Next Photo Please: Towards Visually Consistent Sequential Photo Browsing*

Dong Liu
School of Computer Sci.& Tec.
Harbin Institute of Technology
Harbin 150001, P. R. China
dongliu.hit@gmail.com

Shuicheng Yan
Department of ECE
National Univ. of Singapore
Singapore 117576, Singapore
eleyans@nus.edu.sg

Hong-Jiang Zhang
Microsoft Adv. Tech. Center
49 Zhichun Road
Beijing 100080, P. R. China
hjzhang@microsoft.com

ABSTRACT

Sequential photo browsing has become the most important function in the desktop and online image repository management systems, where existing systems typically display the photos in default orders such as the lexicographic order of the photo filename or the chronological order of the photo taken time. However, these browsing orders, especially when the browsing speed is fast, ignore the vision persistency characteristic of human visual systems, which results in inconsistent visual experience for photo viewers. To address this issue, we construct a photo relationship graph based on various kinds of visual features that complementarily reflect human visual perception. Then the seeking of visually consistent photo browsing sequence is cast into a traveling salesman problem which seeks an optimal path with minimum visual distance within the graph structure. Experiment results on sequential browsing of Flickr photo groups indicate that the proposed method clearly beats the other sequential photo browsing methods in terms of visual consistency.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Performance, Experimentation, Human Factors

Keywords

Vision Persistency, Sequential Photo Browsing, TSP

1. INTRODUCTION

Increasing popularity of digital cameras and mobile photo cameras has resulted in gigabytes of digital photos sitting on our storage devices or on the web. People take photos to record their lives easily with one-click shot and the cost of taking and storing a photo has been reduced to be zero. As a result, digital photos are becoming part of our daily lives.

Typically, digital photos are testimonials of the events, scenes or objects in our lives, which are used to refresh our memories about the life experiences. It is therefore becoming increasingly important to provide good support for the photo owners to browse the photo collections. Currently, the popular commercial desktop photo management softwares such as ACDSee [1] and Microsoft Digital Image Suite [2] all provide user-friendly interface for photo browsing. Besides, photo browsing has also become a key ingredient of the online photo sharing systems such as Flickr [3] and Picasa [4]. There are mainly two paradigms for browsing photo collections in current systems. In the first paradigm, the users are presented with a set of photos in a two-dimensional grid of thumbnails on the screen. However, it can only display photos using thumbnail (miniature) versions and thus prevents the users from the content details of the individual photos. The second paradigm further overcomes the above drawback by displaying photos with their original resolutions in a sequential manner, which has become the most important photo browsing approach in current systems.

The key issue in the sequential photo browsing is the browsing order of the photo sequence, where the existing systems typically display the photos in a collection in one of the following manners: (1) the lexicographic order of the photo filename; (2) the chronological order of the photo taken time. However, these browsing orders totally ignore the *persistence of vision* characteristic [5] of human visual perception systems, which is closely related with visual experience of human beings. Persistence of vision is the phenomenon of the eye by which an afterimage is thought to persist on the retina. That is to say, if an user views an image at one time point, the appearance of the image will continue to have a visual stimuli on the user's biological vision system. The ignorance of this property in sequential photo browsing may cause inconsistent visual experience for the photo viewers, especially when the photos are browsed in a fast speed. As an example, Figure 1(a) illustrates a photo browsing sequence from Flickr[1]. From the figure we can observe that the photos in the browsing sequence are not consistent in the visual contents, and bring abrupt visual variations to the photo viewers.

The issues associated with the existing sequential image browsing methods motivate us to develop visually consistent method to enhance user's visual experience in sequential photo browsing. The photos in Figure 1(a) show a com-

*Area chair: Mor Naaman

[1]http://www.flickr.com/groups/640498@N21/pool/show

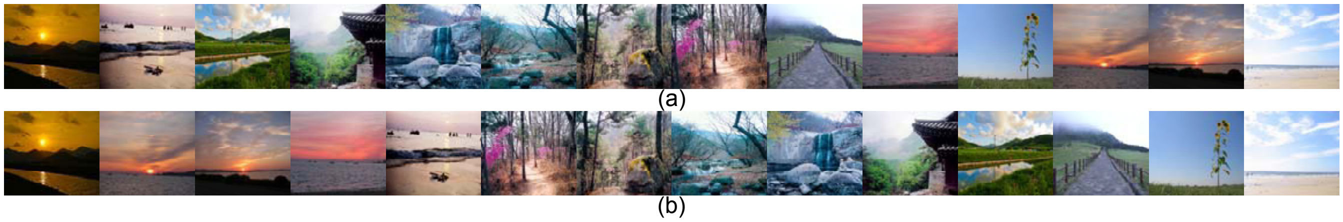

Figure 1: (a) A photo browsing sequence from Flickr. As can be seen, the displaying order is not smooth in terms of visual consistency. (b) A visually consistent photo browsing sequence that orders the displayed photos based on visual contents.

pelling example of where we can improve the visual consistency of the photo browsing sequence. Although the sequence is not optimal in terms of visual consistency, each photo has its visual neighbors in the sequence. If we align the photos in the sequence based on their visual similarities, the subsequent photos will become more smooth in terms of visual consistency, which can be illustrated in Figure 1(b). Finding the visual consistency between the photos and align them accordingly is the basis of the photo browsing method proposed in this paper.

There are two main difficulties in finding visually consistent photo browsing sequence. The first is the image processing required. Note that the images in a collection may have an entirely separate set of visual features that are common among the whole collection, which have a great impact on the visual similarity/distance estimation. For example, some images are similar in terms of color features while others are similar with respect to texture features. Therefore, we look for different kinds of visual features to complementarily reflect the properties of human visual perception.

The second difficulty is that even after we estimate the image similarities/distances based on complementary visual features, we still need a mechanism to seek a photo browsing sequence that reflects visual consistency. Towards this goal, we infer a photo relationship graph, where the photos are linked to each other based on their visual distance. Then the visually consistent photo sequence seeking task can be intuitively cast into a Traveling Salesman Problem (TSP) [6] since we are looking for an optimal path in a connected graph where each vertex (image) has to be visited (ordered) exactly once. We will show that this method can effectively construct visually consistent photo browsing sequence that enhances user's visual experience in photo browsing.

The main contributions of this paper are as follows:

- We propose a visually consistent sequential photo browsing method to visualize the photos in an collection. As our best knowledge, this is the first effort towards improving visual consistency issues in sequential photo browsing.
- We propose to use complementary visual features to construct a graph between the images. We also employ the TSP solver to seek an optimal path in the graph as the desired photo browsing sequence that reflects visual consistency.

2. RELATED WORK

Our work is first related to the visualization methods in image browsing applications. Several research efforts have been dedicated to improve the user's experience in image browsing. Combs et al. [7] have proposed a zoomable user interface to make better use of screen space than the generally applied scrolling. Rodden et al. [8] have further investigate whether it benefits the users to have the sets of thumbnail images arranged according to their mutual visual similarities. The arrangements are created using the multidimensional scaling (MDS) algorithm which treats the inter-object dissimilarities in some high dimensional space, and then attempts to approximate them in a low dimensional output space. Despite these efforts have shown encouraging results, we can see that they only focus on organizing a collection of thumbnail images in a two-dimensional grid plane, whereas there is still a severe lack regrading improving user's experience in sequential photo browsing.

From the aspect of seeking an order from a given image collection, our method is slightly related to the image ranking works in the image retrieval discipline. For example, Jing et al. [9] have proposed the VisualRank algorithm, relying on analyzing the visual link structures among the images, to rank the returned image set of text queries. Liu et al. [10] have constructed the relevance between the tags and the image contents, and utilized the tag relevance information to rank image search result of the tag queries. The focus of these works is to differentiate the images that are relevant to the query keywords from those that are irrelevant, which, however, is essentially different from the sequential photo browsing method introduced in this work, where the images are ranked in term of visual consistency.

3. OUR APPROACH

In this section, we cast the seeking of visually consistent photo browsing sequence into a TSP problem on a graph structure. We first discuss the low-level representations for the images, and then present how TSP can be utilized to seek an optimal path in the graph structure.

3.1 Feature Representations

As aforementioned, the individual low-level features for representing the images in a collection, when used alone, are often too restrictive for the breadth of image contents that need to be handled. Therefore, we turn to the complementary features instead. More specifically, we represent each image with a 418-dimensional feature vector composed of three kinds of low-level global visual features. The first is color moments over 5×5 fixed grid partitions [11], where the first 3 moments of 3 channels of CIE Luv color space are extracted, which results in a 255-dimensional feature vector per image. The second is Gabor texture [12] where we take 8 scales and 8 orientations of Gabor transformation and further use their means and standard deviations to represent the whole image which results in 128-dimensional feature vector. The third is a 75-dimensional edge direction histogram [13], which is taken as the line direction distribution of an image and represents the global information of shape in an image.

3.2 Browsing Sequence Seeking as TSP

We now formulate the seeking of visually consistent photo browsing sequence as a TSP. Assume we are given an image collection $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ where N denotes the total number of images. We first construct a salesman graph G. To do this, we set up a city i for each image $x_i \in \mathcal{X}$. For graph adjacency construction, we aim to build up the edge linkages between the city vertices. Given two city vertices i and j, and their corresponding low-level feature representations x_i and x_j, we are to put one edge if and only if the two cities are nearby enough. In this work, we employ the mutual k-nearest neighbor strategy to create the nearby relationship between the cities. Specifically, two cities i and j are linked with an edge if image x_i is among the k nearest neighbors of image x_j and vice versa. Here the k nearest neighbors are measured by Euclidean distance and the parameter k is fixed as 50. Note that the adjacent relations defined above are symmetric and consequently the constructed adjacent salesman graph is undirected.

The next task is to establish the real-valued (symmetric) distances between any pair of cities, which represent the pairwise visual dissimilarities between the images. In this work, we associate the edge between vertices i and j with a dissimilarity measure d_{ij}, which is calculated as follows,

$$d_{ij} = \begin{cases} \|x_i - x_j\|^2, & \text{if i and j are connected,} \\ +\infty, & \text{otherwise,} \end{cases} \quad (1)$$

where x_i and x_j denote the low-level features of the two images and $\|\cdot\|$ denotes the ℓ_2 norm.

The goal of TSP is to find a *Hamiltonian cycle*, i.e., a sequence which visits each vertex (city) in the graph exactly once and then returns to the starting vertex, with the least weight in the graph [6]. Formally, we define a tour of cities (images) as a cyclic sequence π of cities $\{1, 2, \ldots, N\}$ where $\pi(i)$ represents the city that follows the city i on the tour. The TSP problem is then formulated as the optimization problem to find a permutation π that minimizes the length of the tour defined by

$$\min \quad \sum_{i=1}^{N} d_{i\pi(i)}. \quad (2)$$

For this minimization task, the tour length of $(N-1)!$ permutation vectors have to be compared. This results in a problem which is very hard to solve and in fact known to be NP-complete problem [14]. To find the optimal tour, Held et al. [15] present a dynamic approach to solve the TSP, an optimization problem that is defined as follows: Given a subset of city vertices (excluding the first city) $S \subset \{2, 3, \ldots, N\}$ and $l \in S$, let $d^*(S, l)$ denote the length of the shortest path from city 1 to city l, visiting all cities in S in-between. For $S = \{l\}$, $d^*(S, l)$ is defined as d_{1l}. Then the shortest path for larger sets with $|S| > 1$ is

$$d^*(S, l) = \min_{m \in S \backslash \{l\}} \big(d^*(S \backslash \{l\}, m) + d_{ml} \big). \quad (3)$$

Finally, the minimal tour length for a complete tour which includes returning to city 1 is

$$d^{**} = \min_{l \in \{2,3,\ldots,N\}} \big(d^*(\{2, 3, \ldots, N\}, l) + d_{l1} \big). \quad (4)$$

Using the last two equations, the quantities $d^*(S, l)$ can be computed recursively and the minimal tour length d^{**} can be found. In a second step, the optimal permutation $\pi = \{1, i_2, i_3, \ldots, i_N\}$ of city vertices 1 through N can be computed in reverse order, starting with i_N and working successively back to i_2. The procedure exploits the fact that a permutation π can only be optimal, if

$$d^{**} = d^*(\{2, 3, \ldots, N\}, i_N) + d_{i_N,1} \quad (5)$$

and, for $2 \le p \le N-1$

$$d^*(\{i_2, i_3, \ldots, i_p, i_{p+1}\}, i_{p+1}) = d^*(\{i_2, i_3, \ldots, i_p\}, i_p) + d_{i_p, i_{p+1}}. \quad (6)$$

Because the images are linked together in a graph via complementary visual features that reflect human visual perception, the shortest tour found by TSP solver will correspond to a smooth image sequence in which the subsequent images are visually consistent, which essentially improves the visual experience in sequential image browsing.

Table 1: The details of the photo groups.

Name	Photo #	Name	Photo #
Germany	163	Widelife	1,221
Korea	493	Sunrise	286
Nature	183	London	204
New York	157	Thailand	500
Paris	265	Manasquan	333
Exposure	185	Shanghai	365
Bombay	719	China	500
HongKong	136	NewBook	157

4. EXPERIMENTS

In this section, we conduct extensive experiments to evaluate the performance of the proposed sequential photo browsing method by comparing it with the other sequential photo browsing methods.

4.1 Datasets and Methodologies

We perform the experiments on 16 public photo sharing groups[2] that are collected from Flickr. These photo sharing groups are captured by different people at different locations around the world and contain diverse visual contents, including cityscape, landscape, wide life, etc. Table 1 illustrates the detailed information of these albums. To evaluate the performance of the visually consistent sequential photo browsing method, we employ the following two sequential photo browsing methods as comparison baselines.

- Browsing the photos in each photo group based on the lexicographic order of photo filename, which will be referred to as *browsing by filename*.

- Browsing the photos based on Affinity Propagation (AP) clustering algorithm [16]. For each photo group, we first employ the AP clustering algorithm to divide the photos into a number of clusters. Then the photo browsing sequence can be obtained in a two-step manner: (1) For each cluster, we calculate the average value of all pairwise distances between the images in the cluster and then order the clusters according to the obtained values. (2) The images within each cluster can be ordered according to their distances to the cluster center. We name this method as *browsing by AP* in the experiment.

[2] http://www.flickr.com/groups/

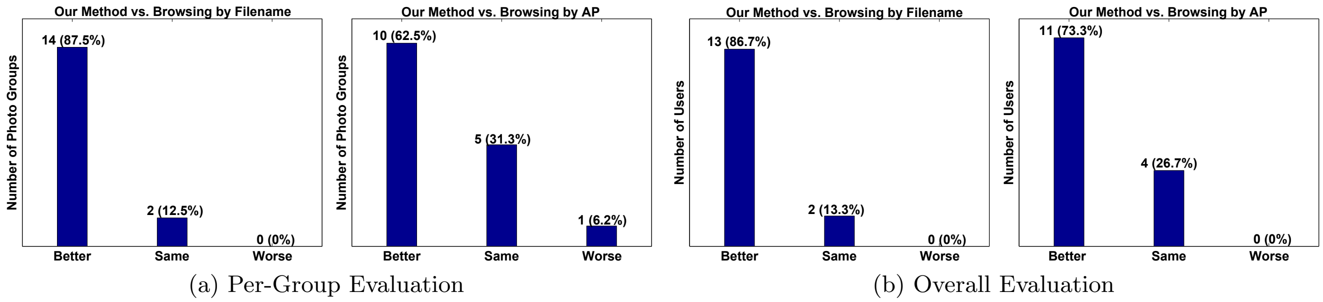

Figure 2: (a) Per-group comparisons between our proposed and browsing by filename and browsing by AP methods on the 16 photo sharing groups. (b) Overall comparison between our proposed method and the other two browsing methods from the 15 users.

For each photo group in Table 1, we use both the two baseline methods and our proposed method to generate photo browsing sequences. Therefore, there are totally 48 (3×16) photo browsing sequences for evaluation. To validate the performance of each photo browsing method, we resort to a human judgement procedure. We invite 15 graduate students to view each photo browsing sequence and accordingly manually label the visual consistency. To simulate the fast browsing in real scenarios, we arrange the browsing as a dynamic sequence where the time interval between two subsequent images is 0.1 seconds. Since it is impractical to ask the evaluators to views hundreds or even thousands photos in a photo group at a time, we only provide a fragment of the photo sequence, which consists of the top 20, median 20 and last 20 photos in the given photo sequence, for displaying. To avoid the bias on the evaluation, the participants do not have any prior knowledge about the method adopted by this work.

We provide two evaluation approaches to justify the performance of three sequential photo browsing methods. (1) Per-group evaluation. In this approach, based on the comparisons of the photo browsing sequences generated by our proposed method and the browsing by filename (or browsing by AP) method, the human evaluators are asked to assign score "2", "1" or "0" to indicate that our proposed method performs better, same and worse than the browsing by filename (browsing by AP), respectively. (2) Overall evaluation. After viewing the photo browsing sequences on all photo groups, the evaluators are asked to make overall comparisons between the proposed method and other two browsing methods. They are asked to judge whether the proposed method works "better" (score 2), "same" (score 1) or "worse" (score 0) than the other sequential photo browsing method.

4.2 Experimental Results and Analysis

The evaluation results for the 15 users are illustrated in Figure 2. In Figure 2(a), we present the average number of the photo groups which have score "2", "1" or "0" assigned by the human evaluators. Form the figure, we can have the following observations. (1) The visually consistent sequential photo browsing method performs better than the browsing by filename method over 14 groups, closely over 2 groups and worse on 0 group. This clearly demonstrates the effectiveness of exploiting visual content information in the task of sequential photo browsing. (2) Comparing to the browsing by AP method, the proposed method performs better over 10 groups, closely over 5 groups and worse only over 1 group, which owes to the fact that the TSP algorithm is able to obtain optimized photo sequence for browsing. Figure 2(b) shows the overall evaluation results. All 15 users consider the proposed method performs no worse than the other two sequential photo browsing methods. Specifically, 86.7% and 73.3% users report that the proposed method is better than the other counterpart.

5. CONCLUSIONS

We have introduced a visually consistent sequential photo browsing method that displays the photos in a sequence in terms of visual consistency. Based on an image relationship graph constructed with complementary visual features, the seeking of visually consistent photo browsing sequence is cast in a traveling salesman problem over the graph structure. Experiment results on real Flickr photo groups have confirmed the effectiveness of the proposed method. For future work, we will apply the proposed photo browsing method into the application of photo-to-video transformation, which will be a promising method to enhance the visual consistency in the video contents.

6. REFERENCES

[1] ACDSee. http://www.acdsee.com/.
[2] Microsoft Digital Image Suite. http://www.microsoft.com/products/imaging.
[3] Flickr. http://www.flickr.com.
[4] Picasa. http://picasa.google.com.
[5] A. Joseph and B. Fisher. The Myth of Persistence of Vision. *Journal of the University Film Association*, 30(4), 1978.
[6] E. Lawler, J. Lenstra, A. Kan and D. Shmoy. The Traveling Salesman Problem. *Wiley*, 1985.
[7] T. Combs and B. Bederson. Does Zooming Improve Image Browsing? In *ICDL*, 1999.
[8] K. Rodden, W. Basalaj, D. Sinclair and K. Wood. Does Organisation by Similarity Assist Image Browsing? In *CHI*, 2001.
[9] Y. Jing and S. Baluja. VisualRank: Applying PageRank to Large-Scale Image Search. *TPAMI*, 30(11), 2008.
[10] D. Liu, X.-S. Hua, L. Yang, M. Wang and H.-J. Zhang. Tag Ranking. In *WWW*, 2009.
[11] M. Stricker and M. Orengo. Similarity of Color Images. In *SPIE*, 1995.
[12] B. Manjunath and W.-Y. Ma. Texture Features for Browsing and Retrieval of Image Data. *TPAMI*, 18(8), 1996.
[13] A. Vailaya, A. Jain and H.-J. Zhang. On Image Classification: City Images vs. Landscapes. *PR*, 31(12), 1998.
[14] D. Johnson and C. Papadimitriou. Computational Complexity, 2009.
[15] M. Held and R. Karp. A Dynamic Programming Approach to Sequencing Problems. *Journal of SIAM*, 1962.
[16] B. Frey and D. Dueck. Clustering by Passing Messages between Data Points. *Science*, 315(5418), 2007.