W2Go: A Travel Guidance System by Automatic Landmark Ranking

Yue Gao *, Jinhui Tang †, Richang Hong ‡, Qionghai Dai †, Tat-Seng Chua †, Ramesh Jain §
† Tsinghua National Lab for Information Science and Technology, Dept. of Automation, Tsinghua University
‡ School of Computing, National University of Singapore
§ Bren School of Information and Computer Sciences, University of California, Irvine
{gaoyue08@mails., daiqionghai@}tsinghua.edu.cn
{tangjh, hongrc, chuats}@comp.nus.edu.sg, jain@ics.uci.edu

ABSTRACT
In this paper, we present a travel guidance system W2Go (Where to Go), which can automatically recognize and rank the landmarks for travellers. In this system, a novel Automatic Landmark Ranking (ALR) method is proposed by utilizing the tag and geo-tag information of photos in Flickr and user knowledge from Yahoo Travel Guide. ALR selects the popular tourist attractions (landmarks) based on not only the subjective opinion of the travel editors as is currently done on sites like WikiTravel and Yahoo Travel Guide, but also the ranking derived from popularity among tourists. Our approach utilizes geo-tag information to locate the positions of the tag-indicated places, and computes the probability of a tag being a landmark/site name. For potential landmarks, impact factors are calculated from the frequency of tags, user numbers in Flickr, and user knowledge in Yahoo Travel Guide. These tags are then ranked based on the impact factors. Several representative views for popular landmarks are generated from the crawled images with geo-tags to describe and present them in context of information derived from several relevant reference sources. The experimental comparisons to the other systems are conducted on eight famous cities over the world. User-based evaluation demonstrates the effectiveness of the proposed ALR method and the W2Go system.

Categories and Subject Descriptors
J.0 [Computer Applications]: General; H.1.2 [Models and Principles]: User/Machine Systems Human factors

General Terms
Algorithm, Experimentation

Keywords
Geo-tag, Landmark Ranking, Tag Analysis, Travel Guide

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

1. INTRODUCTION
With the advent of Web 2.0 technology, there is an explosion of community-contributed multimedia content available online. As we know, the intrinsic attribute of Web 2.0 [7] is to facilitate interactive information sharing, interoperability and collaboration on the internet. By virtue of the attribute, many travel recommendation systems emerge as assistants for tourists all over the world. They usually employ the user-contributed content as the information resources and allow users to edit handily. Two systems that are more popular with extensive user accesses are: WikiTravel [23] and Yahoo Travel Guide [1]. They are supported by Internet Brands and Yahoo respectively. These two systems are to provide elaborate descriptions of places of interests at any level of geographic specificity, from continents to districts of a city.

However, one key question is whether such user-contributed content really credible and favorable for a traveller? We identify three key limitations of such travel recommendation systems. First, as the articles in these systems are manually edited and uploaded by travel editors and some users, the choice of contents and ranking of attractions are highly dependent on the subjective opinion of the travel editors and voting of a small number of users. Hence it is unable to cater to the diverse needs of wide variety of users. Second, there are few entries of locations in such travel recommendations and users mostly cannot obtain desired information from these guides. Third, the maintenance of such systems is labor-intensive and the content cannot be updated automatically. Because of these problems, the contents of travel recommendation systems tend to be static, incomplete, and often subjective. To this end, a system which can automatically produce location based landmark suggestion is desired.

As we know, media-sharing websites/systems, such as Flickr [5], YouTube [25], and Panoramio [16] offer rich dynamic media information including photos, videos, associated tags and other context metadata. These multimedia resources
are triggering a lot of innovative research topics [26, 27] How to leverage these free information, such as the photos and their context information in Flickr, to facilitate travel guidance? Can we utilize the information from Flickr to suggest a complete list of landmarks? Zheng et al. [27] proposed to harvest landmarks from the vast amount of the multimedia content on the web, while how to target specific landmarks to personalized needs of travelers has not been addressed. We argue that the web 2.0 content in Flickr can be more thoroughly explored to support landmark suggestion on account of its advantages of that on the one hand, the increasing content in Flickr is in favor of better performance, while on the other hand, to employ such content can reduce human labor a great deal.

Recently, there is an emerging tendency in public photo-sharing community: geo-tagging. Geo-tagged photos provide location information, which is mainly indicated by latitude and longitude. Flickr launched its geo-tagging interface in 2006, and there are now over 90,000,000 geo-tagged photos in Flickr as reported by Flickr itself. These increasing geo-tagged photos bring more opportunities to semantic analysis and other derived applications. Up to now, geographical information has been used in a variety of applications, including finding the relationship between word concepts and geographical locations [24], extracting events and place semantics [17], and generating representative tags and views for location [8]. Albeit that, geo-tags have potential to mine more semantics from both geographical location and tag information.

Motivated by the vast quantity of photos with semantic and geographical tags available on public media-sharing sites, as well as the detailed descriptions of many attractions in travel recommendation systems, this work presents an automatic landmark suggestion system, named W2Go. W2Go dynamically suggests travel attractions for one location. In the system, both of the two kinds of resources, Flickr and Yahoo Travel Guide, are utilized to facilitate landmark ranking. When user inputs a location. The automatically generated landmarks will contain both the landmarks and their representative views, by taking into account users’ interest as well. The overview of the system is shown in Figure 1. Here, we propose an automatic landmark ranking (ALR) algorithm for W2Go where tag/geo-tag information, metadata of photos in Flickr and user knowledge in Yahoo Travel Guide are incorporated into the analysis. ALR consists of two procedures: candidate landmark selection, landmark ranking and representative views generation. The first step selects popular candidate landmarks from Flickr and Yahoo Travel Guide by taking into consideration the users’ interests. The second step evaluates landmark tags by using a tag-driven algorithm based on geographical distribution of tags, where both geotags and tags are employed. It then performs geographical tag filtering to remove repetitive tags. The final landmark ranking is depended on the impact factors generated from ALR and Yahoo Travel Guide. Then representative views are generated based on the photos from Flickr. Based on the techniques involved, W2Go is able to avoid the subjective bias from travel editors as well as voting of a few travellers. Furthermore, it can deal with any location name irrespective of whether it has been included in existing travel recommendation systems, and W2Go is totally automatic.

The main contributions are as follows.

- A user-friendly location-based travel recommendation, W2Go. It is able to prevent the subjective influence from travel editors as well as voting of a few travellers. Moreover, W2Go can identify and rank landmarks for any location specified by travellers, not just popular site.
- An automatically dynamic landmark recognition and ranking algorithm (ALR), which simultaneously utilizes the semantic and geographical tag information of photos in media-sharing sites and the user knowledge from existing travel recommendation systems.

The remainder of the paper is organized as follows. Related works are reviewed in Section 2. Section 3 introduces the W2Go system architecture and ALR. Elaborative description of candidate landmark selection, landmark ranking and representative view generation are presented in Section 4 and Section 5, respectively. Section 6 demonstrates the experimental results and user based evaluation. Followed by conclusions and further work discussions in Section 7.

2. RELATED WORK

Related works is addressed from two aspects: existing travel assistant systems, and landmark recognition and suggestion using tags and geo-tags. The following discusses the strengths and limitations of these approaches.

2.1 Travel assistant systems

Two of the most famous travel assistant systems are WikiTravel and Yahoo Travel Guide. WikiTravel is a worldwide travel guide, and was founded in July of 2003 [23]. WikiTravel is the most effective travel recommendation system, and it is still being modified by WikiTravellers. For each destination, the articles in WikiTravel generally include all or parts of the following information: history, climate, landmarks, work information, shopping information, food, and how to get there. The interface of WikiTravel is shown in Figure 2.

It is noted that little human interaction is permitted in WikiTravel. There are only static articles in WikiTravel, and only when WikiTravellers have added more information to one article, users can find more new information there. Similarly, only when WikiTravellers write an article about one destination, can users find information about this destination. Thus users may not easily focus on what they are interested in these destinations that have not been written yet. Though human editing landmark suggestion is accurate, it is not convenient when no information has been added into WikiTravel about that place.

The other main travel guide system is Yahoo Travel Guide [1]. Yahoo Travel Guide provides an area based recommendation service. In each country, several main cities are listed. For example, there is a total of 10 cities listed for China, and they are Shanghai, Hongkong, Beijing, Guangzhou, etc. For each city, Yahoo Travel Guide provides information on key attractions where a series of ranked landmarks are shown to users, associated with comments by users. The main interface for each city is given in Figure 3, where the region marked by the red circle is the “Attractions” option. In this option, several ranked attractions/landmarks are provided to users. This ranking is mainly based on the comments/reviews and scores provided by users. It is noted that there are some repetitive attractions. For example, in Beijing, “Great Wall”, a general concept, is listed in the top...
place, while the ninth, the sixteenth, and the seventeenth recommended attractions are three special great walls (Mu Tian Yu, Badaling and Si Ma Tai). Another problem is that only several top ranked attractions are given with plenty of reviews and scores, and others (even in the top 10 attractions) are with few reviews and scores (e.g., no more than 10), and hence their ranking may not be representative.

Also, one major limitation of Yahoo Travel Guide is similar with WikiTravel: it is highly dependent on the subjective opinion of some travel editors, and only few cities are available (Before September 2009 there are fewer cities in Yahoo Travel Guide).

2.2 Landmark recognition and suggestion using tags and geo-tags

Media-sharing websites, such as Flickr [5], YouTube [25], and Panoramio [16], have become increasingly popular these days. Thanks to the rapid development of camera technology, most photos are now with geo-tag information. These media-sharing websites can provide both tags and geo-tags. These algorithms mainly utilize images and tags from Flickr or other image-sharing websites. Most existing landmark recognition algorithms are area-driven. Area-driven methods first focus on a small area for the location, and then recognize the landmarks for that area. [3] and [8] first grouped images into clusters using K-means by geographic information (longitude and latitude). Then in each major cluster (a smaller area as compared with the whole location), images with tags and geo-tags, where SIFT signature is employed as image feature and L1 distance is used. Then it treats any cluster with more than five images as a candidate landmark. Further landmark name refinement is done with the help of geo-referenced Wikipedia. It is noted that this method need high computation cost because of the image matching and clustering procedures. In [2], given a city name, the system first crawls all images with that tag of the city name, and classifies all photos into landmark or non-landmark categories using SVM. For each landmark image group, TF-IDF is again employed to rank tags, and the top-k photos with popular landmark tags are presented to users. The algorithm proposed in [9] estimates landmark locations from geo-referenced images. In this work, their method estimates the camera poses and locations of landmarks using camera images.
sist of the uploader's name (ID). Given Data\{I_1, I_2, ...I_m\} with their associated metadata, our goal can be formulated as follows:

How to make use of Data, tag, geo, and ID from Flickr, as well as user knowledge in Yahoo Travel Guide to dynamically generate travel recommendations for user specified locations?

We take Beijing as an example. If the crawled photos and their associated metadata are available, an ideal result is a ranked list of landmarks for Beijing, such as \{Forbidden City, Summer Palace, Great Wall, Tiananmen, Temple of Heaven\} and so on.

3.2 System architecture

Figure 3 illustrates the flowchart of the proposed system, which consists of two main stages: data crawling and automatic landmark ranking (ALR). Given a location name, first the photos and corresponding tags/geo-tags/metadata are crawled from Flickr, and the associated user knowledge is obtained from Yahoo Travel Guide. ALR consists of two main procedures: candidate landmark selection, landmark ranking and representative views generation. In ALR, first the candidate popular tags are selected from Flickr, and the geo-tag information is utilized to filter out noisy tags. A geographical tag refinement procedure is conducted to further remove repetitive landmark tags. The user knowledge from Yahoo Travel Guide is utilized to refine the results from Flickr. In the landmark ranking stage, both the statistics information from Flickr and user knowledge in Yahoo Travel Guide are taken into account, and candidate landmarks are ranked based on impact factor scores. Additionally, the corresponding article in Wikipedia (if available) is provided to users.

4. CANDIDATE LANDMARK SELECTION

In general, a landmark can be considered as a place where most tourists prefer to go for sight-seeing. Based on these properties, we make the following assumptions.

- The more popular a landmark is, the larger the number of tourists and hence the number of uploaded photos.
- At geo-tag level, most tags from a landmark should be aggregated at a limited number of places. For example, Forbidden City is one of the famous landmarks in Beijing, and all photos from Forbidden City must be taken at the place with appropriate geo-tags.

Different from the existing area-driven geo-tag analysis methods, our landmark tag detection for Flickr tags starts from single tag.

4.1 Candidate popular tag selection

Given a location as input, we first crawl photos along with their associated tags and metadata from Flickr. We aim to select those tags that may be landmarks and also popular for the location. We also build up a vocabulary of irrelevant tags. This vocabulary includes following tag types.

- Single and simple words. Few landmarks can be named by a single and simple word, because it is not differentiable. For example, the Golden Gate Bridge can be provided with a tag “bridge”. However “bridge” cannot be used to identify the Golden Gate Bridge, because there are too many bridges. Other examples include “urban”, “house”, and “skyscraper”.

• Time and number related words.
  Time and number cannot be employed as a landmark’s name but properties, e.g. “October”, “2008”, “top-v1111”
• Camera related words.
  e.g. “Cannon”, “Canon17-40”, “450D”, “black&white”, “supershott”.
• Some popular tags in Flickr.
  e.g. “the unforgettable pictures”, “the perfect photographer”, “flickrs best”, “flicker fly”.

After removing the irrelevant tags by the vocabulary, the relevance metric between the location tag and its associated tags is defined as:

\[
R(tag_s, tag_{loc}) = \frac{P(tag_s|tag_{loc})}{P(tag_s)} \times \frac{\left| (tag_s, tag_{loc}) \right|}{\left| (tag_s) \right| \times \left| (tag_{loc}) \right|},
\]

(1)

where \(tag_{loc}\) is a location tag, and \(tag_s\) is an associated tag with \(tag_{loc}\). \(\left| (tag_s, tag_{loc}) \right|\) is the co-occurrence rate between \(tag_s\) and \(tag_{loc}\), \(\left| (tag_s) \right|\) and \(\left| (tag_{loc}) \right|\) are respectively the occurrence rates of \(tag_s\) and \(tag_{loc}\). The larger the \(R(tag_s, tag_{loc})\) value is, the more important is \(tag_s\).

To reduce the computation cost, we retain \(tag_s\) as the candidate popular landmark tags if and only if \(R(tag_s, tag_{loc})\) is greater than a predefined threshold \(T_R\), i.e., \(R(tag_s, tag_{loc}) > T_R\). Otherwise, we remove \(tag_s\) from the list.

We further employ the landmark information from Yahoo Travel Guide. The recommended attractions’ names are also selected as the candidate popular tags. Assuming that there are \(N_1\) tags of the candidate popular landmark tags in total, we can imagine that most real popular landmark tags are included in the top \(N_1\) tag group.

The reason why we cannot only select the landmark names in Yahoo Travel Guide is as follows. First, there are only a few number of cities in Yahoo Travel Guide. For these cities which have not been added in Yahoo Travel Guide, we cannot select candidate landmarks. Second, for cities added in Yahoo Travel Guide, the landmark names are not complete. There could be many landmark names not involved in Yahoo Travel Guide.

4.2 Noisy tag filtering based on geographical distribution

Although the top \(N_1\) candidate tags may be hot tags for the location, whether all of them are landmark tags cannot be guaranteed. In this step, we filter out those noisy tags (non-landmark tags) based on geographical distribution to find the real landmark tags.

Since a landmark should be located in a small region, we can measure the distribution of any tag with the geo-tag information from Flickr.

Noisy tag filtering using geographical information consists of four steps: removal of noisy photos, estimation of geographical distribution, calculation of tag attributes, and filtering of noisy tag. The images with geo-tags are crawled from Flickr using the location name as query.

4.2.1 Removal of noisy photos

Although the crawled images contain the location tags, there may still be some noisy geo-tags. Moreover, it probably downloads some inappropriate photos and tags on account that the query tag may be polysemy and contain noise.

For the first case, the query tag may point to more than one place. For example, there are two cities named Santiago the capital of Chile and a city in Spain.

For the second case, one image may be annotated with an inexact tag. For example, a photo displaying a favorite Beijing Dish in Tokyo may be labelled as “Beijing”, while the geo-tag is Tokyo.

Based on the analysis above, a geographic information de-noising procedure is employed for removing the inexact tags and photos. Using the geo-information to remove the irrelevant photos is necessary and feasible. If there is more than one place named as the query tag, we find one geo-location \(g_{center}(i)\) for each location respectively, and then geographically group images into clusters, where these geo-locations \(g_{center}(i)\) are used as the cluster centers. Each geo-tagged photo is grouped into the clusters based on the nearest neighbor criteria. Only images in the right cluster are employed as the correct images, and other images are removed.

To remove the wrong tags, we first randomly find a photo’s geo-location in the true location (longitude and latitude), and do a hierarchical agglomerative clustering on these images based on the geographical distance, keeping the distance between two images in one cluster sufficiently small. The cluster including the selected photo is reserved while others are removed as wrong images.

4.2.2 Estimation of geographical distribution for candidate popular landmark tags

In this step, we mine the geographical distribution of candidate popular landmark tags. When one tag geographically locates at a small region, it is probably a candidate tag of landmark.

Actually, one landmark tag may appear in more than one place. For example, “great wall” is a popular landmark in Beijing. There are many different great walls, and two most famous great walls are “Badaling” and “Si Ma Tai”. Thus, only finding the overall distribution of one tag is not accurate enough. We have to detect different landmarks sharing the same tag.

Given a set of photos associated to a tag after de-noising, we first geographically cluster these images and estimate its geographical distribution within each cluster, respectively. Here we employ the hierarchical agglomerative clustering as the clustering method, as it can guarantee that the geographical distance between two photos does not exceed a predefined threshold, which is experimentally selected.

To further remove noisy photos, only clusters with adequate enough are retained for further processing (in our experiments, we use “more than 25% of photos of the original cluster” as the threshold). In each cluster, the distribution of geo-tags is modelled as a Gaussian distribution. Then the geo-distribution for the \(i^{th}\) cluster is \(p_i(x) = N(x|\mu_i, \Sigma_i)\), where \(\mu_i\) and \(\Sigma_i\) are trained from geo-tags. One advantage of employing Gaussian distribution is that we can further predict possible geo-locations for the tag.

4.2.3 Calculation of Tag Attributes

Till now, we have obtained a set of tag distributions and corresponding photo sets. Let the number of obtained clusters be \(N_2\). Here tag attributes are taken into account. Given the \(k^{th}\) photo cluster \(PC(k)\) and the associated tag \(tag_k\), the attributes of \(PC(k)\) include the number of users,
the content score provided by each user, the overall content score for the tag in that region, and the area of the region.

The details about the four attributes are introduced as follows.

- The number of users ($N_{user}$)
  $N_{user}$ is the number of photo uploaders in $PC(k)$. $N_{user}$ indicates how many people have contribution to photos of this region with tag $t_k$. The more users there are, the more popular the region is.

- Content score for each user ($IC_{user}$)
  Each user may not only submit one image for the tag. $IC_{user}$ describes how much content one user provides for the tag in this region. This attribute should be high relevant to the number of images uploaded by the user. On one hand, when the number of photos for the tag becomes larger, the information content should also increase. On the other hand, the relationship between the information content and the number of images provided by one person should not be linear.

- Overall content score for the cluster ($CS$)
  Overall content score describes the overall content provided by all users in the cluster. $CS$ employs both $N_{user}$ and $IC_{user}$ to show how much information content is included from these photos, and it is the sum of all $IC_{user}$ from all owners. $CS$ is calculated by:

$$CS = \sum_{t=1}^{N_{user}} IC_{user}(t).$$

- Area of the cluster ($Area$)
  $Area$ represents the area value covered by images in the region. $Area$ can provide information about whether this region is a landmark with respect to the geographical size. Here $Area$ is computed in the sphere space.

### 4.2.4 Filtering of noisy tag based on tag attributes

We use the attributes calculated in the last step to determine whether one candidate tag can be retained as a landmark tag or not. As described above, a tag may become a real popular landmark tag only when it is located within a small region, and has high importance factor.

Here, the $Area$ and the $CS$ values of each photo cluster are employed to filter tags. $T_{Area}$ is utilized to show whether the region is small enough for one landmark, and $T_{CS}$ describes whether this tag is popular to users. In our tests, $T_{CS}$ is experimentally selected. $T_{Area}$ is selected as an area value of a large landmark. In our experiments, we use twice the area of Summer Palace in Beijing (a large landmark) as $T_{Area}$.

When a tag cluster $cluster_i$ satisfies the following conditions, the associated tag is selected as a popular landmark tag; otherwise it is filtered out as a noisy tag.

$$\begin{cases} CS(cluster_i) > T_{CS} \\ Area(cluster_i) < T_{Area} \end{cases}$$

### 4.3 Tag refinement using geographical relevance

Let $N_t$ be the number of tags selected in the last step, and these tags are denoted as $\{tag_{1t}, tag_{2t}, \ldots, tag_{N_t} \}$. Note that these $N_t$ tags may contain repetitive tags. Here “repetitive tags” means several tags describe the same landmark from different aspects, e.g. the location name and the name of the landmark. This is because these tags were analyzed separately before, and no relevance between tags has been taken into account.

We observe that one landmark may contain more than one representative tags due to different tagging convention. For example, “Si Ma Tai” is one famous great wall in Beijing, while the geographical location is named as “Dong Zhuang He”. Using the above processing procedure, both of these two tags can be retained as popular landmarks.

This type of repetition can be detected using geographic information. When two or more tags describe the same landmark, they should have highly relevant geographical distributions. We have modeled the geographical distribution of tags as Gaussian distribution in previous procedures. Therefore, these $N_3$ tags have $N_3$ Gaussian distributions now.

Given two tags $tag_i$ and $tag_j$, denote the corresponding geographical distributions as $p_i$ and $p_j$, we can employ the symmetric Kullack-Leibler (KL) divergence [10] to measure the distribution difference in distribution between $p_i$ and $p_j$, where:

$$KL(p_i||p_j) + KL(p_j||p_i),$$

$$KL(p_i||p_j) = \int p_i(x) \log \frac{p_i(x)}{p_j(x)} dx.$$

The KL divergence of two Gaussian distributions has a closed formed expression as shown in Equation (7).

$$KL(p_i||p_j) = \frac{1}{2} \left[ \log |\Sigma_j| + T (\Sigma_i^{-1} \Sigma_j) \right] - d + (\mu_i - \mu_j)^T \Sigma_j^{-1} (\mu_i - \mu_j)^T.$$

When the KL divergence between two distributions is small, these two distributions have high probability to be similar. When the KL divergence between two tags’ distributions is below a predefined threshold, these two tags represent the same landmark. Till now, repetitive tags representing the same landmark can be grouped into one cluster. In each cluster, the tag with high frequency is selected for further processing, and all photos in this cluster are associated to the selected tag.

## 5. LANDMARK RANKING AND REPRESENTATIVE VIEWS GENERATION

### 5.1 Candidate landmark ranking

In the last step, assume that a total of $N_4$ landmark tags $\{tag_1, tag_2, \ldots, tag_{N_4}\}$ are selected. To rank these landmarks, two factors of tag attention and content are taken into account. Here two types of information are employed.

The first is the context information from media-sharing website (Flickr), and the second is the user-contributed data from existing travel guide system (Yahoo Travel Guide). As mentioned in Section 1, on one hand, there is much noise in the media-sharing website; on the other hand, the user-contributed data may not be credible. There two types of information can be used as complementary to each other.
<table>
<thead>
<tr>
<th>City</th>
<th>Method</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>W2Go</td>
<td>Forbidden City</td>
<td>Temple of Heaven</td>
<td>Summer Palace</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td>Great Wall</td>
<td>Lama Temple</td>
<td>Forbidden City</td>
</tr>
<tr>
<td>Chicago</td>
<td>W2Go</td>
<td>Millennium Park</td>
<td>Sears Tower</td>
<td>John Hancock Center</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td>Art Institute of Chicago</td>
<td>Museum of Science and Industry</td>
<td>Sears Tower Skydeck</td>
</tr>
<tr>
<td>New York</td>
<td>W2Go</td>
<td>Central Park</td>
<td>Times Square</td>
<td>Statue of Liberty</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td>Time Square</td>
<td>Central Park</td>
<td>St. Patrick’s Cathedral</td>
</tr>
<tr>
<td>San Francisco</td>
<td>W2Go</td>
<td>Alcatraz</td>
<td>Golden Gate</td>
<td>Bay Bridge</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td>Golden Gate Bridge</td>
<td>Alcatraz Island</td>
<td>Japanese Tea Garden</td>
</tr>
<tr>
<td>Paris</td>
<td>W2Go</td>
<td>Eiffel Tower</td>
<td>Arc de Triomphe</td>
<td>Notre-Dame</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td>Tour Eiffel</td>
<td>Musee d’Orsay</td>
<td>Basilique du Sacre-Coeur (Le)</td>
</tr>
<tr>
<td>Hangzhou</td>
<td>W2Go</td>
<td>West Lake</td>
<td>Lingyin Temple</td>
<td>Hangzhou Park</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bremen</td>
<td>W2Go</td>
<td>Bremer Haven</td>
<td>Cathedral of St Peter</td>
<td>Town Hall</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kobe</td>
<td>W2Go</td>
<td>Kobe Port Tower</td>
<td>Akashi Kaikyo Bridge</td>
<td>KOBE Rokko</td>
</tr>
<tr>
<td></td>
<td>Yahoo_TG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 5.1.1 Tag attention factor

Tag attention factor evaluates how popular the landmark is. It is determined by two components: tag relevance in Flickr and the number of times the landmark is reviewed in Yahoo Travel Guide. The number of landmark reviews can be obtained in the attraction page of Yahoo Travel Guide. Tag relevance $R(tag_i, tag_{loc})$ is computed for all $N_t$ candidate tags. Assuming that the number of reviews for the $i^{th}$ landmark is $R(i)$, let the maximal value in all $N_t$ tags be $\text{Max}_t(R(tag_i, tag_{loc}))$, the maximal review number in Yahoo Travel Guide be $\text{Max}_r(\text{review})$, then the tag attention factor for the $i^{th}$ landmark is defined as follows:

$$F_{tag}(i) = \frac{R(tag_i, tag_{loc})}{\text{Max}_t(R(tag_i, tag_{loc}))} + \frac{R(i)}{\text{Max}_r(\text{review})}, \quad (8)$$

### 5.1.2 Content factor

Content factor aims to assess the content information of each landmark tag. Content factor includes the content score in Flickr and the evaluation score in Yahoo Travel Guide for each landmark.

Let the maximum $CS$ score in all $N_t$ tags be $\text{Max}_c$, then the content factor for the $i^{th}$ landmark is defined as follows:

$$F_{con}(i) = \frac{CS(i)}{\text{Max}_c} + \frac{ES(i)}{S_{max}}, \quad (9)$$

where $ES(i)$ is the evaluation score for the $i^{th}$ tag, which is based on the judgment of reviewers. $S_{max}$ is the maximal value of the evaluation score in Yahoo Travel Guide.

Finally, the overall ranking score is calculated as:

$$S_{\text{ranking}}(i) = \alpha \times F_{tag}(i) + \beta \times F_{con}(i), \quad (10)$$

where $\alpha$ and $\beta$ are weighting factors. For a given location, all generated landmarks are ranked using $S_{\text{ranking}}$ in a descending order.

### 5.2 Representative view generation for ranked landmark

Once we have generated a list of landmarks from the original data, we try to find the visual representative views for each landmark. Here a similar method to [4] is employed to select representative views. For a given landmark $LM_i$ with $tag_{LM_i}$, first all photos associated with $tag_{LM_i}$ are grouped into clusters based on local feature (SIFT [13]). We employ Normalized Cuts algorithm [19] for clustering.

After that, the top $K$ ($K$ is set to five in our experiments) largest sub-clusters are selected, and the representative views are chosen from these sub-clusters. To select better views, following criteria are used:

- No face included in the view
- As we aim to show the landmark, we first implement face detection on all the images, and only images without face can be candidate views.
- Large similarity with other images in the same sub-cluster
- The representative view should be similar with other images. Thus it is able to represent other views well.
- In all candidate views, the view with the largest average similarity in the sub-cluster is selected as one representative view.

### 6. EVALUATION

To evaluate the performance of W2Go system, we compare it with a baseline method and the Yahoo Travel Guide.

#### 6.1 Dataset

Eight cities are selected as the testing locations to evaluate the effectiveness of the proposed W2Go system using ALR. They are Beijing, Chicago, New York, San Francisco, Paris, Hangzhou, Bremen, and Kobe. These cities are chosen due to two reasons. First, these cities are world famous, and people are familiar with them. So the experimental evaluation can be accurate. Second, these cities are big enough, and there are adequate images with geo-tags. It is noted that the last three cities have not been included in Yahoo Travel Guide.

For each city, we crawl geo-tagged images from Flickr with tags and metadata using the Flickr API [6].

#### 6.2 Evaluation of W2Go using ALR

We conduct two user studies to evaluate W2Go. The first user study is to evaluate the usefulness of W2Go, and the second user study is to compare W2Go with other methods/systems.

We have invited 25 people to join the user study. These users come from Singapore, China, and United State, respectively. There are 16 men and 9 women, and these subjects cover wide background, including graduate students,
researchers, sales people, teachers, and business people. Their ages range from 18 to 47.

In the experiment, given a location name, the proposed W2Go system first automatically crawl the tag/geo-tag information and metadata from Flickr as well as the user knowledge from Yahoo Travel Guide. Generally, around 100,000 tags and associated photo information, and up to tens of thousands of geo-tagged photos are crawled from Flickr as the dataset. Note that in some cases, only thousands of photos attached with geo-tags are available. Thus the precise number of crawled photos depends on the resources. ALR runs on the dataset, and generates a list of ranked landmark. One examples of output is shown in Figure 5. Top three landmark suggestions for each city are given in Table 1.

6.2.1 Evaluation of effectiveness

To evaluate the effectiveness of W2Go using ALR, participants were asked to answer the following questions, by providing a rating score of between 1 (bad) to 10 (excellent):

1. All selected tags should represent landmarks. Based on the set of tags given, how do you rate the performance of W2Go on landmark tag selection?

2. The output landmarks should be the important landmarks in that location. Based on the set of landmarks returned, how do you rate the performance of W2Go on important landmark ranking?

3. The tag refinement by geographical analysis aims to merge two or more tags representing the same landmark. Based on the set of tags given, how do you rate the performance of W2Go on tag refinement?

4. W2Go is used for travel assistant. In your opinion, how helpful is W2Go as a travel recommendation system?

Figure 6 shows the evaluation results from all 25 participants. The results show that most users think that the selected tags represent the real landmarks well.

Figure 7 shows the evaluation results from all 25 participants for the second question. The ranked list of landmarks provided were regarded by 76% of users to be important (with scores of ≥ 7).

For the third question, users were again required to give a score from 1 to 10. This judgment depends on the W2Go’s tag refinement performance using geographical analysis. A higher score means less tag repetition in the final results. Figure 8 describes the evaluation results. The results show that 88% of participants were happy that W2Go has “few repetitive tags” (with a score ≥ 8).

The fourth question evaluates the effectiveness of W2Go in providing travel assist. Figure 9 describes the evaluation results. W2Go was rated to be “very helpful” (with scores of ≥ 8) by 40% of participants, and be “somewhat helpful” (with scores of ≥ 6) by another 32% of all participants.

6.2.2 Comparison with other methods

From Flickr tags, top ten tags with high co-occurrence rates are selected as the popular landmark recommendations for each city. This result is used as the baseline, named as Flickr_CO. In Yahoo Travel Guide, there is an ATTRACTION selection for each city. Top 10 ATTRACTIONS are selected as the second compared method, named Yahoo_TG.

In this overall evaluation, comparisons between the effectiveness of W2Go and Flickr_CO and Yahoo_TG were conducted. Users were asked to give a score of between -5
(bad) to 5 (excellent) for the question: “how is the landmark tags returned by W2Go compared with Flickr_CO and Yahoo_TG, respectively?” The evaluation results are shown in Figure 10 and Figure 11.

The top three landmark suggestions for all eight cities are provided in Table 1. Figure 11 shows the comparison between W2Go and Yahoo_TG. As indicated from the data, 20% participants agreed that W2Go and Yahoo_TG are with similar performance, 20% users held the view that W2Go is slightly worse, while 60% of participants voted that W2Go is better than Yahoo_TG. Generally, W2Go is somewhat better as compared with Yahoo_TG. For the last three cities, W2Go generates a series recommended landmarks, while no information can be found in Yahoo Travel Guide.

As Yahoo_TG requires manually edition, only a few cities are included in Yahoo_TG system’s travel guide. For example, only several big cities from China are included in Yahoo_TG, such as Beijing, Shanghai, and Hongkong. When other cities that have not been added into the system are requested, Yahoo_TG cannot provide any answers. Moreover, even when a city has been included in Yahoo_TG system, the landmarks listed in Yahoo_TG should be separately entered manually. Thus the system needs to be updated frequently for the landmark list to be up-to-date. For example, Water Cube and Bird’s Nest (National Stadium) are two new landmarks in Beijing, and they were built around 2008. These
two landmarks are relatively important now, but they have not been included in the top 50 landmarks in Yahoo_TG (until October 30th, 2009).

For United States, although there are many cities included in the system for many states, most of the cities are only with few landmarks, with some even none. For example, there is no attraction suggestion for Surfside Beach, SC, and only 9 attractions for Harrisonburg, VA (until October 30th, 2009). Compared with Yahoo_TG, W2Go works automatically, and no human assistance is needed. Therefore, W2Go can deal with all locations with geo-referenced images all over the world.

6.2.3 Discussion

As shown in the experimental evaluation and the comparison with other methods, the proposed ALR is effective. It is worthy of noting that there are also some limitations, which are listed and discussed as follows.

1. Falsely recognize an event or activity tag as a landmark.
   Since some famous events or activities are always hosted in one place, these tags associated with them may be falsely detected as landmarks. For example, “Midi Music Festival” was hosted in Beijing’s Haidian Park, which is China’s largest rock music festival. It was falsely recognized as a landmark. Another good example is the event “Golden Star Award” in New York. Removal of this type of errors depends on more accurate recognition of tag semantics.

2. Two adjacent but different landmarks are falsely merged as one landmark.
   This type of errors comes from the geotag based analysis. When two landmarks are quite close, their geographical distributions may have large overlaps and high KL divergence. They can thus be merged into one landmark. One example is that “Aon Center” and “Grant Park” in Chicago are two adjacent landmarks, while they are falsely merged as one landmark.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a dynamic travel recommendation system named W2Go by employing automatic landmark ranking (ALR) techniques. The system not only utilizes photo statistics from Flickr, but also user knowledge from Yahoo Travel Guide. ALR is designed to be able to select candidate landmarks from these two resources and then rank them. After that, representative views are generated for these landmarks. The prototype system of W2Go is built up based on ALR, and the user based evaluations, which 25 participants are involved, demonstrate that the system can provide effective assistance to travellers.

There are two limitations of ALR. First, it is hard to remove tags of events and activities from the landmark list as they are location-sensitive. Second, two geographical adjacent landmarks are prone to be merged in some cases. Our future work will target at the two limitations mentioned above. Furthermore, we will study the activities of travellers and users from travel guide systems to improve W2Go.

8. ACKNOWLEDGEMENTS

This work was supported by the National Basic Research Project of China (No.2010CB731800), the Science Fund for Creative Research Groups of NSFC (No.60721003), NSFC-Guangdong Joint Fund (No.U09035001), the National High Technology Research and Development Program of China (No.2009A0101Z329), and was partially supported by NRF/IDM Program of Singapore, under Research Grants NRF2007IDM-IDM002-047.

9. REFERENCES