Automatic Summarization of Travel Photos Using Near-Duplication Detection and Feature Filtering

Wei-Ta Chu
Dept. of CSIE
National Chung Cheng University, Taiwan

wtchu@cs.ccu.edu.tw

Chia-Hung Lin
Dept. of CSIE
National Chung Cheng University, Taiwan

lchu96m@cs.ccu.edu.tw

ABSTRACT

We try to address part of the challenge proposed by CeWe. A photo summarization method is developed to select representative photos. Based on the observation that the most important objects/views are often captured several times, we exploit near-duplicate detection techniques to represent a sequence of photo as a graph, and then the graph structure is analyzed to facilitate importance ranking. We focus on summarizing hundreds of photos taken in journeys lasting for one or two weeks. The qualitative and quantitative measurement results demonstrate the effectiveness of the proposed method.

Categories and Subject Descriptors

I.5.2 [**Pattern Recognition**]: Design Methodology – *classifier design and evaluation, feature evaluation and selection.* H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *abstracting methods.*

General Terms

Algorithms, Measurement, Experimentation.

Keywords

Feature classification, representative selection, photo summarization.

1. INTRODUCTION

Travel is not only a good way to relax tense life, but also a symbol of vogue in the modern age. With the popularity of consumer appliances, large volume of multimedia content can be easily captured in journeys and uploaded to social community platforms for sharing. Massive content burdens users in browsing and managing, and therefore an effective system that is able to summarize a journey is urgently needed.

In this work, we focus on finding compact presentation of personal photo collections captured in journeys. In a journey, we observe that the important objects, such as famous landmarks or canonical views of a scenic spot, are often captured several times. Therefore, we propose a representative selection method based on near-duplicate detection (NDD) techniques and demonstrate promising results [1]. Scale-invariant feature transform (SIFT) features play the most important role in near-duplicate detection. However, different types of SIFT features pose different influences on NDD. Recently, we further classify SIFT features

into that on artificial objects, which impact NDD more, or in natural scenes. We effectively find representative presentation of a journey. An online demo is available in http://mclab.cs.ccu.edu.tw/research/MMchallenge2009.

2. FRAMEWORK

Figure 1 shows the system framework. Given a photo collection taken in a journey, say two-week tour in Paris, we first segment the collection into smaller clusters, each of which contain the photos taken in the same scenic spot. This segmentation can be easily conducted because we can check the time gaps between temporally adjacent photos, and dynamically determine the boundaries of different scenes.

For a cluster of photos, prominent feature points are extracted by the difference of Gaussian (DoG) detector and are described by SIFT descriptor. For human beings, near-duplication is claimed to occur mainly because there are similar artificial objects between photos. Therefore, we especially classify SIFT features into that on artificial objects or in natural scenes to differentiate their influences on near-duplication detection [3].

After examining near-duplicate relationships between photos, we represent the relationships into a graph, in which each node denotes a photo, and an edge linking two nodes represent that these two photos are detected as near-duplicate pair. Based on the graph, we evaluate the centrality value of each node, which is motivated by social network analysis, to be the degree of representative of each photo.
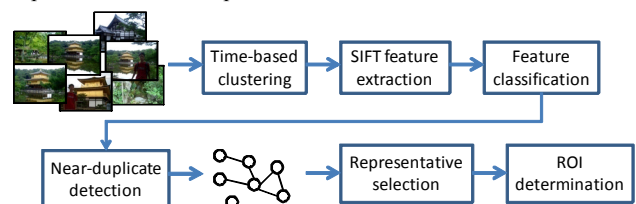

Figure 1. Framework of the proposed summarization and presentation methods.

According to graph structure and each node's centrality value, we dynamically determine number of representative photos in a cluster. A summary of a journey is constructed by a set of representative photos that present a sequence of photo clusters in temporal order. In addition, ROI in each representative photo is extracted based on the spatial distribution of matched feature points, which are byproducts of NDD. Representative photos facilitate browsing a journey in a macroscopic way, while ROIs facilitate browsing in a microscopic way. In this paper, we only describe and evaluate the representative selection module.

3. NDD WITH FEATURE FILTERING

If there are near-duplicate objects in two photos, the orientation of SIFT matched lines between these two photos would concentrate on a few specific directions. This characteristic is modeled by a binary classifier, i.e., support vector machine (SVM), which examines the matching situations between two photos and determines whether they are near-duplicate or not [3].

We improve this work by differentiating different types of features [2]. First, each photo is divided into 40×40 regions. The matched feature points in each region are then aggregated to describe a region. Because artificial objects often have specific geometry structure, some dimensions in aggregated feature vectors would have much larger value than others. We model this characteristic by an SVM classifier, which classifies each region into the one on artificial objects or that in natural scenes. Only features on artificial objects are considered in NDD. We have demonstrated that the method with feature filtering has superior performance in NDD and the following applications [2].

4. REPRESENTATIVE SELECTION

After transforming near-duplicate relationships between photos into a graph, we calculate degree centrality of each photo. According to graph structure, we dynamically determine the number of representative photos for each scenic spot. Three cases would be considered: (1) if only one near-duplicate group exists in this cluster, i.e., one subgraph, only the photo with the largest centrality value is selected as the representative photo; (2) if there is only one subgraph, the nodes with the first few largest centrality values are selected; (3) if there are more than two subgraphs, one or a few nodes with the largest centrality values are selected for each subgraph.

After the processed described above, a sequence of representative photo, which is displayed in temporal order, is finally generated to present the progress of a journey. Note that the uniqueness of the proposed method is that representative photos are selected based on NDD. With this method, this system can effectively select the photos containing important objects. In contrast to conventional content-based method, we select photos that convey clear semantics and provide more impact on recalling travel experience.

5. RESULTS

We evaluate the proposed method based on four tours, as shown in Table 1, each of which contains hundreds of photos taken in a journey lasting for at least four days. For each scenic spot, one or a few representative photos are selected, and a summary composed of a sequence of representative photos is generated.

We ask photo owners to give one to five points to judge the performance of representative selection for each photo cluster. Larger score means higher satisfaction. Figure 2 shows the demo system and the interface for them to make judgement. The selection performance for each journey is then obtained by averaging. Overall, satisfaction of each journey is 3.58, 4.05, 3.63, and 3.50, respectively.

To quantitatively measure the effectiveness of photo summary, we first determine the ground truth of each journey, including determining meaningful scenic spots, and determining appropriate representative photos for each scenic spot. The precision and recall rates between the ground truth and the generated summaries are calculated. Overall, the precision rates are 42%, 42%, 44%, and 50%, respectively. The recall rates are 87%, 89%, 68%, and 67%, respectively. Higher recall rates mean that many of scenic spots in ground truths are correctly segmented, and the corresponding representative photos are correctly detected. However, not every photo cluster contains meaningful scenic spots. For example, some may take photos outside the bus when transporting and no photo in this photo cluster is able to present a scenic spot. This kind of photo cluster degrades the precision rate.

Table 1. Information of evaluation data.

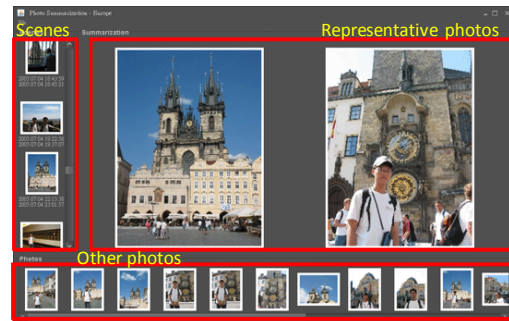| Data set	# photo	# photo clusters	Duration
France – Italy	683	33	8 days
Czech Rep. – The Netherlands	438	36	9 days
California, USA	627	44	10 days
Beijing, China	170	13	4 days

Figure 2. Snapshot of the demo system.

6. CONCLUSIONS

We have presented a photo summarization method facilitated by representative selection based on near-duplicate detection. The generated summaries more appropriately contain important objects so that photo owners can easily recall their travel experience and have high satisfaction.

7. ACKNOWLEDGEMENT

8. REFERENCES

[1] W.-T. Chu and C.-H. Lin, "Automatic selection of representative photo and smart thumbnailing using near-duplicate detection," Proceedings of ACM Multimedia Conference, pp. 829-832, 2008.

[2] W.-T. Chu, C.-H. Lin, and J.-Y. Yu, "Feature points classification for representative photo selection," submitted to ACM Multimedia Conference, 2009.

[3] W.-L. Zhao, C.-W. Ngo, H.-K. Tan, and X. Wu, "Near-duplicate keyframe identification with interest point matching and pattern learning," IEEE Trans. on Multimedia, vol. 9, no. 5, pp. 1037-1048, 2007.