USER STUDY OF THE FREE-EYE PHOTO BROWSING INTERFACE

Kan Ren, Janko Ćalić
I-Lab
University of Surrey
Guildford GU2 7XH, UK

Risto Sarvas
Helsinki Institute for Information Technology
PO Box 9800
02015 TKK, Finland

ABSTRACT

The striking proliferation of user-generated as well as broadcasted visual content prompted a high demand for effective content management tools and interfaces for search and browsing of visual media. This paper presents a novel intuitive interactive interface for browsing of large-scale image collections. It visualises underlying structure of the dataset by its size and spatial relations. In order to achieve this, images are initially clustered using an unsupervised graph-based clustering algorithm. By selecting images that are hierarchically laid out on the screen, user can intuitively navigate through the collection or search for specific content. The conducted experimental results based on user evaluation of photo search, browsing and selection demonstrate good usability of the presented system and improvement when compared to the standard methods for interaction with large-scale image collections.

1. INTRODUCTION

Nowadays, we are witnessing a striking proliferation of digital content, generated by users themselves, the omnipresent capture devices that surround us as well as the growing multimedia industry. This trend will enforce radical transformation of the way content is maintained, managed and exploited. Driven by the continually changing environment and the need for effective management of large-scale multimedia datasets, there is a strong demand for efficient and flexible way of interaction with the digital content. Personal media devices such as digital cameras or video recorders have become a commonplace. Users can easily take hundreds of photos and video clips on a daily bases. Currently, personal photos only have capture date and time as the default metadata. This implies that the user’s local storage is filled with photos and video clips in an unordered manner. The problem of browsing and retrieving content from such collections is becoming a major challenge of multimedia management systems.

There are two major approaches that tackle this problem. One approach is to ask users to manually annotate visual content every time they import the media. This approach has been proven unfeasible, mainly due to the proliferation of everyday digital media produced by a common user. The other option is to generate annotation automatically using content-based media analysis, computer vision and machine learning. However, due of the problem of the semantic gap between the low level features such as colour, texture, etc. and high-level semantic understanding of the media, the content-based retrieval cannot deliver satisfying results.

The work presented in this paper makes a shift towards more user-centered design of interactive image and video search and browsing interfaces by augmenting user’s interaction with content rather than learning the way users create related semantics. This shift enables not only efficient retrieval of the desired content, but offers more intuitive access to vast image collections and often gives unexpected perspective of the explored dataset. Finally, this approach facilitates more intuitive and effortless browsing, enabling exploitation of the system by a wider user base. The conducted user-centric evaluation of the browsing interface (dubbed FreeEye) demonstrated efficient and intuitive navigation though large personal photo collections, thus facilitating familiarisation with the content and effortless selection of a thematic subset.

The paper is structured as follows. The work related to this area is presented in Section 2, while the Section 3 brings the methodology used in generating the browsing interface, describing the utilised image clustering and interface design methods. In order to evaluate the presented system, Section 4 describes the experimental setup and discusses the achieved results, while in Section 5 we reflect upon the results and outline the future plans.

2. RELATED WORK

There has been a lot of effort put in the scientific research as well as commercial development of user-friendly photo browsing applications. The majority of current photo management applications are based on the time domain clustering, having the temporal metadata readily available from the digital cameras [1] [2]. But the disadvantage of this approach is that the user needs to annotate manually additional metadata, which can be incoherent given the fact that events can span more groups and vice versa. Triggered by the prolifera-
3. INTERACTIVE IMAGE BROWSING INTERFACE

In order to interactively browse large photo collections, the browsing interface follows the idea of ranked image representation, where more relevant images should be more apparent and thus displayed bigger. This is supported by a hierarchical layout of images on the screen. When user selects an image from the displayed dataset by clicking, the image is relocated to the repository and arranged on the screen. By doing this, the user practically moves the centre of perspective from which the collection is explored.

Fig. 1. Building blocks of the image browsing interface

The image browsing system comprises two main modules: image clustering engine and the interface generation, as depicted in the Figure [1]. The image rank in a generated display is proportional to the similarity measure between user-selected central image and other images from the dataset. The choice of the similarity metric is completely independent of the proposed clustering engine and interactive interface, enabling generic applications of this system. In this paper we conducted user-tests exploiting two descriptors: a three dimensional RGB colour histogram and the timecode field from the photo’s Exif data.

3.1. Image Clustering

To achieve system scalability and algorithm complexity nearly linear to the number of key-frames, a graph based clustering algorithm is utilised [8]. Its ability to preserve detail in low-variability clusters while ignoring detail in high-variability regions maintains notion of global features of the dataset in the process of making greedy decisions locally. The graph $G = (V, E)$, is formed so that each image corresponds to a node $v_i \in V$, and the images are connected by undirected edges $(v_i, v_j) \in E$. Weights of each edge $w(v_i, v_j)$ measure the dissimilarity between the two corresponding images.

The graph node grouping is defined by a graph predicate $D(c_1, c_2) : Ext(c_1, c_2) > mInt(c_1, c_2)$, which evaluates if the two clusters $c_1$ and $c_2$ should stay disconnected by comparing inter and intra cluster differences.

The intra cluster difference within cluster $c$ is defined as the largest weight in the minimum spanning tree $MST(c, E)$ of the cluster $c$:

$$ Int(c) = \max_{e \in MST(c, E)} w(e) \quad (1) $$

The joint intra cluster difference measure $mInt(c_1, c_2)$ is therefore given as:

$$ mInt = \min(Int(c_1) + \tau(c_1), Int(c_2) + \tau(c_2)) \quad (2) $$

The inter cluster difference $Ext(c_1, c_2)$ is the minimum distance between the two nodes that are members of different clusters:

$$ Ext(c_1, c_2) = \min_{v_i \in c_1, v_j \in c_2} w(v_i, v_j) \quad (3) $$

The threshold function $\tau(c) = k/|c|$, where $k$ is some constant parameter and $|c|$ denotes the size of $c$, controls the degree to which the difference between the two components must be greater than their internal differences. The intra component difference is defined as the minimal weight edge connecting the two components. The technique adaptively adjusts the merging criterion based on the degree of variability in neighbouring regions of the dataset. The node grouping is iteratively repeated until there is no more component merging.

3.2. Interface Design

The interactive interface is generated following two main objectives: i) to visually convey data structure extracted in the
image clustering stage and ii) to achieve intuitive interaction with this structure. The interface design follows support of the hierarchical groups generated by the clustering engine. An example of the interface is given in Figure 2. The centre image is maximised and displayed at 16% of the displayed window size. If the user clicks on an image, the image will move to the centre of the refreshed screen, and the remaining display layout will reform in order to represent images in the vicinity of the central image. The immediate neighbourhood is represented with 12 most similar images from the same cluster encircling the central image. These images are displayed at 4% the window size. The next layer encircling the central cluster contains 36 images displayed at 1% of the window size, separated into two parts: four edges and four corners. The 32 images located at the four edges are representing the centres of clusters closest to the central image. To support knowledge discovery and help users locating other areas of interest, four random pictures from the set of unrepresented images are located at four corners of the screen. For the initial screen, the system chooses random 49 cluster centres to display, or if the cluster centres are not enough for 49, the rest will be the random images from the user database. This screen can give the user a general content summary about their database. Every time the user clicks, the system iteratively re-arranges all images as described above. In case the user gets stuck in an unwanted cluster, the ‘Esc’ button will randomise the content and get the user back to the initial browsing status.

4. USER EVALUATION

To evaluate the designed photo browsing tool we conducted ten user trials. Of ten recruited participants, five were women and five were men, all aged 24-32, and all but two were advanced computer users. For each trial, a participant contributed with a set of personal digital photos taken during the last 12 months. The number of photos contributed by each participant ranged from 1063 to 1775. For each participant three separate tasks were set. All experiments were conducted in the full screen mode at 1440x900 resolution on a 15 inch screen. The first task was to select photographs from a long event (lasting more than 5 days) to be sent by email to a friend of family. The second task was to select photographs for a yearbook photo album, representing events within last 12 months. The final task was to create a selection of photos as a gift to a close person. For each task the participants were asked to think about specific people they would show the photographs to. The selected photographs were not actually sent or shown to anyone outside the trials. In order to evaluate the usability of the FreeEye system, after each task the participants were asked a set of questions about the tool, the event, and photographs. The participants were also asked to give a score from 1-5 on how well the tools represented the events, how well the tool helped them to find photographs, how this version tool compared to the version they previously used, and how the tool compared to their regular ways of selecting photographs. In addition, two versions of the tools were tested, one that calculates image similarity solely on the RGB colour descriptor, and one that linearly combines it with the timestamp when the photos were taken.

The quantitative results of the comparison between the colour only (NT) and colour with time (WT) image similarity are given in Table 4. The values represent the average difference in time required for completion of each task in seconds, average time between two clicks and average time required to select a photo. The subjective results are given in Table 4, stating the user satisfaction with the two versions of system descriptors.

<table>
<thead>
<tr>
<th>Evaluation question</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>How well the tool helped in selection</td>
<td>3.9</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>How well the selected photos reflected the event</td>
<td>4.3</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>How does it compared to regular way of selection</td>
<td>3.8</td>
<td>3.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. User satisfaction results

4.1. Evaluation Results

In the task of summarising a long event, the participants searched for photos of their holiday trips, hikes or other types of travel. For the yearbook task, the whole set of images was used and no temporal or event restrictions were given. The participants...
selected about 18 photos in long event task and personal gift task to be sent to friends, family, or people who were in the photographs. In the case of the yearbook, the participants selected around 30 photos to make the book mainly for themselves and planned to show it to friends and family.

The participants were satisfied with the selected photos and the way they represented the event. In the long event and yearbook task (task 1 and 2) they reported that they felt that almost no photos were missing in both different version tools. In the personal gift event almost all of them still felt that no photographs were missing, but only one participant felt that she missed 5 photographs.

Overall, the tool scored high in user satisfaction in our trials for both descriptors. As shown in Table 2, the overall average score for how well the tool helped the user in selecting photographs was 3.9 & 4.1 on a scale from 1-5 (1=very bad, 5=very good). Compared to the participant’s regular ways of selecting photographs for similar tasks it scored 3.8 on a scale of 1-5 where 3 was as good as their regular one and 5 was much better. Generally the tool was thought to be good in recollecting events and photographs taken. The way in which it showed forgotten photographs was mentioned as a positive thing (score 4.3). One of the main issues the participants had with the tool was that if they had a particular photograph in their mind, it was not always easy to find it. Compared to the other tasks, the Task 1 (long event) was considered harder because, unlike the yearbook task and personal gift task, the long event was restricted in time. However, after including the time feature to the image similarity measure, the user satisfaction has notably improved (from 3.9 to 4.1).

Our research interest focuses on building a user interface that leverages available information to facilitate the photo selection process, not to automate it. Selecting photographs from increasingly large personal collections is a common task for a variety of situations. For that reason we have built a tool where the user is in charge and does the final selection. In our tool we used only the visual and temporal similarity information to help the user select photos for emailing, uploading, or making a photo album. The two analysed features were considered helpful and the scores show that the participants were very satisfied with the tool and the selected photographs.

The tool in its simplicity has potential as a general user interface for selecting media from a large collection. What we learned from our trial was that our tool seems to work well with personal collections: the participants knew their own photographs or took part in the events which were in the photos, which helped them to feel in control. The strength of our tool is that it is a general tool that is not coupled with any particular task or with any particular system. The other main strength is that according to our user trial, people found it inspiring and fun.

In future research, we will focus on adding other interactive functionalities and additional image features, such as location, people and tags, to improve user satisfaction.

6. REFERENCES


