ITEMS: Intelligent Travel Experience Management System

Chih-Chieh Liu[2], Chun-Hsiang Huang[1], Wei-Ta Chu[1], and Ja-Ling Wu[1,2]
[1]Department of Computer Science and Information Engineering
[2]Graduate Institute of Networking and Multimedia
National Taiwan University
{ja,bh,wtchu,wjl}@cmlab.csie.edu.tw

ABSTRACT
An intelligent travel experience management system, abbreviated as ITEMS, is proposed to help tourists organize and present the digital travel contents in an automatic and efficient manner. Readily available metadata are adopted to reduce the overhead of user intervention and manual annotation. Robust image similarity metrics are also incorporated to utilize the power searching capability of WWW search engines. The proposed system automatically identifies the embedded geo-information of personal media, and accordingly integrates media with map and text-based schedule to facilitate travel experience management and presentation. We show several prototypes in two application scenarios and demonstrate the effectiveness of the proposed system.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Retrieval models.* H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *Synchronous interaction.*

General Terms
Algorithms, Management, Experimentation.

Keywords
Automatic image annotation, photo presentation.

1. INTRODUCTION
Due to the advent of high-quality digital imaging devices and large-capacity storages, recording travel experience in the form of digital photographs and audio/video clips have become indispensible activities throughout a tour. However, managing hundreds or even thousands of media files after traveling are now nightmares for most travelers. Since manually organizing and annotating travel experience to facilitate future presentation are time-consuming and labor-intensive processes, most people simply stock unorganized travel data in storage devices and never revisit them at all.

Currently, many companies have provided album services for web users to store and publish personal media contents [1][2]. Apparently, scenery photos, tour video or travel notes are one of the major types of content sources for these websites. However, though some vendors have provided preliminary tagging mechanisms, semantic-level media annotation, organization and presentation still require intensive user intervention. While currently most services engaged to simplify the flow of content uploading and publishing, automatic content analysis/organization capabilities have not been successfully integrated into such services. Most web album users still have to iteratively manipulate a great deal of digital contents. Moreover, tedious slideshow remains to be the only choice of dynamic presentation in most services.

In this paper, an intelligent travel experience management system, abbreviated as ITEMS, is proposed to help users organize and present their travel media in an automatic and semantic-meaningful manner. Contents are organized and annotated with the help of metadata that can be conveniently and readily obtained. Metadata that serve as the bases for automatic annotation may come from the web or from the official information provided by tourism bureaus. After automatic annotation, various presentation methodologies, especially the geographically demonstrated multimedia tour (denoted as *GeoTour* in subsequent discussions), can be generated on the fly. In addition to implementation details and extensive experimental results, the benefits and limitations of the ITEMS will also be discussed.

This paper is organized as follows. Section 2 illustrates the application scenario and overview of the ITEMS. In Section 3, details of the automatic image annotation processes are illustrated. Functional blocks related to travel experience presentation are specified in Section 4. Extensive experimental results are provided in Section 5. Limitations and potential extensions of the ITEMS are discussed in Section 6. Section 7 concludes this paper.

2. SYSTEM OVERVIEW
The relationships between the proposed system and relevant entities are illustrated in Figure 1. Note that the input data of ITEMS can be roughly classified into two categories: travel contents and relevant metadata. Travel contents consist of various forms of multimedia generated during the user's journey, including digital photographs, video clips, audio/speech recording, as well as textual travel notes. Since the proposed system is aimed at solving content management tasks for the user, the input travel contents are expected to be in their originally unorganized status. The only one requirement of the input content is that the

MIR'07, September 28-29, 2007, Augsburg, Bavaria, Germany.

accompanying time information must be preserved for further analysis. This requirement is reasonable because, in most multimedia file format, time information is readily available since acquirement of each media file.

On the other hand, the most important metadata required by ITEMS is a digital travel schedule. A digital travel schedule must clearly specify time slots in a trip and their corresponding scene sites. In real-world scenarios, the digital travel schedule can be either provided by travelers using easy-to-use software or offered by traveling agencies following interoperable document standards. Furthermore, additional information related to a scenic site, such as textual description or scene photos, can be obtained from the web using search engines or provided by travel agency/ sightseeing offices in governments. In addition, if an annotated map is available, users can enjoy a novel content presentation experiences based on geographic relationships. Figure 2 shows a schedule example that is written in XML and consists of the names and introduction of visited scenic spots in each day.

Note that the assumption that annotated maps and additional scene information are available is feasible since, for each scene site, one version of these metadata suffices to be utilized by all of its tourists. In other words, the commercial values provided by utilizing these additional metadata for automatic content management is much larger than the necessary costs.

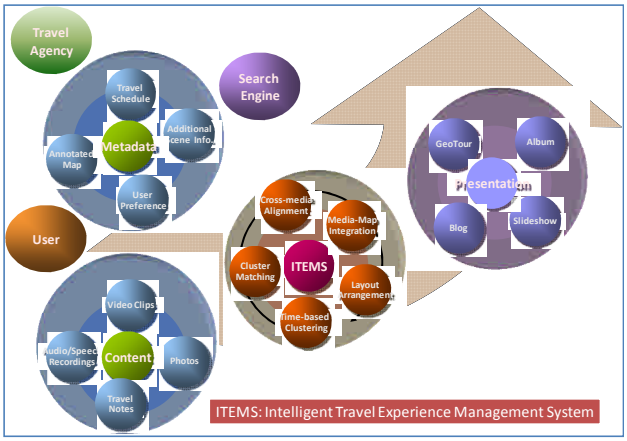

Figure 1. The proposed travel experience management system.

```
- <schedule>
  - <day>
    - <spot>
        <name>Les Invalides</name>
        <intro>Les Invalides in Paris, France consists of a complex of buildings in the 7th arrondissement containing
          museums and monuments, all relating to the military history of France, as well as a hospital and a retirement
          home for war veterans, the building's original purpose. It is also the burial site for some of France's war
          heroes.</intro>
      </spot>
    - <spot>
        <name>Arc de Triomphe</name>
        <intro>The Arc de Triomphe is a monument in Paris that stands in the centre of the Place Charles de Gaulle, formerly
          the Place de l'E'toile, at the western end of the Champs-E'lyse'es. The arch honours those who fought for France,
          particularly during the Napoleonic Wars, and today also includes the tomb of the unknown soldier.</intro>
      </spot>
    - <spot>
        <name>Notre Dame</name>
        <intro>Notre Dame de Paris, often known simply as Notre Dame in English, is a Gothic cathedral on the eastern half
          of the I^le de la Cite' in Paris, France, with its main entrance to the west. It is still used as a Roman Catholic
          cathedral and is the seat of the Archbishop of Paris. Notre Dame de Paris is widely considered one of the finest
          examples of French Gothic architecture. It was restored and saved from destruction by Viollet-le-Duc, one of
          France's most famous architects. Notre Dame translates as "Our Lady" from French.</intro>
      </spot>
    </day>
  - <day>
    - <spot>
        <name>Palace of Versailles</name>
```

Figure 2. An XML-based digital travel schedule, which consists of the names and introduction of visited scenic spots in each day.

With the help of metadata, the ITEMS can organize and annotate travel contents with none or least user intervention. Input photographs are firstly clustered according to time information stored in file headers and labeled with possible scene site information according to schedule. Then, clustered photos are matched with additional scenery photos provided by travel agency or obtained via search engine. In cases where audio and video contents are provided, matched photos serve as basis of cross-media alignment. Finally, clustered media data can be associated to annotated maps to realize geographical presentation of travel contents. If textual travel notes are provided, automatic text-photo alignment and layout arrangement can be proceeded to generate media-rich blog articles.

In this paper, we will focus on the geography-based presentation of travel contents, denoted as GeoTour in following sections. The readers who are interested in slideshow scheme may refer to [15] for a novel tiling slideshow scheme. Automatically generating travel experience-based web articles is one of our in-progress research topics.

3. AUTOMATIC IMAGE ANNOTATION

3.1 Problem Formulation

Since Yeh et al. [3] proposed their idea, automatic image annotation based on web information has inspired a new direction for image understanding. Conceptually, the automatic image annotation problem has been modeled as follows.

Given a test image i, find a keyword k^* from a keyword pool such that posteriori probability $p(k|i)$ is maximized.

$$k^* = \arg\max_{K} p\left(k\,|i\right), \quad K = \left\{k_1, k_2, ..., k_M\right\}, \tag{1}$$

where there are M candidate keywords from the pool. In previous works [3][4][5], the test image is first used as a query image to perform content-based image retrieval to a well-develop image database. Images in this database are elaborately collected and convey rich metadata. Keywords relevant to this test image are then collected from the corresponding metadata of retrieved images. Based on these keywords, these systems perform web-based image retrieval to obtain relevant images and then use them to rank keywords. Figure 3 shows the conventional framework. To rank keywords, different variations has been proposed, such as comparing the test image with retrieved images [3], keyword merging [4], or search result clustering [5].

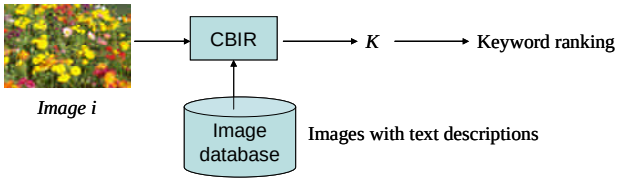

Figure 3. The framework of previous image annotation works.

In previous works, a well-developed bootstrap image database is needed because, in most cases, the required keywords for web searching are not available. However, in the case of travel contents management, automatic annotation for travel photos can be facilitated since people often have a travel schedule before traveling. The schedule is often provided by travel agencies or

prepared by the travelers themselves and usually consists of scenic spots in the temporal order of visiting. Therefore, if both a traveling schedule and the photos taken in this journey can be given, automatically annotating travel photos can be facilitated. This prerequisite is different from that of general image annotation problems.

Moreover, people often take a great deal of photos within a single scenic spot. Although photos that are temporally close are expected to represent the same scenic spot, their visual appearances may vary significantly. Therefore, existing content-based image matching and keyword merging approaches are not suitable for our system. It would be more reasonable to annotate a cluster of photos with the same spot name.

Assume that there are M spot names $K=\{k_1,k_2,\ldots,k_M\}$ in the travel schedule, and there are N clusters of photos $I = \{IC_1,IC_2, \ldots,IC_N\}$. Assume that every scenic spots were visited and some photos were taken in each spot, i.e. $M \leqq N$. The annotation problem is to find the most matched image cluster IC^* corresponding to a given keyword k_j:

$$IC^* = \arg\max_{i=1,\ldots,N} p\left(IC_i \middle| k_j\right). \quad (2)$$

Based on the spot name (keyword)[1] k_j, we retrieve relevant photos from web-based search engines. Assume that $IS_j=\{s_1,s_2,\ldots,s_m\}$ is the set of retrieved photos based on the keyword k_j, the probability $p(IC_i|k_j)$ is approximated as:

$$\begin{aligned} IC^* &= \arg\max_{i=1,\ldots,N} p\left(IC_i \middle| k_j\right) \\ &\approx \arg\max_{i=1,\ldots,N} p\left(IC_i \middle| IS_j\right). \end{aligned} \quad (3)$$

In contrast to previous works, we do not need a well-developed image database. However, the reference photos retrieved from the WWW are obviously noisy. We need to develop a robust matching method to compare the photos taken by users with photos retrieved from the web.

3.2 Time-based Clustering

Tourists often take a great deal of photos when they arrive at a scenic spot. On the other hand, we rarely take photo during the traffic from one spot to another. Therefore, photos taken in different spots can be distinguished by checking the changes of shooting frequency.

To characterize the shooting frequency, we exploit a time-based clustering algorithm proposed in [6]. Photos are first sorted by their creation time. This algorithm dynamically determines noticeable time gaps through checking the temporal context of photos in a sliding window, say 10 photos, and reveals the change of shooting pace. After this process, photos that are categorized into the same cluster are assumed to be taken at the same scenic spot.

[1] In this paper, "spot name" and "keyword" are used interchangeably. "Photo" and "image" are also used interchangeably.

3.3 Cluster Matching

3.3.1 Features

Given clusters of photos and the retrieved images based on spot names, the next problem is to estimate and maximize $p(IC_i|IS_j)$. Note that our goal is to "assign a cluster of photo to a spot name", given the limitations that some clusters of photos may be taken between two spots, e.g. in transportation, and don't belong to any spot name.

Using the spot names k_j in the travel schedule as a query, we can obtain reference images IS_j from web-based search engines. The probability $p(IC_i|IS_j)$ is then estimated based on the similarity between the reference images and clusters of user's photos. The previous works [3][4][5][7][8] primarily exploit content-based image similarity to do this estimation. However, the same task for travel photo annotation poses two different challenges:

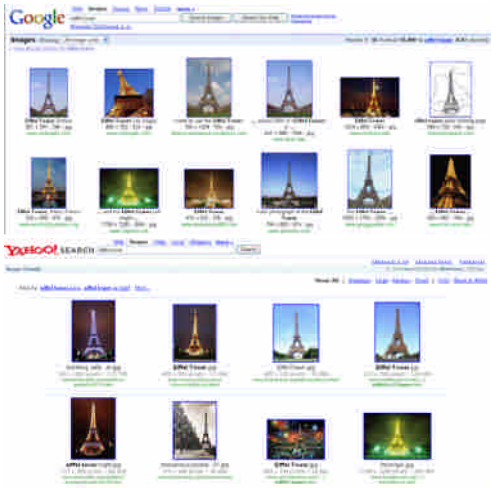

Figure 4. Sample results of searching "Eiffel Tower" from Google and Yahoo! Image search engines.

1) In our work, we retrieve top ten search results from Google, Yahoo!, and Flickr. Although using web-based search engine is convenient, the retrieved images often significantly vary in visual appearance, scales, or shoot angles. Figure 4 shows some sample results of searching "Eiffel Tower" from Google and Yahoo! image search engines. Due to large visual variations in scales and viewing angles, it is hard to use conventional content-based image features, such as dominant color or edge information, to well estimate the image similarity.

2) Even if a reference image is visually similar to a user's photo based on color or edge information, they don't necessarily belong to the same scenic spot. Users are brilliant in recognizing scenic spots and are very sensitive to matching errors, especially famous landmarks like Eiffel Tower or Arc of Triumph. For travel photo annotation, a small matching error may significantly impede efficient management and presentation.

With these concerns, a more reliable feature for image matching is required. We would rather emphasize the accuracy of matching than pursuing zero misses. In this work, we utilize the scale-invariant feature transform (SIFT) [9] as the basis for image

matching. This feature is invariant to image noises, rotation, scaling, and small changes in different viewpoints. It is demonstrated to be reliable in image registration and is widely used in object tracking researches.

If there are n photos in the cluster IC_i and m photos in the retrieved images IS_j, we approximate the similarity between IC_i and IS_j as follows.

$$p\left(IC_i \middle| IS_j\right) \approx \max_{\substack{p=1,\ldots,n \\ q=1,\ldots,m}} \left(mSIFT\left(c_p, s_q\right)\right), \tag{4}$$

where $mSIFT(c_p,s_q)$ denotes the number of matched SIFT points between the image c_p and s_q. To alleviate the influences of many noisy reference images, we use the largest number of SIFT matched points of any image pair to represent the similarity between IC_i and IS_j. Figure 5 shows the number of matched points in different situations. We can obviously see that photos contain the same landmark, though they are in different scales and viewing angels, would have more SIFT matched points. In the real implementation, we utilize the package developed in [16] to calculate SIFT features.

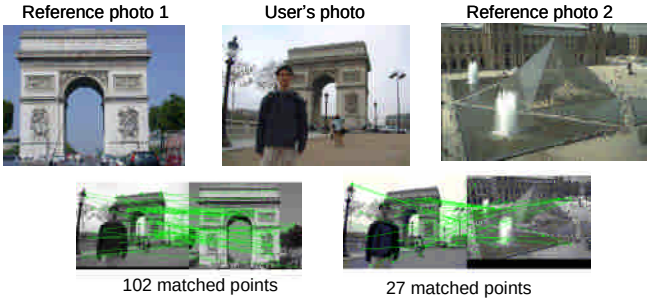

Figure 5. Number of SIFT matched points in different situations.

3.3.2 *The Matching Algorithm*

For the ease of description, we denote the largest number of matched points between IC_i and IS_j as $S_{i,j}$. The algorithm to match reference images and user's photos are described as follows.

Given M clusters of reference images $IS = \{IS_1, IS_2, \ldots, IS_M\}$ and N clusters of user's photos $IC = \{IC_1, IC_2, \ldots, IC_N\}$. Find a subset of IC that consists of M clusters (IC^*) such that the total matched points between IC^* and IS are maximal.

$$IC^* = \arg\max \sum_{IC_\ell \subset IC} S_{\ell,j}, \quad \left\|IC^*\right\| = M. \tag{5}$$

Figure 6 shows an illustrative example about matching five clusters of user's photos and three scenic spots. Note that both sets of clusters are sorted according to temporal order, and we assume that only one scenic spot name is related to a cluster of user's photos. After checking different assignment combinations, we will select the assignment that causes the maximal matched points. In this example, it is obvious that IC_1, IC_3 and IC_4 are best assignment candidates that match each reference clusters retrieved from three spot names respectively. Therefore, this process finally annotates IC_1 as spot 1, IC_3 as spot 2 and IC_4 as spot 3.

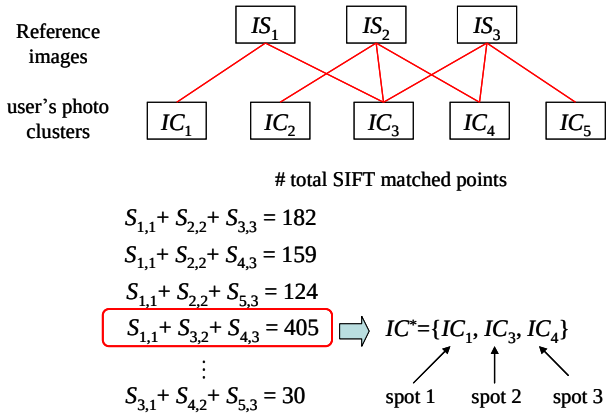

Figure 6. An illustrative example of cluster matching.

Although the matching process described above can successfully find out the alignment between spot names and user's photos, calculating the SIFT feature points and performing nearly exhaustive matching is computation intensive. To enhance the performance of image matching, we devise a lightweight filtering method to efficiently decrease the image pairs required for comparison.

3.4 Lightweight Filtering

The lightweight filtering should be based on features that are computationally efficient. Therefore, we first study the relationships between the number of SIFT matched points and some image distance metrics derived from conventional content-based features. The content-based features we extract include intensity histogram, edge histogram, and color layout. Because extracting intensity histogram is the fastest, we take it as the instance in following discussions. Figure 7 shows the relationship between distances calculated based on SIFT matched points and intensity histogram. A point in this figure denotes two types of distances between a reference image and a test photo.

After more than 12000 tests on SIFT-based image matching, we found that only image pairs with more than 100 SIFT matched points are likely to be in the same place, i.e. the regions I and IV of Figure 7. According to Figure 7, there are many image pairs with large intensity distance and few SIFT matched points, i.e. the region II, but very few image pairs with large intensity distance and large SIFT matched points, i.e. the region I. Therefore, we can calculate the intensity histogram distances between targeted image pairs and ignore those with large distance, i.e. the ones falling into the regions I and II. Note that the deciding the horizontal boundary between the upper part and the lower part is a tradeoff. If the boundary moves up, e.g. 0.7, fewer matches would not be missed, but more image pairs should be compared. The opposite situation appears if the boundary moves down.

According to our experiments, more than one fourth of image pairs can be filtered out by checking intensity histogram distance in advance. Calculating distance based on intensity is significantly faster than that based on SIFT (about 300 times faster). Therefore, we can largely reduce the time needed to do clustering matching described in the previous section.

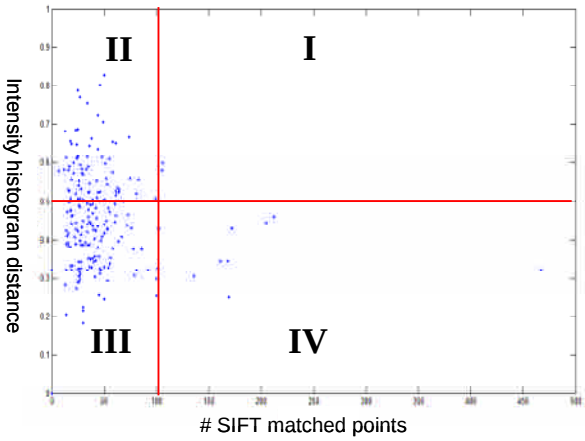

Figure 7. The relationship of the number of SIFT matched points and intensity histogram difference.

4. TRAVEL EXPERIENCE PRESENTATION

The automatic annotation approach described above eases the need of well-annotated image databases and effectively recognizes the scenic spots conveyed by personal photos without the assistance of any special equipment like GPS. In this section, we will demonstrate that the proposed approach is not conflicting with previous works like [3] and [4]. We will show two common scenarios of personal travel experience management.

- Global scenario: In this case, we may travel around many scenic spots in a few cities. There is no well-developed image database to be the confident reference basis. But as long as the travel schedule and personal photos are available, we can leverage web-based image search to find out reference images and use them to automatically annotate personal photos.
- Local scenario: In contrast to the global case, we may travel around a large scenic spot for several hours or a half-day. Commonly, there would be an administrative institute that can provide very detailed information, including the map of this scenic spot and the locations of distinguishing buildings or landscapes. In this case, we can simply replace the web-based searching results by an official image database and carry out automatic annotation of personal photos in a similar manner.

After automatic annotation, we can integrate photos with geographic information like maps to reconstruct the tour according to the user's photos. We name this presentation methodology as *GeoTour* because geographic information is automatically recognized and incorporated in photo presentation.

4.1 GeoTour

4.1.1 Cross-Media Alignment

Although photos are undoubtedly the primary media to capture travel experience, more and more tourists begin to capture video or audio clips in a journey. Once where a cluster of photos were taken can be identified with the schemes illustrated in Section 3, we can propagate this information to other temporally adjacent media and identify where a video or a voice clip was recorded accordingly.

4.1.2 Media-Map Integration

After identifying corresponding scenic spot names of user's captured media, we can find the longitude and latitude of the scenic spots through consulting a geographic database like [10]. This geo-information is then used to locate user-captured media on a map.

4.1.3 GeoTour in the Global Scenario

Figure 8 shows the framework for generating GeoTour in the global scenario and local scenario. In the global case, we find the reference images from the web. After cluster matching, we integrate geographic information and multiple travel media to generate a map-based presentation.

Figure 9 shows an example of a GeoTour in global scenario. The center of the map shows current scenic spots, and the right part of this interface shows the user's photos that are identified as capturing in this spot. The bottom part shows text descriptions from the travel schedule or the travel notes written by users. Users can browse photos from one spot to another, with the auxiliary information of map and travel schedule.

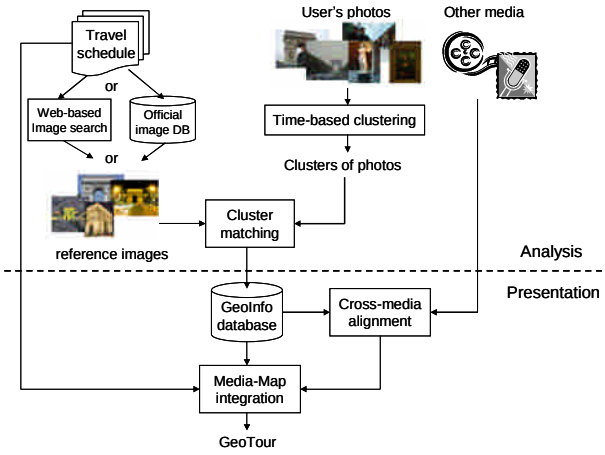

Figure 8. The system framework of generating GeoTour.

Figure 9. An example of a GeoTour in global scenario.

Instead of using a static map, we can also integrate the Google Earth service to show a "global" tour. The visited scenic spots can be labeled on the interface, and the touring process can be recorded as a video clip. Figure 10 presents some video frames captured from Google Earth's My Tour functionality.

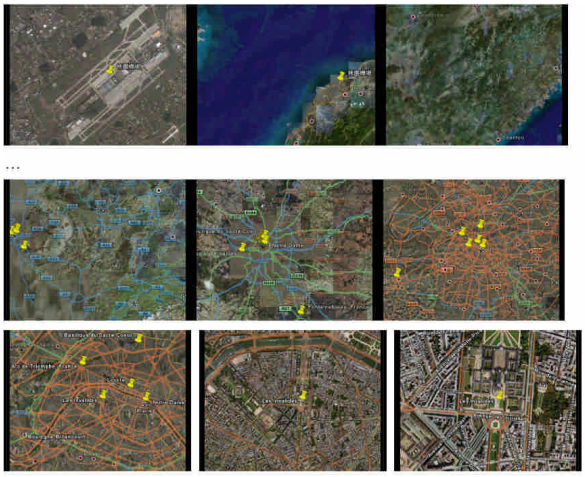

Figure 10. An example of a GeoTour in the global scenario that integrates Google Earth.

4.1.4 GeoTour in the Local Scenario

In the local case, the reference images are provided by tourism administrative offices or related organizations. Furthermore, we can directly use the official map for integrated presentation. Figure 11 shows an example of a GeoTour in the NTU campus.

Figure 11. An example of a GeoTour in local scenario.

A major difference between global and local scenarios is that there may be fixed touring paths within a scenic spot, such as the red routes shown in Figure 11. If we can identify the photos taken at distinguished buildings, photos that were taken temporally in-between photo clusters corresponding to two buildings can be located by interpolation calculation. In this way, we can handle the photos that were taken when a tourist moves from one location to another, and reconstruct personal tour on the official map to a more detailed degree. This interesting presentation style may bring higher commercial values when promoting a scenic spot. After a tourist's traveling this scenic spot, he can just put his camera memory card into a machine and then a customized personal-tour recording can be generated. This customized tour shows how he moved in this scenic spot and conveys information about the geographic information about each photo.

4.2 Other Presentation

Actually, there are growing interests in using geo-information to improve photo browsing. In [11], the WWWX exploits GPS information and arranges images on an interactive 2D map. In [12], the World Explorer shows the aggregate knowledge based on user-tagged Flickr images in the form of representative tags for arbitrary areas in the world. In [13], they present a system for interactively browsing and exploring photos of a scene using a novel 3D interface. We have to notice that the proposed approach is especially suitable to personal media management, especially for the people who are lazy or have no time to organize their travel experience. The proposed method can be integrated to the referred works.

In addition to geography-based presentation, results of automatic image annotation can also facilitate other kind of experience management or presentation. One promising application is travel blog generation [14]. The proposed approach automatically locates user's photos and integrates the materials captured in traveling to generate a blog. Moreover, location information can also be integrated into an on-line album or a slideshow [15]. Many variations can be made for vivid presentation of the travel experience.

5. EXPERIMENTAL RESULTS

5.1 Results of Cluster Matching

Here, we primarily demonstrate the cluster matching results in the global scenario. Table 1 shows the ground-truth information about the evaluation data captured by three amateurs in different tours. Five, seven, and thirteen scenic spots were visited in these three tours, respectively. Figure 12 shows some sample photos in the evaluation data set.

Table 2 shows the accuracy of cluster matching. The accuracy values n_1/n_2 represent that n_2 clusters of user's photos are identified as the visited spots, and among them, n_1 clusters are correctly identified. Without the helps of GPS information and well-developed image databases, we effectively identify where photos were taken.

Table 1. Information of the evaluation data.

	Sydney	Paris	New York City
# photos	120	58	155
# visited spots	5	7	13
# image clusters	17	18	27
Spots	Queen Victoria Building, Sydney Opera House, Sydney Harbour Bridge, Royal Botanic Garden, …	Les Invalides, Arc de Triomphe, Notre Dame, Palace of Versailles, Louvre, …	Time Square, Central Park, Metropolitan Museum, Statue of Liberty, Brooklyn Bridge, United Nations, …

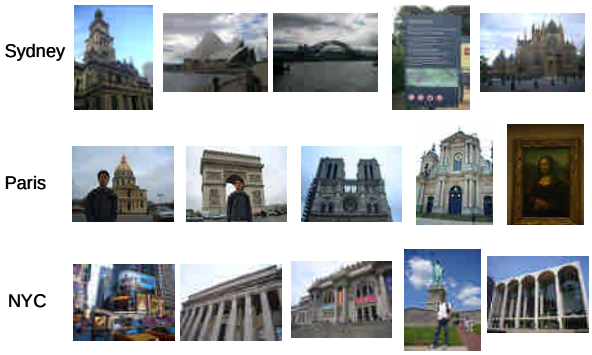

Figure 12. Some sample photos in the evaluation data set.

Table 2. Accuracy of cluster matching.

	Sydney	Paris	New York City
Accuracy	3/5	7/7	12/13

Apparently, the annotation result in the Sydney dataset is worse than others. Therefore, a failure case is shown in Figure 13 to illustrate the poor performance. Note that though the user visited the Sydney Opera House according his schedule, the viewing angles of his photos are significantly different from the canonical ones retrieved from the web.

Table 2 shows the matching results with the lightweight filtering process described in Section 3.4. Because we strictly set the constraint for filtering, it's very rare that really matched image pairs are erroneously filtered out. In the reported results, the accuracy of matching results with and without lightweight filtering are the same, while we save 1/4 matching time with filtering.

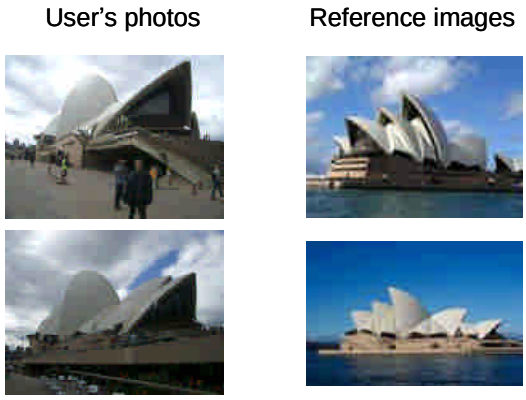

Figure 13. A failure case in matching.

5.2 Relationship between Reference Images and User's Photos

We also studied the relationship between the number of reference images and the number of correctly annotated images. We use a dataset of 117 photos, in which 13 distinguished buildings were visited, and change the ratio of the number of reference images to the number of testing images. Figure 14 shows the relationship. We can see that increasing the number of reference images does not significantly improve the number of correct recognition. In other words, the proposed approach can work well even only a few reference images in canonical views are utilized.

Note that photos categorized as in same cluster with the recognized image are also labeled by the recognized spot names. Therefore, although it seems that small ratio of photos are "exactly matched" with the reference image, most photos can be successfully labeled due to clustering and annotation propagation.

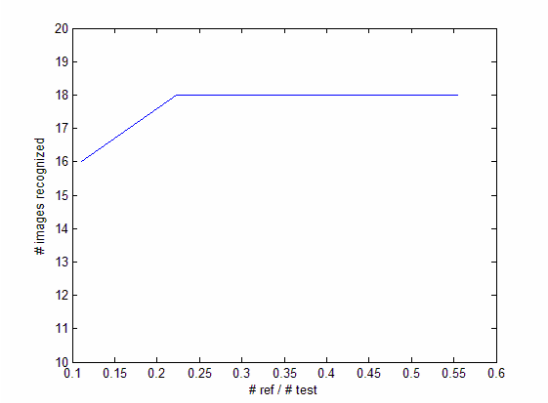

Figure 14. The relationship between the amount of correct recognition and the ratio between reference and test images.

6. DISCUSSION

We discuss the limitations of the current approach and propose some issues needed to be studied more. The first issue is the shortage of SIFT-based image matching. Although SIFT is invariant to scale and small viewing angle variations, objects that have few apparent and unique corner information, such as mountains, sky and beaches, are hard to be characterized. Therefore, only better matching performance in cityscape or artificial objects can be obtained. To tackle with this shortage, we may further integrate statistics-based concept detectors with the travel schedule. For example, although it's hard to match "Luzern Lake" based on SIFT features, we can apply a lake detector to the photos taken in a tour of Switzerland and annotate those with a lake with the specific spot name "Luzern Lake".

The second issue is that we often assign multiple annotations to a photo. The photos that were taken in Eiffel Tower or Arc de Triumph can also be labeled as Paris. The resolution issue in presentation and annotation will be addressed in the future.

7. CONCLUSION

In this paper, we implement a prototype system to show how we annotate and integrate media captured in travels. Simple approaches are exploited to perform annotation propagation and integrated presentation. Actually, travel experience can be maintained more elaborately if we discuss more about cross-media synchronization or automatic grabbing relevant information from other resources to enhance users' personal experience.

8. REFERENCES

[1] Flickr, http://www.flickr.com

[2] Picasa, http://picasa.google.com/

[3] Yeh, T., Tollmar, K., and Darrell, T. Searching the web with mobile images for location recognition. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, 76-81.

[4] Wang, C., Jing, F., Zhang, L., and Zhang, H.-J. Scalable search-based image annotation of personal images. In *Proceedings of ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2006, 269-277.

[5] Wang, X.-J., Zhang, L., Jing, F., and Ma, W.-Y. AnnoSearch: image auto-annotation by search. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, 1483-1490.

[6] Platt, J.C., Czerwinski, M., Field, B.A. PhotoTOC: automating clustering for browsing personal photographs. In *Proceedings of IEEE Pacific Rim Conference on Multimedia*, pp. 6-10, 2003.

[7] Cheng, P.-J., and Chien, L.-F. Effective image annotation for searches using multilevel semantics. *International Journal of Digital Library*, vol. 4, 2004, 258-271.

[8] Joshi, D., Wang, J.Z., and Li, J. The story picturing engine - a system for automatic text illustration. *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, 2006, 68–89.

[9] Lowe, D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 20, 2003, 91-110.

[10] GeoNames, http://www.geonames.org/

[11] Toyama, K., Logan, R., Roseway, A., Anandan, P. Geographic location tags on digital images. In *Proceedings of ACM Multimedia*, 2003, 156-166.

[12] Ahern, S., Naaman, M., Nair, R., Yang, J. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In *Proceedings of International Conference on Digital Libraries*, 2007, 1-10.

[13] Snavely, N., Seitz, S.M., and Szeliski, R. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, vol. 25, no. 3, 2006, 835-846.

[14] Cemerlang, P., Lim, J.-H., You, Y., Zhang, J., and Chevallet, J.-P. Towards automatic mobile blogging. In *Proceedings of IEEE International Conference on Multimedia and Expo*, 2006, 2033-2036.

[15] Chen, J.-C., Chu, W.-T., Kuo, J.-H., Weng, C.-Y., and Wu, J.-L. Tiling slideshow. In *Proceedings of ACM Multimedia Conference*, 2006, 25-34.

[16] SIFT++: A lightweight C++ implementation of SIFT, http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html