{z\in Z^{gl}}\theta_{d,z}\varphi^{gl}_{z,w} + p(x=loc|d)\sum_{z\in Z^{loc}}\psi_{l_d,z}\varphi^{loc}_{z,w} + p(x=tl|d)\sum_{z\in Z^{tl}}\phi_{(l_d,t_d),z}\varphi^{tl}_{z,w}\Bigg] \quad (1)$$

Generally, the number of background themes is one [20]; thus the log-likelihood of D is reformulated as

$$L(D) = \sum_{d\in D}\sum_{w\in W} n(d,w) \times log\Bigg[p(x=bg|d)\varphi^{bg}_{w} + p(x=gl|d)\sum_{z\in Z^{gl}}\theta_{d,z}\varphi^{gl}_{z,w} + p(x=loc|d)\sum_{z\in Z^{loc}}\psi_{l_d,z}\varphi^{loc}_{z,w} + p(x=tl|d)\sum_{z\in Z^{tl}}\phi_{(l_d,t_d),z}\varphi^{tl}_{z,w}\Bigg] \quad (2)$$

Similar to [20], φ^{bg}_w is estimated as

$$\varphi^{bg}_w = \frac{\sum_{d\in D} n(d,w)}{\sum_{w'\in W}\sum_{d\in D} n(d,w')} \quad (3)$$

Other parameters are estimated using the Expectation-Maximum (EM) algorithm [1] and the details can be referred in our previous work [22].

To enable STTM to work in an online way, we design the incremental STTM, namely ISTTM.

3.2 Incremental spatio-temporal theme modeling

Before describing ISTTM, we introduce some necessary notations. For the data stream D_s at the time slice s, the new document $d_{new} \in D_s$ and the set of words from D_s is W_s. The new words from D_s is $w_{new} = W_s \backslash W^a_{s-1}$ and the old words from D_s is $w_{old} \in W_s \cap W^a_{s-1}$, where W^a_{s-1} denotes the accumulated words from all the data streams up to the time slice $s-1$.

In ISTTM, we first initialize the model in the way of STTM at the first data stream D_1. The trained model is used as a prior for the new data stream D_s. Both new documents and new words from D_s are folded into the model. First, new documents d_{new} corresponding to the new time interval t_{new} lead to the re-estimation of $\theta_{d,z}$ and $\phi_{(l,t),z}$. The theme distribution on the landmark $\psi_{l,z}$ is generally time-invariant, thus we keep $\psi_{l,z}$ fixed. Second, we should treat differently the word distribution on different kinds of themes. Specifically, general themes and local themes are time-invariant ones and thus their theme-word distributions generally do not change over time. Therefore, $\varphi^x_{z,w_{new}} = 0, x \in \{gl, loc\}$. Since the new data stream D_s probably introduces new temporal themes, $\varphi^{tl}_{z,w}$ needs to be re-estimated.

In order to re-estimate these parameters at the current data stream D_s, the original theme-word distribution $\varphi^{tl}_{z,w}$ at the time slice $s-1$ is first re-normalized proportionally as follows:

$$\left(\varphi^{tl}_{z,w_{old}}\right)_{s-1} = \frac{\left(\varphi^{tl}_{z,w_{old}}\right)_{s-1}}{\left(\sum_{w_{old}}\varphi^{tl}_{z,w_{old}}\right)_{s-1}} \quad w_{old} \in W_s \cap W^a_{s-1} \quad (4)$$

We then assign initial probabilities $(\varphi^{tl}_{z,w_{new}})_s$ to the new words. In our work, we simply initialize the $(\varphi^{tl}_{z,w_{new}})_s$ to a small ϵ as

$$\left(\varphi^{tl}_{z,w_{new}}\right)_s = \epsilon, \quad w_{new} = W_s \backslash W^a_{s-1} \quad (5)$$

Finally, we re-normalize the existing trained words $(\varphi^{tl}_{z,w_{old}})_{s-1}$ to $(\varphi^{tl}_{z,w_{old}})_s$ as shown in Eq. (6) and the new words for $(\varphi^{tl}_{z,w_{new}})_s$ as in Eq. (7).

$$\left(\varphi^{tl}_{z,w_{old}}\right)_s = \frac{\left(\varphi^{tl}_{z,w_{old}}\right)_{s-1}}{1+\sum_{w_{new}} \varphi^{tl}_{z,w_{new}}} \tag{6}$$

$$\left(\varphi^{tl}_{z,w_{new}}\right)_s = \frac{\varphi^{tl}_{z,w_{new}}}{1+\sum_{w_{new}} \varphi^{tl}_{z,w_{new}}} \tag{7}$$

As for $\theta_{d,z}$ and $\phi_{(l,t),z}$, we randomly initialize them. After initialization, the EM algorithm is applied to the model as follows:

E-Step:

$$P(z,x|d,w) = \begin{cases} \frac{P(x=bg|d)^{(n)}\varphi^{bg}_w}{(P_{d,w})^{(n)}} \\ \frac{P(x=gl|d)^{(n)}(\theta_{d,z})^{(n)}\varphi^{gl}_{z,w}}{(P_{d,w})^{(n)}} z \in Z^{gl} \\ \frac{P(x=loc|d)^{(n)}\psi_{l_d,z}\varphi^{loc}_{z,w}}{(P_{d,w})^{(n)}} z \in Z^{loc} \\ \frac{P(x=tl|d)^{(n)}(\phi_{(l_d,t_d),z})^{(n)}(\varphi^{tl}_{z,w})^{(n)}}{(P_{d,w})^{(n)}} z \in Z^{tl} \end{cases} \tag{8}$$

M-Step:

$$(\theta_{d,z})^{n+1} = \frac{\sum_{w\in W_1\cap W_s} n(d,w)P(z,x=gl|d,w)}{\sum_{z^{â€™}\in Z^x}\sum_{w\in W_1\cap W_s} n(d,w)P(z,x=gl|d,w)}, z\in Z^{gl} \tag{9}$$

$$(\phi_{(l_d,t_d),z})^{n+1} = \frac{\sum_{d|t_d=t,l_d=l}\sum_{w\in d} n(d,w)P(z,x=loc|d,w)}{\sum_{z^{â€™}\in Z^x}\sum_{d|t_d=t,l_d=l}\sum_{w\in d} n(d,w)P(z,x=gl|d,w)}, z\in Z^{tl} \tag{10}$$

$$(\varphi^{tl}_{z,w})^{n+1} = \frac{\sum_{d\in D_s} n(d,w)P(z,x=tl|d,w)}{\sum_{w'\in W_s}\sum_{d\in D_s} n(d,w)P(z,x=tl|d,w)}, z\in Z^{tl} \tag{11}$$

Similarly, φ^{bg}_w is also set as

$$\varphi^{bg}_w = \frac{\sum_{d\in D_s} n(d,w)}{\sum_{w'\in W_s}\sum_{d\in D_s} n(d,w)} \tag{12}$$

Note that we utilize the prior from the previous stream $s-1$ to initialize the distribution $\varphi^{tl}_{z,w}$ at the current data stream s and thus improve the convergence rate.

Through the ISTTM, we get one sequence of STTM models and a set of temporal themes from each STTM model. One advantage of ISTTM is that each temporal theme built at the data stream D_{s-1} transforms naturally to its follow-up theme at the data stream D_s, which generates the theme thread [10] and facilitates the following periodic theme identification.

4 Periodic theme identfication

In order to identify periodic themes, we examine each temporal theme thread. For themes from each temporal theme thread, the similarity between themes is calculated as follows:

$$sim(z_{k,s_i}, z_{k,s_j}) = \frac{\mathbf{w}_{z_{k,s_i}} \cdot \mathbf{w}_{z_{k,s_j}}}{\|\mathbf{w}_{z_{k,s_i}}\|\|\mathbf{w}_{z_{k,s_j}}\|}, \tag{13}$$

where $\mathbf{w}_{z_{k,s_i}}$ and $\mathbf{w}_{z_{k,s_j}}$ denote the word vector of the theme thread k at the data stream s_i and s_j, respectively. Note that since we allow that vocabularies grow over time, the theme-word distributions from each theme thread are made comparable by filling in zero-probabilities for the differing words. For two themes z_{k,s_i}, z_{k,s_j} with high similarity, we find the corresponding temporal and location pair set TL_{s_i} and TL_{s_j}, where $TL_{s_i} = \{(t,l)|z_{k,s_i} = \max_{k=1,\ldots|K|}\psi_{(t,l),z_{k,s_i}}, t\in T, l\in L\}$ and $TL_{s_j} = \{(t,l)|z_{k,s_j} = \max_{k=1,\ldots|K|}\psi_{(t,l),z_{k,s_j}}, t\in T, l\in L\}$. Note that each data stream has the same number of time intervals T. In our experiment, we consider one year as one time slice and simply divide one month into four time intervals; thus $T=48$ for each data stream. If $TL_{s_i}\cap TL_{s_j}\neq\phi$, then z_{k,s_i} and z_{k,s_j} are marked as one periodic theme.

5 Theme visualization

In order to visualize these temporal themes, we try to find representative photos for each of them. Here, we employ a method similar to [27] to calculate the relevance scores of retrieved photos. We define the original photo set $P=\{p_1,...,p_i,...,p_N\}$, where N is the number of photos and each photo p_i is associated with one document d_i. The corresponding relevance score vector $\mathbf{r}=\{r_{p_1},...,r_{p_i},...,r_{p_N}\}$, where r_{p_i} is the relevance score of the photo p_i. Particularly, we use the following regularization scheme to calculate $\mathbf{r}$:

$$\begin{aligned} Q(\mathbf{r}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}\left(\frac{r_{p_i}}{\sqrt{D_{ii}}} - \frac{r_{p_j}}{\sqrt{D_{jj}}}\right)^2 \\ &+ \lambda\sum_{i=1}^{N}(r_{p_i} - sim(z_q,d_i))^2 \\ \mathbf{r} &= \arg\min(Q(\mathbf{r})), \end{aligned} \tag{14}$$

where λ is the regularization coefficient. W_{ij} is the visual similarity between two photos p_i and p_j. $D_{ii}=\sum_{j=1}^{N} W_{i,j}$. $sim(z_q,d_i)$ is the similarity between the query theme z_q and the document d_i. W_{ij} can be computed based on a Gaussian kernel function with a radius parameter σ, i.e.,

$$W_{ij} = exp\left(\frac{\|\mathbf{v}_{p_i} - \mathbf{v}_{p_j}\|^2}{\sigma^2}\right), \tag{15}$$

Table 1 The statistics of landmarks

#Landmark	#Landmark
Arc De triomphe	Big Ben
Brookly bridge	Buckingham palace
Eiffel tower	Washington monument
Forbidden city	Golden gate bridge
Great wall	Kiyomizu-dera
London eye	Lincoln memorial
Statue of liberty	Notre dame
Summer palace	Sydney opera house
Tokyo tower	Tower bridge
Trafalgar square	White house

Table 2 Number of photos per year

♯Year	2010	2011	2012
♯Photo	134,325	140,618	128,727

where $\mathbf{v}_{p_i}$ denotes the visual feature vector of photo p_i. $sim(z_q, d_i)$ can be computed as follows:

$$sim(z_q, d_i) = \frac{\mathbf{w}_{z_q} \cdot \mathbf{w}_{d_i}}{\|\mathbf{w}_{z_q}\| \|\mathbf{w}_{d_i}\|}, \tag{16}$$

where $\mathbf{w}_{z_q}$ and $\mathbf{w}_{d_i}$ denote the word vector of the theme z_q and document d_i, respectively.

We use the iterative method [27] to solve $\mathbf{r}$ and then select photos with the high relevance score to represent the temporal theme.

6 Evaluation

6.1 DataSet

The experiment is conducted on 20 well-known landmarks, listed in Table 1. The photos together with their associated tags, the title, descriptions and the taken-time are crawled from Flickr by searching the names of the landmarks and constraining the time-taken date from Jan 01, 2010 to Dec 31, 2012. Since the landmark names and relevant city names are meaningless in mining themes, we first removed them. Then the text is cleaned by removing the stopwords and camera-related words such as “Cannon” and “35 mm”. The obtained dataset consists of 403,670 photos and 22,514 unique words. Without loss of generality, we partitioned the data into 3 slices, and each slice corresponds to one year. The number of photos per year is shown in Table 2.

For each photo, we extract 809-D visual features [30], including 81-D color moment features, 37-D edge histogram features, 120-D wavelet texture features, 59-D LBP features and 512-D GIST features. For the parameter setting, the number of landmarks $L = 20$. For each time slice, we further divide it into 48 time intervals, where each month is simply divided into four time intervals, thus $T = 48$. The number of background themes, general themes, local themes and temporal themes are set empirically to 1, 50, 50 and 150, respectively, and they are fixed across all the streams. We trained all the models using the EM algorithm with exactly the same stopping criteria: the average change in the expected log likelihood is <1 on one server with an Intel 4-Core 2.27 GHZ processor and 32 G RAM.

6.2 Qualitative evaluation

6.2.1 Case study on discovered themes

We demonstrate the effectiveness of ISTTM by providing some example themes. These examples are shown in both Tables 3 and 4, where “#j” denotes the j-th theme index. Each theme is represented by 10 top-ranked words, sorted by their respective theme-word distributions φ_w^{bg} and $\varphi_{z,w}^x, x \in \{gl, loc, tl\}$. Meanwhile, we manually label discovered temporal themes according to $\phi_{(l,t)}$ and local themes based on ψ_l. As shown in Table 3, we can see that some temporal themes denote the distinctive views at certain moments, such as #15 Cherry Blossom at kiyomizu-dera. Others characterize some important activities (e.g., #78 Fleet Week at Golden Gate Bridge and #101 Olympic Games at Tower Bridge) or events (e.g., #66 Occupy Wall Street at Brooklyn Bridge). Table 4 shows some examples of background theme, general themes and local themes. We can see that the theme words from the background theme are non-informative. Global themes tend to appear in many landmarks, such as #21 Transportation. In contrast, local themes are mainly related to the styles of certain landmarks. For example, Local theme #30 indicates the style “bay” of Golden Gate Bridge while Local theme #9 characterizes the style “harbor” of Sydney Opera House.

6.2.2 Case study on detected periodic themes

We examine each theme thread to identify the periodic theme. Since each data stream corresponds to one year, the periodic interval is one year. There are 21 theme threads, where at least two themes from each theme thread are meaningful. Through our periodic theme identification method, 10 periodic themes are identified and some examples are shown in Table 5. We can see that the theme “Independence Day” from America is an annual event, which happens in fourth of July each year at White House and Lincoln Memorial. “Cherry Blossom” is an important view in 01/04–07/04 each year at Kiyomizu-dera.

Table 3 Examples of temporal themes

Temporal ♯69 kiyomizu-dera 01/04/2010 ∼ 07/04/2010 Cherry blossom	Temporal♯78 Golden gate bridge 01/10/2010 ∼ 07/10/2010 Fleet week	Temporal ♯66 Brooklyn bridge 01/04/2011 ∼ 07/04/2011 Occupy WallStreet	Temporal ♯101 Tower bridge 07/08/2012 ∼ 14/08/2012 Olympic games
Blossom	Fleet	Street	Olympics
0.10198	0.04988	0.060124	0.20724
Cherry	Alcatraz	Occupy	Olympic
0.061244	0.04974	0.049851	0.11943
Blossoms	Blueangels	Wall	Rings
0.032415	0.043113	0.045948	0.0606
Wolowski	Embarcadero	Occupywallstreet	Olympicrings
0.030428	0.037701	0.045125	0.03296
April	Blue	September	Games
0.025449	0.03641	0.041864	0.031622
Janek	Bay	Nypd	Londonolympics
0.020769	0.036206	0.037644	0.027342
Pink	Fleetweek	Ows	Olympicgames
0.018254	0.034576	0.025723	0.026696
Cherryblossomfestival	Coittower	Bridges	Ceremony
0.016751	0.029835	0.025201	0.020415
Cherryblossoms	Chrissyfield	October	Wenlock
0.015096	0.023515	0.017748	0.016768
Flowering	Airshow	Wallstreet	Paralympic
0.013996	0.022564	0.016247	0.011217

6.3 Quantitative evaluation

6.3.1 Experiment on online theme modeling

To evaluate the performance of ISTTM quantitatively, we consider the following three baselines for comparison:

- Naive_PLSA: This method uses a new random initial setting to re-estimate all the parameters of STTM at each time slice. It only differentiates between background themes and non-background themes.
- Online_PLSA [6]: similar to ISTTM, but only differentiates between background themes and non-background themes.
- Naive_STTM: similar to Naive_PLSA, but differentiates among four kinds of themes.

- We first utilize the perplexity [14] as our goodness-of-fit-measure to assess the generalization performance of our model. The smaller the value of perplexity, the better the performance. For the test set D_{test}, the perplexity is defined as:

$$\text{perplexity}(D_{\text{test}}) = \exp\left(-\frac{\sum_d \sum_w n(d,w) log P_{d,w}}{\sum_d \sum_w n(d,w)}\right), \tag{17}$$

where $n(d,w)$ denotes the counts on the test data D_{test} and

$$\begin{aligned} P_{d,w} = P(x=bg|d)\varphi_w^{bg} + P(x=gl|d)\sum_{z\in Z^{gl}}\theta_{d,z}\varphi_{z,w}^{gl} \\ + P(x=loc|d)\sum_{z\in Z^{loc}}\psi_{l_d,z}\varphi_{z,w}^{loc} \\ + P(x=tl|d)\sum_{z\in Z^{tl}}\phi_{(l_d,t_d),z}\varphi_{z,w}^{tl} \end{aligned} \tag{18}$$

For the current data stream, we hold out 80 % of the data for training and 20 % of the data for test. In the training stage, naive method performs a random initialization while online method utilizes the theme-word distribution parameters from the previous data stream for initialization. In order to compute perplexity, we obtain the topic-word distributions φ_w^{bg} and $\varphi_{z,w}^x, x \in \{gl, loc, tl\}$ in the training stage. In the testing stage, we fix $\varphi_{z,w}^x$ and re-estimate $\theta_d, \psi_l, \phi_{(l,t)}$ and $p(x|d)$, where $x \in \{bg, gl, loc, tl\}$, and then compute $P_{d,w}$. Each individual experiment is repeated 10 times and we present the average value of the hold-out perplexities.

The result is shown in Table 6. We can see that (1) the online method is always better than the naive method. This is because online methods considers the prior knowledge from previous data stream and thus have better generalization ability; (2) the STTM method is always better than the PLSA method. This indicates that the STTM model

Table 4 Examples of background themes, general themes and local themes

Background♯1	General♯4	General♯21	Local♯9 Sydeny opera house	Local♯30 Golden gate bridge
Sky	Morning	Cars	Harbour	Fog
0.0050265	0.046831	0.040378	0.15081	0.062074
Night	Riverthames	October	Sydneyharbour	Beach
0.004769	0.033011	0.013263	0.040352	0.033625
View	Night	Bus	Harbourbridge	Water
0.0038357	0.02881	0.013202	0.038824	0.028499
Tree	Sunny	Paul	Bridge	Ocean
0.0036178	0.025875	0.01275	0.028961	0.020438
National	Fort	Spectacular	Northern	Skyline
0.0035946	0.025162	0.012445	0.026585	0.019021
Building	Sunrise	Start	Harbourbridge	Clouds
0.0033589	0.023811	0.012191	0.018112	0.014198
Day	Day	Telephone	Circular	Boats
0.003315	0.018474	0.01152	0.017697	0.013566
People	Shots	People	Ferry	Harbour
0.0031897	0.015303	0.010291	0.016901	0.012384
Architecture	Grey	Red	Cruise	Sunset
0.0030551	0.012018	0.010235	0.016382	0.011373
Bridge	January	Driving	Yachts	Harbor
0.0029571	0.01147	0.010109	0.015984	0.01133

considers multiple kinds of themes to better learn predictive models from the documents. Our method learns the model by differentiating multiple kinds of themes in an online way and thus achieves the best performance among all the methods.

Besides the perplexity, the execution time required on training models is used to evaluate the efficiency of our online theme model. Table 7 summarizes the comparison results between our method and other three baselines at each time slice. PLSA-like methods and STTM-like methods adopt the same initialization at the fist time slice, respectively. For the STTM-like method, the computation time of our method is less than Naive_STTM. Compared with the Naive_STTM method, which uses a new random initial setting to re-estimate all the parameters, our method utilizes the previous estimated parameters as the initialization and thus reduces largely the time. Since our method considers the update on more parameters, our method uses more time than PLSA-like methods. However, the perplexity of our method is the best. Therefore, our method can better balance the effectiveness and efficiency.

6.3.2 Experiment on theme visualization

Since our goal is to analyze the landmarks mainly using discovered temporal themes, we first give an evaluation on the number of semantically meaningful temporal themes and then evaluate the semantic consistency for each semantically meaningful theme based on the retrieved photos according to Eq. 14.

We ask 10 users to label semantically meaningful temporal themes with the help of the retrieved photos. They can also consult both Flickr dataset and external resources to help make judgment. In order to compare the performance between STTM-like methods and PLSA-like methods, we select the Online_PLSA baseline and utilize the number of discovered temporal themes as the metric. The statistics are shown in Table 8. We can see that our method performs better than Online_PLSA method. This is because our method can make the differentiation between the temporal themes and other two kinds of themes: global themes and local themes. Therefore, our method properly filters out these words using general and local themes, which do not belong to temporal themes.

Meanwhile, we ask them to label the top 10 retrieved relevant photos for each meaningful temporal theme; MAP@10 is used to evaluate Online_PLSA and our method. Table 9 shows the statistics at different time slices. Again, our method achieves the better performance than Online_PLSA results.

With the discovered temporal themes, represented by tags and photos, we organize these information and present the summary in the way of timeline. The timeline of two example landmarks Trafalgar Square and Washington Memorial are shown in Figs. 3 and 4, respectively. From

Table 5 Examples of periodic themes

♯Year	2010	2011	2012
♯12 Landmark: Lincoln memorial White house	Fireworks	July	July
Time: 01/07–07/07	0.19254	0.14219	0.27004
Temporal theme: Independence day	July	Fireworks	Fireworks
	0.11396	0.13013	0.25093
	Independenceday	River	Firework
	0.085132	0.056072	0.033864
	People	Day	Day
	0.068934	0.054896	0.033045
	Lanape	Vca	Display
	0.06554	0.034785	0.028434
	Washingtonmemorial	Independenceday	River
	0.062551	0.029865	0.022619
	Capitolhill	Independence	Fire
	0.034836	0.028821	0.022612
	Idependence	Ptomac	Nght
	0.022132	0.018055	0.014485
	Capitolbuilding	Night	Independence
	0.013757	0.017815	0.010237
	Washingtonmemorialfireworks	Colors	Exploratoriu
	0.012893	0.016064	0.0089089
♯15 Landmark: Kiyomizu-dera	Blossom	Cherry	Spring
Time: 01/04–07/04	0.10198	0.065808	0.14562
Temporal theme: Cherry blossom	Cherry	Blossom	Cherry
	0.061244	0.057621	0.071316
	Blossoms	Festival	Sakura
	0.032415	0.039937	0.0705
	Wolowski	Cherryblossoms	Blossom
	0.030428	0.029478	0.043135
	April	National	Blossoms
	0.025449	0.029136	0.04284
	Janek	Blossoms	Grounds
	0.020769	0.028957	0.039217
	Pink	Spring	Recent
	0.018254	0.026951	0.036513
	Cherryblossomfestival	Tidal	Cherryblossoms
	0.016751	0.02266	0.035013
	Cherryblossoms	Tidalbasin	Trees
	0.015096	0.021863	0.028461
	Flowering	Bloom	Festival
	0.013996	0.021224	0.024585
♯144 Landmark: Sydney Opera House	Vivid	Vivid	Vivid
Time: 01/06 $\sim$ 14/06	0.1317	0.12416	0.090264
Temporal theme: Vivid Sydney	Festival	Festival	Festival
	0.032968	0.063627	0.02789
	Vividsydney	Light	Light
	0.030583	0.056038	0.026594
	Sails	Lights	June
	0.026555	0.04442	0.023565

Table 5 continued

♯Year	2010	2011	2012
	Light	Sails	Vividsydney
	0.018779	0.039644	0.02224
	National	Night	Mozart
	0.016438	0.029567	0.01993
	Lighting	Circularquay	Ideas
	0.014657	0.028637	0.019608
	June	Lighting	Creative
	0.013973	0.027505	0.015562
	Macquarie	Vividsydney	Sails
	0.013232	0.023396	0.015333
	Laurie	June	Piano
	0.012333	0.019402	0.014913
	Katoomba	Colour	Colour
	0.01226	0.015694	0.014123

Table 6 Perplexity comparison on different methods

Method	2010	2011	2012
Naive_PLSA	1,103.63	3,791.35	1,473.76
Online_PLSA	Nil	1,433.89	1,297.85
Naive_STTM	187.91	423.38	371.11
Our method	Nil	388.29	336.55

Table 7 Execution time (in seconds) of all methods

Method	2010	2011	2012	Average
Naive_PLSA	20,553.84	45,105.63	43,401.97	36,353.81
Online_PLSA	20,553.84	31,586.85	28,871.80	27,004.16
Naive_STTM	59,373.6	65,995.24	56,309.14	60,559.33
Our method	59,373.6	40,953.47	32,593.37	44,306.81

Table 8 Statistics on discovered temporal themes between Online_PLSA and our method

Method	2010	2011	2012
Online_PLSA	25	24	16
Our method	36	38	37

Table 9 MAP@10 between Online_PLSA and our method

Method	2010	2011	2012	Average
Online_PLSA	0.8081	0.8155	0.7295	0.7844
Our method	0.8776	0.8673	0.8311	0.8587

such summary, users can easily know what happened near landmarks at some important moments.

7 Conclusions

In this work, we propose a novel Incremental Spatio-Temporal Theme Model (ISTTM) to discover temporal themes in an online way. ISTTM is capable of differentiating temporal themes from the other three kinds of themes, i.e., general themes, local themes and the background theme, and treating differently the update of the distribution on different types of themes. The Temporal Theme Analysis for Landmark (TTAL) framework is presented based on the proposed ISTTM. It first learns the temporal themes and then identifies the periodic themes from the mined temporal themes. Finally, the themes are represented by relevant tags and the selected photos and these visualized themes

Fig. 3 The timeline of trafalgar square

Fig. 4 The timeline of Washington memorial

are organized together to construct the landmark timeline. Future research issues include the identification of emerging themes and the detection of correlated themes from different landmarks.

Acknowledgments This work is supported in part by National Program on Key Basic Research Project (973 Program, Project No. 2012CB316304), Beijing Natural Science Foundation (Grant No. 4131004) and National Natural Science Foundation of China (Grant No. 61225009, No. 61201374, No. 61432019).

References

1. AlSumait, L., Barbará, D., Domeniconi, C.: On-line lda: adaptive topic models for mining text streams with applications to topic detection and tracking. In: Data mining, 2008. ICDM'08. Eighth IEEE international conference on, pp. 3–12. IEEE (2008)
2. Avrithis, Y., Kalantidis, Y., Tolias, G., Spyrou, E.: Retrieving landmark and non-landmark images from community photo collections. In: Proceedings of the international conference on multimedia, pp. 153–162. ACM (2010)
3. Blei, D., Carin, L., Dunson, D.: Probabilistic topic models. Signal Process. Mag. IEEE **27**(6), 55–65 (2010)
4. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp. 127–134. ACM (2003)
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
6. Brants, T.H., Tsochantaridis, I., Hofmann, T., Chen, F.R.: Methods, apparatus, and program products for performing incremental probabilistic latent semantic analysis. US Patent 7,529,765 (2009)
7. Canini, K.R., Shi, L., Griffiths, T.L.: Online inference of topics with latent dirichlet allocation. In: International conference on artificial intelligence and statistics, pp. 65–72 (2009)
8. Chien, J.T., Wu, M.S.: Adaptive bayesian latent semantic analysis. Audio Speech Lang. Process. IEEE Trans. **16**(1), 198–207 (2008)
9. Chou, T.C., Chen, M.C.: Using incremental plsi for threshold-resilient online event analysis. Knowl. Data Eng. IEEE Trans. **20**(3), 289–299 (2008)
10. Gohr, A., Hinneburg, A., Schult, R., Spiliopoulou, M.: Topic evolution in a stream of documents. In: SDM, vol. 9, pp. 859–872. SIAM (2009)
11. Guan, T., He, Y., Gao, J., Yang, J., Yu, J.: On-device mobile visual location recognition by integrating vision and inertial sensors (2013)
12. Hoffman, M., Bach, F.R., Blei, D.M.: Online learning for latent dirichlet allocation. In: advances in neural information processing systems, pp. 856–864 (2010)
13. Hoffman, M.D., Blei, D.M., Cook, P.R.: Content-based musical similarity computation using the hierarchical dirichlet process. In: ISMIR, pp. 349–354 (2008)
14. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Mach. Learn. **42**(1–2), 177–196 (2001)
15. Ji, R., Duan, L.Y., Chen, J., Yao, H., Yuan, J., Rui, Y., Gao, W.: Location discriminative vocabulary coding for mobile landmark search. Int. J. Comput. Vis. **96**(3), 290–314 (2012)
16. Ji, R., Gao, Y., Zhong, B., Yao, H., Tian, Q.: Mining flickr landmarks by modeling reconstruction sparsity. ACM Trans. Multimed. Comput. Commun. Appl. (TOMCCAP) **7**(1), 31 (2011)
17. Ji, R., Yao, H., Liu, W., Sun, X., Tian, Q.: Task-dependent visual-codebook compression. Image Process. IEEE Trans. **21**(4), 2282–2293 (2012)
18. Kennedy, L., Naaman, M.: Generating diverse and representative image search results for landmarks. In: Proceeding of the 17th international conference on World Wide Web, pp. 297–306. ACM (2008)
19. Li, Y., Crandall, D., Huttenlocher, D.: Landmark classification in large-scale image collections. In: Computer vision, 2009 IEEE 12th international conference on, pp. 1957–1964. IEEE (2009)
20. Mei, Q., Liu, C., Su, H., Zhai, C.: A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In: Proceedings of the 15th international conference on World Wide Web, pp. 533–542. ACM (2006)
21. Min, W., Bao, B.K., Xu, C.: What happened near big ben: event-driven landmark mining from flickr. In: Advances in multimedia information processing-PCM 2012, pp. 769–778. Springer (2012)
22. Min, W., Bao, B.K., Xu, C.: Multi-modal spatio-temporal theme modeling for landmark analysis. IEEE multimedia p. 1 (2014)
23. Min, W., Xu, C., Xu, M., Xiao, X., Bao, B.: Mobile landmark search with 3d models. IEEE Trans. Multimed. **16**(3), 623–636 (2014)
24. Pang, Y., Hao, Q., Yuan, Y., Hu, T., Cai, R., Zhang, L.: Summarizing tourist destinations by mining user-generated travelogues and photos. Comput. Vis. Image Underst. **115**(3), 352–363 (2011)
25. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based landmark and event detection on tagged photo collections. Multimed. IEEE **99**, 1–1 (2010)
26. Rudinac, S., Hanjalic, A., Larson, M.: Generating visual summaries of geographic areas using community contributed images (2013)
27. Wang, M., Yang, K., Hua, X., Zhang, H.: Towards a relevant and diverse search of social images. Multimed. IEEE Trans. **12**(8), 829–842 (2010)
28. Xiao, X., Xu, C., Wang, J., Xu, M.: Enhanced 3-d modeling for landmark image classification. Multimed. IEEE Trans. **14**(4), 1246–1258 (2012)
29. Zheng, Y., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T., Neven, H.: Tour the world: building a web-scale landmark recognition engine. In: Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on, pp. 1085–1092. IEEE (2009)
30. Zhu, J., Hoi, S., Lyu, M., Yan, S.: Near-duplicate keyframe retrieval by nonrigid image matching. In: Proceedings of the 16th ACM international conference on multimedia, pp. 41–50. ACM (2008)