Visual Summarization for Place-of-Interest by Social-Contextual Constrained Geo-clustering

Yayun Ren, Xueming Qian, Shuhui Jiang

School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China

renyyy@stu.xjtu.edu.cn; qianxm@mail.xjtu.edu.cn; shjiang0531@gmail.com

***Abstract*—With the rapid development of social networks, more and more users choose to share their own photos with their friends. Especially, users prefer to share the photos they took during traveling, thus there emerges many user generated content for place-of-interests (POIs). So based on the user contributed photos, we can summarize each POI by mining location-of-interest (LOI, which represents the attractive viewpoints of POI) and selecting some representative images from them. It is important for scheduling a traveling, and in this paper, an effective POI summarization approach is proposed by an improved geo-clustering with visual and views verification, which helps us to have a representative and comprehensive perception for POI. In our approach, we firstly collect POI related photos from social media, and filter the raw data by the combination of tags and geo-locations. Secondly, we mine LOIs for each POI by the improved geo-location clustering method. Finally, we employ visual and views verification to select images from LOIs to summarize the POI. We conduct a series of experiments based on Flcoickr dataset. Experimental results demonstrate the effectiveness of our proposed method.**

I. INTRODUCTION

With the appearance of Web 2.0, more and more users take varieties of activities by means of network. Some are more willing to share their lives on the networks, including some pictures taken during their travels. Taking Flickr as an example, after launching its geo-tagging interface in 2006, the number of photos uploaded on Flickr has increased to more than 6 billion. These photos have supplementary information, e.g., tags, geo-tags and views (times that the picture has been viewed by other users), which can not only facilitate the big media management but also give us a wealth of resources to carry out travel planning. In Fig. 1, we give an example of Flickr image.

Passengers planning to travel often browse some representative images about different place-of-interests provided in some websites and then determine where to go. The representative images are convenient for the candidate visitors to know more details about the POIs. If we can well find representative attractions of a given POI from the crowded community contributed pictures, it will have a beneficial effect on our tourism. It will also save our time to obtain the first-hand materials of the POI. In recent years, much work aim to mine POI worldwide based on these

2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), Oct 19-21, 2015, Xiamen, China.
978-1-4673-7478-1/15/$31.00 ©2015 IEEE

Fig. 1. An example of Flickr image. This picture was shot at 27.217998°N, 77.98645°E in Taj Mahal, and it had already been viewed 205 times after it was uploaded.

metadata. Some of the exiting works use a variety of methods to carry out representative pictures recommendation [1-2],[8], and the corresponding travel recommendations [6],[9-10].

In general, for a POI, if a viewpoint in it is attractive, there will be more photos on the web uploaded by people who have visited it in the past. We also call viewpoint of POI as a location-of-interest (LOI). The LOI is a subset of a POI.

We aim to mine LOIs from the community-contributed photos and their metadata, simultaneously considering the different views of pictures. However, there are mainly three challenges in visual summarization for POI: 1) How to find candidate LOIs from the large volume of the images with noisy tags and geographical locations. 2) How to determine suitable LOIs using multimodal metadata information from social media. 3) How to generate representative images for a given LOI so as to summarize the POI comprehensively.

To solve the challenges mentioned above, we propose an improved geo-location clustering with visual and views verification method based POI summarization approach. The main contributions of this paper are summarized as follows:

- We propose a social user weighting approach based on the traditional mean shift algorithm to mine LOI. It can embody the contribution of worldwide users in LOI mining.
- We also introduce the weight of views, which makes pictures with more click times have more opportunities to be selected during POI summarization.
- We fuse the visual and views information in visual summarization, which improves the performance of the recommended images both on relevance and diversity.

The rest of paper is organized as follows: Related work on visual summarization and representative images generation are introduced in Section II. In Section III, "Place-Of-Interest Dataset Filtering", "Location-Of-Interest Mining" and "Visual Summarization" are described in details. The comparison experiments of our approach and discussion are given in Section IV. Finally, we conclude and discuss future work in Section V.

II. RELATED WORK

The most relevant work is representative image recommendation. And many research efforts have been dedicated to representative image recommendation. Many of them [2], [4], [6], [8] consider both visual and text features to solve their problems. For instance, Lyndon et al. [2] used content-based methods to choose diverse and representative images for a given landmark. Their methods focused on statistics, and have got a certain effect. However, statistical method needed large enough accurate data. Xue et al. [4] modelled an image's viewpoint in horizontal, vertical, scale and orientation aspects, and then, they used 4-D vectors to construct the viewpoint vectors for each image. They selected Identical Semantic Points (ISPs) from SIFT points of the image to capture major and unique parts of a landmark. Simon et al. [6] used multi-user collections from the Internet to construct scene summarization. Given a set of images with the specified tags, they generated canonical views by clustering of images visual properties, and extracted representative tags for each cluster. Chu et al. [8] proposed a novel method to automatically select representative photos by feature classification. They mainly used the local features to distinguish artificial objects and natural objects, and chose the near-duplicate parts as the representative landmark.

Community-contributed photos offer GPS information. This is also important source for representative image selection [1],[3],[7],[9],[10]. Rudinac et al. [3] found representative images by modelling multimodal graph from visual features, text associated with the photos, users' information and their social networks. Hays et al. [7] estimated the geographic location of a single image by searching for those visually similar samples in the given dataset in a nearest fashion. Zheng et al. [9] built a landmark recognition engine by modelling and recognizing world-scale level landmarks. They only showed strong visual relevance between the traditional representative views of landmarks. Jiang et al. [1],[10] put forward a new method, using mean shift method to mine High Frequency Shooting Locations (HFSLs) for the landmark. They then sorted the SIFT matching pairs of pictures at these HFSLs, and chose the representative pictures. Their work is the most relevant to ours, but they have not considered social users' shooting behaviours and the meaning of different views of pictures.

Based upon the above review of related work, a new method to summarize POI by LOI mining is proposed. Since many works have not taken the impact of users and views into account, we introduce two weights - user weight and views weight on the basis of traditional mean shift, which helps to reduce the proportion of pictures of users who have uploaded many similar pictures, and to increase the probability of more viewed pictures in the clustering process. That also facilitates us in the subsequent representative images selection. Thus, the final selected pictures are constructed to POI summarization.

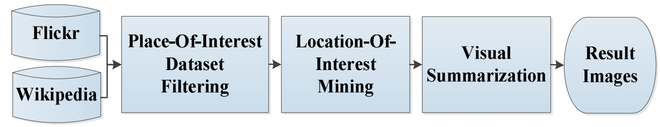

Fig. 2. Flow chart of the proposed visual summarization by location-of-interest mining.

III. APPROACH

Flickr[1] is a well-known image sharing and online album website. There are many beautiful photos about POIs taken by world-wide users, with supplementary information, e.g., tags, geo-tags and views. Thus from the community-contributed photos, we can summarize the POI by selecting some representative images.

The framework of our approach is shown in Fig. 2. The inputs of our system are the tag and official location of a POI. It consists of the following 3 steps: 1) Crawling geo-tagged images from social media with the user contributed information (tags) and dataset filtering to remove noisy images with the help of official location. 2) Carrying out location-of-interest mining for the POI via improved clustering on the crawled image set, taking social users shooting behaviours and different views into account. And 3) carrying out visual summarization with visual and views verification intra LOIs to select representative images for each LOI, and use diversity comparing inter LOIs to choose diverse images for summarizing a POI. Hereinafter we turn to introduce our approach in detail.

A. Place-of-Interest Dataset Filtering

In recent years, more and more people use their smartphones with the GPS positioning function to take pictures, and it is easier for them to share the photos with their friends. From the community contributed photos crawled from Flickr, we can summarize the POI by selecting some representative images. Moreover, Wikipedia has edited the official geographic locations of POIs, which is very convenient for us to recognize the places of the shared photos by consulting them on Wikipedia[2]. Thus we collect images from Flickr and corresponding location information from Wikipedia.

Due to the fact that in the large scale social user contributed photos, many of them are irrelevant to the target POI. Only from the tags, we cannot select appropriate images to summarize the POI [1], [2], [10]. So, we need to filter the crawled dataset to remove irrelevant images. Our POI filtering consists of the following steps:

1) Tag Based Image Filtering: Given a place-of-interest, first, we use the POI related tags, such as "Big Ben", to find photos associated with the input tags. However, it should be pointed out that there is no limit about matching length of tag, some pictures with the tag "big bend" will also be selected. These pictures have no correlation with "Big Ben". Moreover, the photos of a person whose name is "Big Ben" which appeared in the surrounding texts of he shared photos will also

[1] http://www.Flickr.com/
[2] http://en.wikipedia.org/

TABLE 1
OFFICIAL LOCATION OF TEN POIS AND PICTURE NUMBERS AFTER DIFFERENT FILTERING STAGES

POI	Official Location	Picture Numbers after Filtering	
		Tag Based	Location Based
#1 Tower Bridge	51°30'20"N 0°04'32"W	622	356
#2 St Paul's Cathedral	51°30'49"N 0°5'53" W	6097	457
#3 Eiffel Tower	48°51'29.6"N 2°17'40.2"E	1,785	1,048
#4 Angkor	13°26'N 103°50'E	4,287	1,243
#5 Big Ben	51°30'2.72"N 00°07'28.78"W	1,542	451
#6 Cologne Cathedral	50°56'29"N 6°57'29"E	4,328	1,994
#7 Colosseum	41°53'24.61"N 12°29'32.17"E	624	279
#8 Golden Gate Bridge	37°49'11"N 122°28'43"W	1,815	1026
#9 Statue of Liberty	40°41'21"N 74°2'40"W	929	372
#10 Taj Mahal	27°10'30"N 78°02'31"E	1,042	294

in our dataset. We need further processing in next step.

2) Location Based Image Filtering: We further remove pictures without GPS information and we only retain pictures containing valid latitude and longitude information. We utilize the official geographical position of the POI from Wikipedia to constrain the tag filtered images. Only the photos with their geo-tags who are near to the POI are retained.

In TABLE 1, we give the official location of ten POIs used in our experiment and the picture numbers after different filtering stages.

B. Location-of-Interest Mining

Once we have obtained a set of filtered POI dataset, we turn to mine LOIs by a geo-clustering approach fusing with social user factor and pictures' views factor. Thus, in this section, we firstly introduce the traditional mean shift method. Then illustrate the new proposed improved mean shift clustering algorithm to mine LOIs.

1) Traditional Mean Shift Algorithm

The mean shift algorithm is often utilized to group geo-tagged photos into clusters, from which representative images can be selected [1], [6-8].

We assume that the number of geo-tagged images after filtering is *N*. The location information of each picture is denoted as $x_i (i = 1,2,\cdots,N)$. It is a 2-dimenional vector including latitude and longitude. Each x_i can be treated as a point in two-dimensional space. The man-shift algorithm consists of the following steps:

Firstly, we randomly select an initial starting point x and set the bandwidth h, which is set empirically. Then we calculate the mean shift vector $M_h(x)$ expressed as follows:

$$\begin{cases} M_h(x) = \dfrac{1}{k}\sum_{x_i \in S_h}(x_i - x) \\ S_h \equiv \{x_i : (x_i - x)^T (x_i - x) \le h^2\} \end{cases} \quad (1)$$

where S_h is the circle whose radius is h and centroid is x, and k is the number of points that falling in the region S_h .

Secondly, we update the starting point as follows: $x \leftarrow x + M_h(x)$. Continue to update the starting point until traversing all of the pictures.

2) Improved Mean Shift Algorithm

a) User Weight

In traditional mean shift clustering based interesting viewpoints detection, all images are viewed equally important. However, it does not take the social user factor in social media into account. Different users have different shooting styles and sharing behaviours. For example, in Fig. 3, we list some photos taken by user A (whose total uploading pictures number is 58) and user B (whose total number is 5) in Golden Gate Bridge, San Francisco, USA.

If a user A upload one thousand images and another user B only share 10 images at the same place, then the centroid determined by traditional mean shift based clustering will be biased to user A. To overcome this, we cannot simply consider that the importance of all of the pictures is equal. A user weighting based approach can be utilized to make a balance between users. We defined user weight $W_{u_{x_i}}$ of picture x_i as follows:

$$W_{u_{x_i}} = \frac{1}{\sqrt{N_u}} \quad (2)$$

where, $N_u$denotes the total number of pictures upload by user u(who has uploaded picture x_i). We have calculated the photo numbers uploaded by different users, which varies from 1 to a few hundred, so we use its square root $\sqrt{N_u}$ to represent the user's weight. And from our assumption, a bigger N_u indicates less diversity of the user uploading photos, so we use inverse of $\sqrt{N_u}$. $W_{u_{x_i}}$is then normalized to (0,1]. Then a user based mean shift vector is as follows:

$$M_h(x) = \frac{1}{k}\sum_{x_i \in S_h} W_{u_{x_i}}(x_i - x) \quad (3)$$

From (3), we find that when users upload only one picture,

(a). Pictures Uploaded by User A.

(b). Pictures Uploaded by User B.

Fig. 3. Pictures Uploaded by User A and B. In (a), photos in the first row and third row are shot in the same position, and their differences are very small. Meanwhile, among 5 pictures shown in (b), each has different perspectives.

i.e. $W_{u_{x_i}} = 1$, then, (3) is identical to (1).

b) Views Weight

In Flickr, the views of photos are also provided as shown in Fig. 1. Views of different pictures vary dramatically from 0 to more than 10,000, showing the popularities of pictures to some extent. Since people are more willing to click to see some of the pictures, we also take views into account during LOI mining and representative images selection. For each image, we use the log of its viewing times to represent its weight $W_{v_{x_i}}$. In order to avoid the problem caused by views are 0, we give a definition as follow:

$$W_{v_{x_i}} = log\left(V_{x_i} + 1\right) \quad (4)$$

Then, we quantify the views weight of each picture to [0, 1]. After this step, the clustering centers will not shift fiercely with the change of views. The views based mean shift vector is as follow:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} W_{v_{x_i}} (x_i - x) \quad (5)$$

From (5), if the views are 0, (5) is identical to (1). That is to say, we only need to consider the original probability density distribution and user factor under this circumstance.

c) Factor Fusing

By considering both the user factor and views factor in our geo-clustering, we simply combine the two factors by linear summarization. We can rewrite (1) as follows:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (W_{u_{x_i}} + W_{v_{x_i}})(x_i - x) \quad (6)$$

Thus, our proposed mean shift based geo-clustering approach does not just considering the position distribution of the pictures as the traditional mean shift algorithm. It highlights the importance of views, which is carried by pictures themselves. At the same time, it avoids that most of the cluster centers are shifted to users uploading too many pictures.

We summarize the latitude and longitude information of all the pictures in the POI dataset, and take it as the input of improved mean shift clustering. In this paper, we set the bandwidth as 0.001, which is roughly 100 meters. This is because 1° of latitude is equivalent to about 111 km on Earth's surface. Under this bandwidth, we can get a lot of clusters, which are called LOIs.

Note that, after improved mean shift clustering, in some clusters, there are only one or two photos. These clusters are more likely to be noises rather than LOIs, so we just delete them by a picture threshold and a user threshold. Only the clusters whose numbers of photos and users are higher than certain thresholds are chosen as candidate LOIs. So we obtain many different LOIs for each POI.

C. Visual Summarization

In this section, we aim to summarize POIs with a set of representative images and with sufficient diversity. We take both the intra LOIs visual clustering and inter LOIs ranking into account for visual summarization. It consists of the following two steps: 1) Intra LOIs, we choose different numbers of representative images using topical album [4], and compare the views of candidate images, we choose the more views ones to summarize each LOI. 2) Inter LOIs, we select LOIs by diversity comparing to make sure the selected representative images are with abundant diversity.

1) Intra LOIs

There exist diverse viewpoints in the mined LOIs, which is because when people are shooting pictures, they may change their focal distances and shoot pictures from different angles. Thus, pictures shot in the same location may include different viewpoints. Clustering of geo-location can help find the interesting locations, and visual verification in the obtained LOIs is also unreasonable. Here, we employ visual classification to classify pictures in LOIs, and then use views to choose the most representative ones. Assuming that, we got M LOIs, expressed as $L_m (m = 1,2, \cdots, M)$. The number of images in LOI L_m is N_m.

a) Visual Gathering

After POI dataset filtering and LOI mining, images in the LOIs are relatively similar, so we use local features to implement the next analysis. Here, we use 128-dimension Scale Invariant Feature Transform (SIFT) feature. Each of the 128-dimension SIFT descriptors of an image is quantized to a visual vocabulary by hierarchical quantization [4]. In this paper, each picture is represented by 10000-dimensional BOW histogram.

We use the method proposed in [4] to partition images in LOI $L_m (m = 1,2, \cdots, M)$ to viewpoint albums, which are collections of some same pictures. Thus, we obtain many viewpoint albums for each LOI.

b) Views Re-ranking

Through visual gathering, pictures in the same viewpoint album are visually alike. It's likely that the more the views, the more popular the pictures will be. To make the top ranked summarization results with high visual diversity, we iteratively select one reprehensive image from each viewpoint album. We rank the candidate images by their views. Images with the highest views will be chosen to represent the viewpoint album. In these candidate images, we choose the highest views to represent the LOI.

2) Inter LOIs

After choosing representative images for each LOI, we turn to rank these images inter different LOIs iteratively.

Each LOI $L_m (m = 1,2, \cdots, M)$ has its center, which is expressed as a two-dimensional vector (C_x, C_y). C_x presents latitude and C_y presents longitude. For the *m-th* LOI, whose centre is (C_{xm}, C_{ym}), we calculate the Euclidean distance from each LOI to the official location as (7). The official location of a POI is represented as (C_{xo}, C_{yo}), which is crawled from Wikipedia. Then we rank LOIs according to their geographical distance $D(m, o), m = 1,2, \cdots, M$.

$$D(m, o) = \sqrt{(C_{xm} - C_{xo})^2 + (C_{ym} - C_{yo})^2} \quad (7)$$

Firstly, we select a LOI which is nearest to the official location. We put this LOI into the ranked sequence (*RS*), which is initialized to be the LOI whose centroid is closest to the official location. We also remove it from the completed LOI set $\Omega_L = \{L_m; m = 1,2, \dots, M\}$. Then we find a LOI L_c that has the largest distances from all of the ranked LOIs (L_k) in RS as follows:

$$max \sum_{k \in RS} D(c, k),\ c \in \Omega_L - RS \qquad (8)$$

where $D(c, k)$ is the Euclidean distance between LOI L_c and L_k. Then we push c into the ranked sequence RS and remove c from Ω_L, i.e. RS=RS+c, and $\Omega_L = \Omega_L - c$. This process continues until Ω_L is empty. Therefore, we can get a ranked sequence RS of LOIs. And all of the LOIs' representative images construct the visual summarization of POI.

IV. EXPERIMENTS

In this section, we conduct a series of experiments to evaluate the performance of proposed visual summarization method for POIs, and compare it with the existing approaches.

A. Datasets

There are many sources of community-contributed photos on the web. We collect near 7 million Flickr photos by open API. These photos are uploaded by 7,387 users and the heterogeneous metadata are associated with the photos. We also collect the official geographic location of these relevant POIs from Wikipedia.

We select 10 popular POIs to evaluate the performance of our method. They are as follows:#1 Tower Bridge, #2 St Paul's Cathedral, #3 Eiffel Tower, #4 Angkor, #5 Big Ben, #6 Cologne Cathedral, #7 Colosseum, #8 Golden Gate Bridge, #9 Statue of Liberty, and #10 Taj Mahal.

B. Performance Measures

We invite 20 volunteers to evaluate the summarization for each of the 10 POIs. And the volunteers are familiar with the POIs. We invite them to evaluate the results from two aspects as[1] and [4]:

Relevance: Can the summarization well represent this POI (0-10)?

Diversity: Are the differences between the result pictures large enough (0-10)?

C. Evaluation

To show the effectiveness of the proposed landmark summarization, we make comprehensive comparisons of our approach (denoted as OURS) and Canonical Views (denoted as CV) [5], Clustering, Ranking and Ranking (denoted as CRR) [2], Identical Semantic Points (denoted as ISP) [4], and High Frequency Shooting Location (denoted as HFSL) [1]. We show the performance by the criteria of "Relevance" and "Diversity" in Fig. 4. We also give the average value of the two performances in the last column of the histogram, named "avg". From Fig. 4, we can note that the results of five methods are sufficiently representative, while they also have different characteristics. Generally speaking, for the majority of ten POIs in terms of performance, OURS, HFSL and CRR achieved good relevance performance because these approaches summarize images according to image content overlap. CV works well with diversity, since it penalizes canonical views for being too similar. ISP works well with diversity, because it models an image's viewpoint in horizontal, vertical, scale and orientation aspects. OURS gets

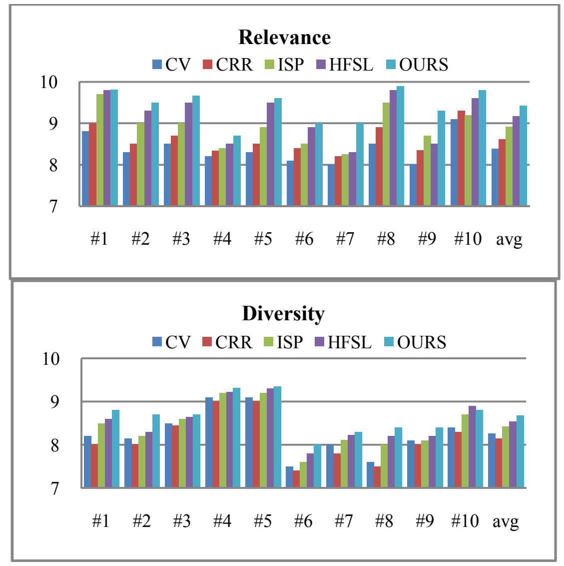

Fig. 4. Performance comparison for the criteria of "Relevance" and "Diversity" of the methods: CV, CRR, ISP, HFSL and OURS on the ten POIs. The y-axis is the score and the x-axis is the POI.

much higher scores than the other four methods in Relevance, Diversity. In terms of "avg", our method has achieved the highest score.

Compared to other methods, our approach considers the visual relationship between the pictures in different LOIs. We introduce visual verification, so the dug LOIs, pictures taken in the same LOIs are similar enough and pictures shot in different LOIs are diverse enough. Due to space limitation, we give some representative pictures of five POIs in Fig. 5. Our method is relatively good. For Colosseum, because it is a circular building, the results of the three approaches are similar. And OURS works better in diversity. For Taj Mahal, OURS gives three pictures of the mosque besides the mausoleum located in the center of the POI. We have got a comprehensive sensory of Taj Mahal.

Moreover, after comparing views of different results, pictures obtained by our method have been viewed for many times, while others have been viewed less.

D. Discussion of the Impact of User and Views Weights

We discuss the impact of the user and views weights in the improved mean shift clustering in this section. Here, we use four different methods to mine LOIs, traditional mean shift (denoted as **MS**), only adding user weight (denoted as **MS_u**), only adding views weight (denoted as **MS_v**), and adding both user and views weights (denoted as **MS_uv**).

We also invited the volunteers to evaluate the final summarization results of these four methods, considering the relevance, diversity of recommended pictures. And calculate the average scores for ten POIs, as shown in Fig. 6. Comparison of four methods, after adding user and views weight, the recommended results tend to be better than traditional mean shift clustering.

For most of the POIs, e.g., POI #2,#8,#9, summarization results have been significantly improved, and achieved the highest score when the two weights are all used. While, for

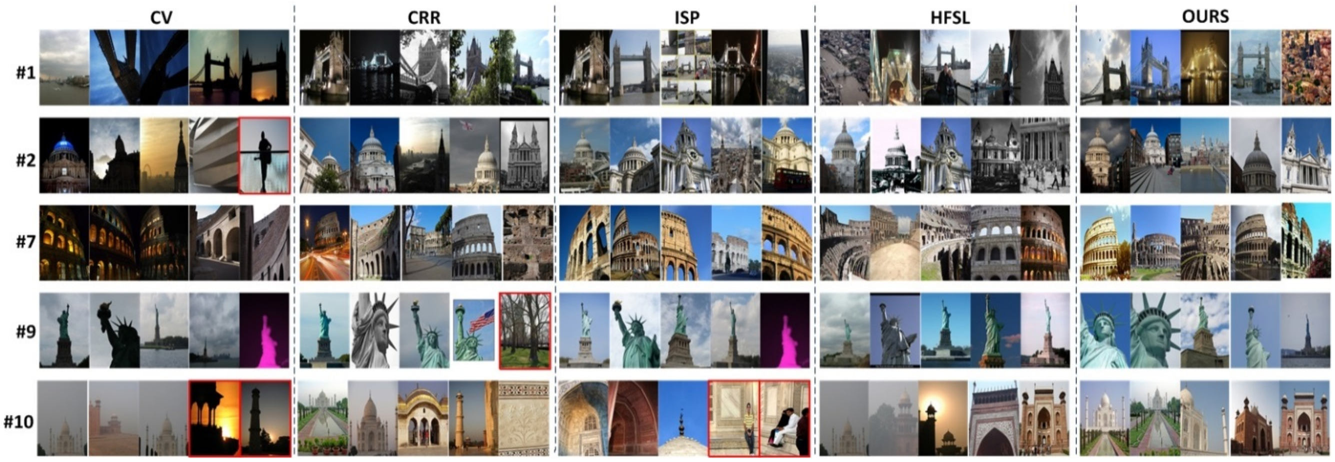

Fig. 5. Five results of five POIs by CV, CRR, ISP, HFSL and OURS. From left to right are the results of CV, CRR, ISP, HFSL and OURS. Images in red frames are irrelevant to the POIs.

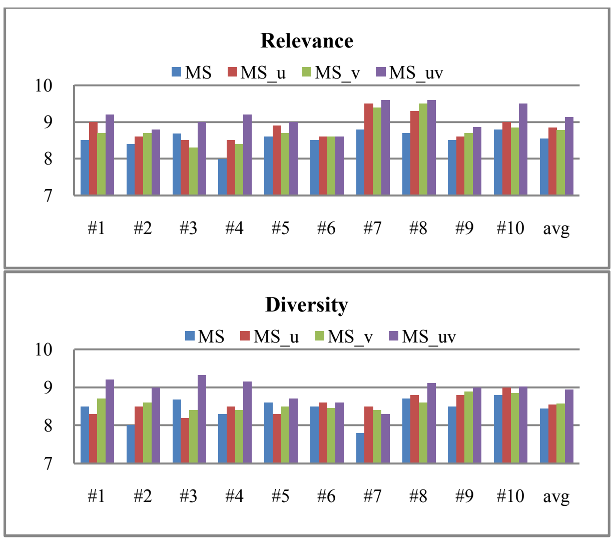

Fig. 6. Performance comparison for the criteria of "Relevance" and "Diversity" using four different methods for geo-location clustering. The y-axis is the score and the x-axis is the POI.

the #1 Tower Bridge, and #7 Colosseum, when views weight is added, they exhibit poorer performance. This is because the merely using of views weight for clustering, some pictures with human portraits but have being viewed for many times are selected. For other POIs, after adding user and views weight, the recommended results are more reasonable and convinced. We compare the views of each picture in the visual summarization results, pictures in our result are viewed more times than the other two methods. It should be pointed out that, views of pictures are linked to their upload time. However, the more the views are, the more popular the results will be.

V. Conclusion

In this paper, we proposed a new way to generate visual summarization for POI via LOI mining. We used tag based and location based filtering to construct candidate POI dataset. We introduced user weight and view weight in the traditional mean shift algorithm to mine LOIs from Flickr data. In this way, photos uploaded by different users are treated discriminately, and photos with many views played more important roles in LOI mining. We also applied inter and intra LOIs analysis to perform visual summarization, offering more relevant and diverse results. Our approach mainly analysis images in given POIs, if expanded to the worldwide scale, it is also applicable. In the future, we will use more cross-platform data to complete the personal travel recommendation.

Acknowledgment

This work is supported in part by the Program 973 No.2012CB316400, by NSFC No.60903121, 61173109, 61332018, and Microsoft Research Asia.

References

[1] S. Jiang, X. Qian, "Generating representative images for landmark by discovering high frequency shooting locations from community-contributed photos," *IEEE International Conference on Multimedia & Expo*, pp.1-6, 2013.

[2] L. Kennedy, M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proceedings of the 17th int. conf. on World Wide Web*, pp. 297-306, 2008.

[3] S. Rudinac, A. Hanjalic, and M. Larson, "Finding representative and diverse community contributed images to create visual summaries of geographic areas," in *Proc. of the 19th ACM int. conf. on Multimedia*, pp.1109-1112, 2011.

[4] X. Qian, Y. Xue, X. Yang, Y.Y.Tang, X. Hou, and T. Mei, "Visual summarization of landmarks via viewpoint modeling," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.PP, issue. 99, pp. 1-13, 2014.

[5] I. Simon, N. Snavely, S.M. Seitz, "Scene summarization for online image collections," *IEEE Int. Conf. on Computer Vision*, pp.1-8, 2007.

[6] L. Cao, J. Luo, A. Gallagher, T. Huang, "A worldwide tourism recommendation system based on geotagged web photos," *IEEE Int. Conf. on Acoustic Speech and Signal Processing*, pp.2274-2277, 2010.

[7] J. Hays and A. A. Efros, "Im2gps: estimating geographic information from a single image," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1-8,2008.

[8] W. Chu, C. Lin, and J. Yu, "Feature classification for representative photo selection," in *Proc. of the 17th ACM int. conf. on Multimedia*, pp.509-512, 2009.

[9] Y. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T. Chua, and H. Neven, "Tour the world: building a web-scale landmark recognition engine," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[10] S. Jiang, X. Qian, J. Shen, and T. Mei, "Travel recommendation via author topic model based collaborative filtering," *IEEE Trans. Multimedia*, vol.17, no.6, pp.907-918, 2015.