Noname manuscript No.
(will be inserted by the editor)

Interactive Search and Browsing Interface for Large-scale Visual Repositories

Kan Ren · Risto Sarvas · Janko Ćalić

Received: date / Accepted: date

Abstract Due to the rapid proliferation of user generated visual content, as well as staggering influx of video broadcasted online, the interfaces for search and browsing of visual media have become increasingly important. This paper presents a novel intuitive interactive interface for browsing of large-scale image and video collections. It visualises underlying structure of the dataset by its size and spatial relations. In order to achieve this, images or video key-frames 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 extensive experimental results based on user evaluation of photo search, browsing and selection as well as interactive video search demonstrate good usability of the presented system and improvement when compared to the standard methods for interaction with large-scale image and video collections.

Keywords Image and Video browsing · Interactive interfaces · Unsupervised clustering

1 Introduction

The ever-increasing amount of digital content, generated by users themselves, the omnipresent capture devices that surround us as well as the growing multimedia industry, has transformed the way content is maintained, managed and exploited. Driven by the continually changing environment and the need for effective management of large-scale

K. Ren and J. Ćalić
I-Lab, University of Surrey
Tel.: +44-1483-684739
Fax: +44-1483-686011
E-mail: k.ren@surrey.ac.uk and j.calic@surrey.ac.uk

R. Sarvas
Helsinki Institute for Information Technology
Tel.: +44-1483-684728
Fax: +44-1483-689550
E-mail: risto.sarvas@hiit.fi

multimedia datasets, there is a strong demand for efficient and flexible way of interaction with the digital content. Personal media devices such as digital camera or video recorders have become a commonplace. Users can easily take hundreds of photos and video clips on a daily bases. However, only a few generate high-level annotations at the time of its import into their personal computers. Currently, the photos only have capture date and time as a default metadata, while video clips by default do not have any metadata assigned. This implies that the users 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 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 the 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 visual data 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 search and browsing interface 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. Section 3 brings the methodology used in designing the browsing interface, starting with image clustering and describing the interface layout. In order to evaluate the presented system Section 4 describes the experimental setup, while Section 5 discusses the achieved results. In Section 6 we reflect upon the results and outline the future plans, while the references are given in Section 7.

2 Related Work

There has been a lot of effort put in the scientific research as well as commercial development of user-friendly image and video browsing applications. Most of the browsing applications are based on the time domain clustering of the personal photo collections, having the temporal metadata readily available from the digital cameras. The applications simply cluster the images based on the time when the photo or video clip was generated [1] [2] [3]. But the disadvantage of this approach is that the user needs to type manually an event name for a group of photos, which can be inexact and unreasonable given the fact that events can span more groups and vice versa.

Triggered by the proliferation of global positioning system (GPS) technology, some of the new applications are using the image similarity based on the location where the operator took the photos [4] [5]. Being an emerging technology in this context, GPS

modules are still rarely built into the camera, so users often need to assign the location information manually.

Recently, some commercial applications introduced semi-automated annotation of images by using the face recognition technology [5]. The application first detects face region in the photo and then attempts to identify and tag the image by using face similarity algorithm. However, this approach is unfeasible for many photos, such as landscape photos, animals, etc., since people are not always the major subjects in a captured scene.

There have been a number of approaches to develop visualisation that would augment the usability of interfaces to large image collections. In [6], Huynh et. al. introduced a method that trades off screen space for better presentation of temporal order in photos. In addition, some systems utilised methodologies to analyse the underlying data structures to present image collections [7] in a more accessible way.

However, the interaction with large visual collections has not been addressed in an intuitive way. Derived from its definition in [8], intuition implies correlation between system inference and the users expectations. By following this definition, we developed an intuitive interactive interface (dubbed *FreeEye*) for browsing of large image and video collections, based on the efficient image clustering method and interactive hierarchical interface.

In order to facilitate interactive browsing of video content by the means of the *FreeEye* interface, the proposed system efficiently extracts a set of representative key-frames from the sequences present in the repository by unsupervised clustering methodology. There is a number of similar approaches that utilise unsupervised clustering in the process of key-frame extraction. An efficient clustering method has been utilised in [9] [10], where K-means algorithm is used to classify data into a fixed number of groups, starting from a random initial partitioning. In [11], an unsupervised clustering based approach was introduced to select key-frames within predetermined shot boundaries. Similarity comparison using a shot histogram analysis and subsequent clustering is carried out within each shot to automatically select the most representative key-frames.

Focusing on the frame saliency and importance in the video summarisation context, a number of graph-based methods have been proposed [12] [13] [14]. However, the efficiency of these approaches heavily depends upon the size of the dataset, due to a high complexity of the spectral analysis exploited in their graph representation. Nevertheless, there have been proposals to analyse visual similarity in the graph-based context with almost linear complexity to the number of nodes in the graph representation. Developed for efficient image segmentation, the algorithm presented in [15] introduces a graph predicate that keeps the notion of global features while making fast decisions locally.

3 Interactive Search and 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 centre, while the remaining data is retrieved from the repository and arranged on the screen. By doing this, the user practically moves the centre of perspective from which the collection is explored.

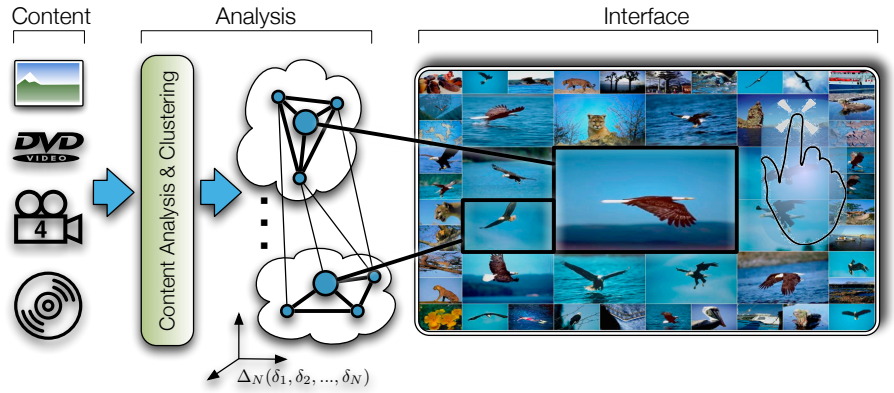

Fig. 1 Block scheme of the system comprising content input, clustering engine and interactive interface

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 a three-dimensional HSV colour histogram has been chosen as the similarity metric, but additional metrics such as photo's timestamps, GPS locations and tag co-occurrence were tested. To achieve system scalability and algorithm complexity nearly linear to the number of images, a specific graph based clustering algorithm is utilised, as described in more detail in Section 3.1.

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. The centre image is maximised and displayed at 100% of its 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 50% their original size. The next layer encircling the central cluster contains 36 images displayed at 25% 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. Every time the user clicks, the system re-arranges all images as described above.

3.1 Image Clustering

There are two clustering methods utilised in this system. The standard K-means algorithm is used to efficiently and robustly remove redundancy of the video sequences and

Fig. 2 Interface design follows the hierarchical structure extracted at the clustering stage

select key-frame candidates as salient representatives, resulting in an over-segmented dataset, as described in Section 3.1.1. In addition, to extract key-frames and cluster images to analyse the underlying structure of the overall image dataset, a fast graph-based algorithm is described in Section 3.1.2.

3.1.1 Video frame selection

In order to achieve fast frame grouping and minimize the perceptual redundancy, K-means clustering algorithm is utilised. This approach is used not only to exploit the algorithms efficiency and robustness, but also to perform unsupervised processing once the number of clustering groups has been estimated.

The traditional approach to the K-means clustering algorithm used for key-frame extraction is to treat every frame of the video as a point in an N-dimensional feature space. Here, we use 128-bin HSV colour histogram as the feature vector. Using an iterative approach, the cluster centres and group memberships are updated in each iteration in order to minimise intra-group distances. Finally, the point nearest to the clustering centre is selected as the key-frame candidate of the group.

The main drawbacks of this algorithm are a need to determine the number of cluster centres a-priori and the initialisation the centres. Since our goal is only to remove redundancy by over-segmenting the dataset, the number clusters K have been established empirically. For a video with the frame-rate of 30fps, a ratio of 90 frames per

group generates on an average one key-frame candidate for 3 seconds of video, which is a reasonable assumption. By doing this, we define exactly the number of key-frame candidates for a given video clip. The number of clusters is thus derived as $K = L/90$, where L is the length of video in frames. The initial positions of cluster centres bias the final grouping of the frames. Thus, in order to achieve an even representation of the visual content, the locations of the cluster centres will be distributed uniformly throughout the video sequence.

During the clustering, only the intra-frame distances are calculated. By doing this, the iterative calculation of exact position of the cluster centre is avoided, resulting in a faster algorithm convergence. Following this approach, the number of key-frame candidates will be higher than the number of shots in the video and many perceptually similar frames will be generated. However, by applying the unsupervised clustering method described in following Section 3.1.2 that balances locally greedy search and global optimisation, a set of video key-frames is generated and used for image-based search and browsing.

3.1.2 Graph-based clustering

As we already stated, having in mind he goal of system scalability and algorithm complexity nearly linear to the number of key-frames, a specific graph based clustering algorithm is utilised [15]. Although initially formulated in the image segmentation context, this algorithm can be extended to a more generic dataset scenario. 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.

Following a common approach to graph based image clustering, this method forms edges of a graph $G = (V, E)$, where each image corresponds to a node $v_i \in V$ in the graph, and certain 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 $D(c1, c2)$, which evaluates if the two clusters c_1 and c_2 should stay disconnected by comparing inter and intra cluster differences, as depicted in Figure 3 and following equations:

$$D(c_1, c_2) : Ext(c_1, c_2) > mInt(c_1, c_2) \tag{1}$$

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

$$Int(c) = \max_{e \in MST(c,E)} w(e) \tag{2}$$

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

$$mInt = min(Int(c_1) + \tau(c_1), Int(c_2) + \tau(c_2)) \tag{3}$$

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

$$Ext(c_1, c_2) = \min_{v_i \in c_1, v_j \in c_2} w(v_i, v_j) \tag{4}$$

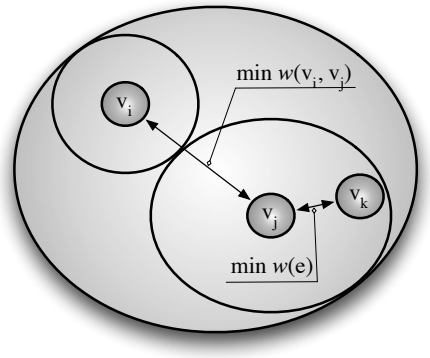

Fig. 3 Graph predicate compared inter and intra cluster similarities, maintaining notion of global features while making greedy decisions locally

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.

4 Experimental results

The proposed interactive search and browsing interface has been evaluated in three different application scenarios. The first scenario comprised an image search task in minimal time, focusing on the overall intuitiveness and efficiency of the system. The second user study comprised three sub-tasks of selecting a set of personal photos depicting an event, a holiday and the whole year. Finally, in order to investigate usability of the system on video medium, an interactive video search task defined by a sentence has been set, and the achieved list of retrieved results was tested for its precision and recall.

In order to conduct these experiments, the proposed system is designed to record the user history, e.g. images they selected, timings and locations of user clicks, user satisfaction, etc. Thus, the users initial screen shows the favourite 45 images from the user history. Otherwise, if the user has never used the tool before, it will display random 49 images from the database on the initial screen.

4.1 Interactive image search

The image repository used is a selection of cca. 3000 colour images from the Corel image database. In order to test the effectiveness of the search and browsing tool, the database subset includes multiple semantic concepts such as the wild animals (leopard, eagle, fox, etc.), nature scenery (forest, ocean, etc.), historical buildings (western temples, Asian buildings, etc.), portrait, plants (flower, garden, etc), etc. A sample

The subjective tests were conducted by inviting 26 people to join the challenge *Find me a postcard* [16]. The challenge comprised finding 5 images from a set of 3000 only by means of interactive interface described above. Of 26 people involved, 18 persons were male, and 17 had the advanced computer knowledge. All users were using the tool for the first time and the only requirement was to have a basic knowledge of manipulation with a mouse. The gender, racial and cultural diversity of the subjects was balanced.

The task was to find the 5 fixed images in the predefined order. The content of the five images was varied, as presented in the Figure 4. We recorded the full browsing system state for every user step, which included indexes of all images on the screen, their positions, user selection and timing. This has enabled us to fully reproduce the browsing process for each user and analyse achieved results.

Fig. 4 Images used as queries in the interactive image search task

The basic statistics of the experimental results shows that the average time for a user to finish the whole experiment is 8 minutes and 20 seconds in 50 mouse clicks. This gives an average of around 100 seconds time and 10 mouse clicks needed for a user to find an image from the database of 3000 images. Assuming that in the case of thumbnail presentation users need to inspect all images from the data set, the average number of images inspected by using the FreeEye tool is 6 times smaller.

In order to evaluate the interface intuitiveness, the user history records are studied. The Figure 5 shows users browsing paths and distribution of user clicks for all 26 users in all 5 tasks. The left column of the Figure 5 presents the distance between the desired image and the central image for each user click from the start of the task until the desired image is found. The right column in the Figure 5 shows a histogram of user clicks needed to find the desired image for all 5 tasks.

From all 5 browsing paths, it is observable that after only a few clicks, the trend of the distance curves is to fall towards zero. This means that the users were rapidly converging towards the goal of the task just after a couple of clicks, implying systems intuitive character. This trend is obvious in the 2nd, 3rd and 5th task, while the initial task and the 4th task that was a more difficult one, demonstrated the same convergence, but required more user clicks.

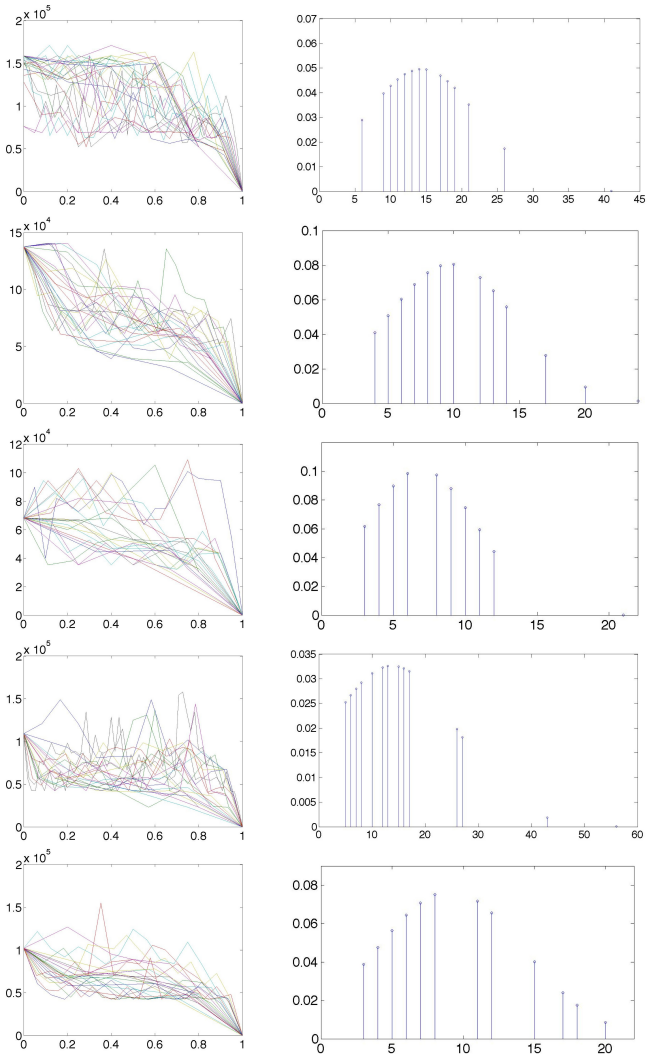

Fig. 5 Convergence of the five search tasks and the corresponding distribution of clicks for all users in the interactive image search task

Since the timing and user clicks directly depend upon the difficulty of the task, we studied the distribution of the number of user clicks required to find the desired image in the database. From the histograms shown in the right column of the Figure 5, it can be observed that the distributions become increasingly skewed in a positive sense (right-skewed) as the users progress through the tasks. This represents that more users require less iterations to find the desired image as they use the interface. This characteristic demonstrates that without any assistance, users intuitively learn how to efficiently use the interface, regardless of the task difficulty. The same conclusions were made while studying the distribution of time required to find the desired image for each user.

Fig. 6 Spatial distribution of user clicks in the interactive image search task

In addition to the click and time statistics, we have studied the spatial distribution of positions of images selected by users. As depicted in the Figure 6, where the region brightness represents frequency of its selection, images in the second level (neighbouring frames of the central image) are selected more often than images in the third level. However, some of the random images in the four corners were occasionally selected, mainly to move away from the currently displayed set of images and test where they would take the user in his search attempt. Furthermore, the top area of the second level was slightly more popular than bottom area, while the right side was a more popular than left side.

4.2 Photo selection task

In order to evaluate the proposed system in a photo selection scenario, we conducted five user trials [17]. The recruited participants were 3 women and 2 men aged 24-32, and all but one had a computer science background. For each trial the participant brought a set of their own digital photos. The number of photos brought by each participant ranged from 1385 to 1664. For each participant there were three separate tasks. The first task was to select photographs from a short-time event (1-2 days) to be sent by email to someone. The second task was to select photographs from a long time event (more than two days) to be uploaded to a web page or shown to someone. The third task was to select photographs for a book representing events and happenings in the past 6-12 months. 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. After each task the participants were asked a set of questions about the tool, the event, and photographs. The participants were

Table 1 Quantitative results of the user study

	selected photos	time spent	clicks	sec/click	sec/photo
Task 1	10.4	1:52	16.8	6.65	10.7
Task 2	15.6	5:36	49.2	6.82	21.5
Task 3	23.4	6:16	56.6	6.65	16.1

Table 2 User satisfaction results

	Task 1	Task 2	Task 3	ALL
How well the tool helped to select?	3.9	3.1	4.1	3.7
How well the selected photos reflected the event?	4.5	3.9	4.6	4.3
Compared to regular way of selection	4.2	3.3	4.1	3.9

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, and how the tool compared to their regular ways of selecting photographs. The answers to these questions are summarised in Table 2. For each task the number of clicks and the time spent was measured, as well as the number of photos selected (see Table 1).

The short events the participants searched photos for were a birthday party, roller skating, and holiday trips. For the long events the participants all had a trip: hiking, traveling, and a long roller skating trip. For the yearbook task whole set of images was used and no temporal or event restrictions were given. The participants selected about 10-20 in each task to be sent to friends, family, or people who were in the photographs. In the case of the yearbook, the participants made the book mainly for themselves and planned to show it to friends and family.

The participants were satisfied on how well the photos they selected represented the event. In the long event task (task 2) they reported that they felt that they missed some photographs they would have liked to have. In the short event they felt that no photographs were missing, and in the yearbook task one participant reported that he got almost all of them, and another participant felt that she missed 5-6 photographs. As seen in Table 2, the participants were very happy with the photographs they had selected in tasks 1 and 3. In task 2 they thought they had missed some, but felt content anyway.

Overall, the FreeEye tool was scored high in our trials. As shown in Table 2, the overall average score for how well the tool helped the user in selecting photographs was 3.7 on a scale from 1-5 (1=terrible, 5=very good). Compared to the participants regular ways of selecting photographs for similar tasks it scored 3.9 on a scale of 1-5 where 3 was as good as their regular one and 5 was much better. All but one of the participants used Windows operating systems user interface to select their photographs, and the tool was considered better than Windows OS (average score of 4.1). The one participant used Picasa and he thought the tool was as good as Picasa (score of 3).

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. 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 easily found. Especially Task 2 (long event) was considered harder to do than the other tasks because there were more pictures than in a short event and unlike the yearbook task, the long event was restricted in time. The quantitative data in Table 1 supports this: more time was spent

per chosen photograph than in the other tasks, although the time spent between clicks was not significantly different.

4.3 Video browsing and interactive search

In order to conduct evaluation of the proposed system in the context of video browsing and summarisation, the experiments were based on the TRECVID benchmarking tasks and content. This content is provided by NIST as the benchmarking material for evaluation of video retrieval systems. Specifically, material targeting the TRECVID interactive search task in 2009 has been used, and one representative screenshot of the interactive video search task is given in Figure 7.

Fig. 7 A screenshot of the interface in the interactive video search task

The interactive search task is defined by a sentence describing required content to be retrieved from the repository. These sentences are called topics. There were 24 search topics, and the FreeEye system was used to browse through the video key-frames and locate all instances that match the topic. The Figure 8 shows timing of the interactive searches per topic.

As envisaged, the precision of the results is very high, as depicted in Figure 9, due to the human decision maker. However, the recall results are very limited, since of possible 10619 relevant shots for the 24 topics, user detected only 1176. Nevertheless,

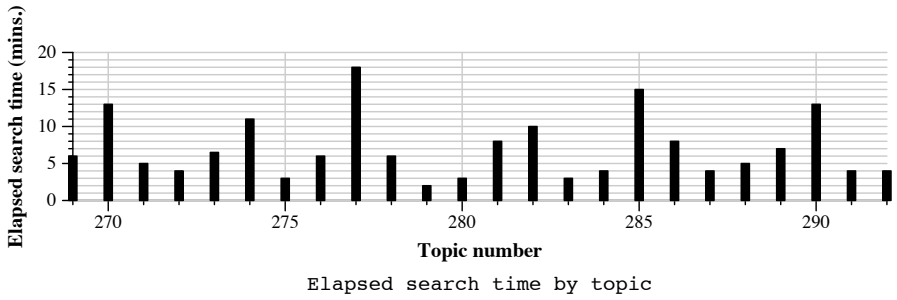

Fig. 8 Timing of the interactive video search task

of all TRECVID participants in 2009, we have scored top to average results, without involving any training or recognition.

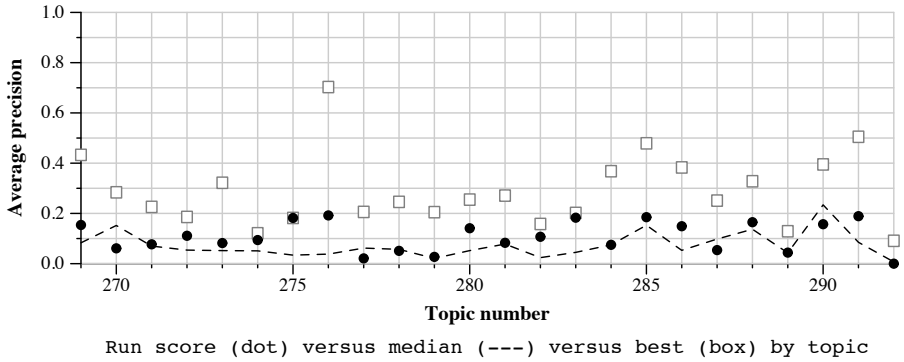

Fig. 9 The precision scores for interactive video search task in TRECID 2009

5 Conclusions

In this paper we have introduced a novel interactive interface for intuitive search and browsing of image and video collections. The presented interface is targeting a multitude of applications: from browsing of personal photo collections, selecting a year photo book to video summarisation and content-based retrieval. From the initial experimental results, the system is very usable and intuitive, while offering pleasant browsing of visual data and often offering new perspectives of the same dataset by making surprising links between the data subsets. In addition, the users could manipulate the visual interface without any specific introduction. Finally, the knowledge discovery element of four random images in the corners of the display has been proven as a very useful tool of the interface.

Having in mind that our research interest is in building a user interface that leverages available information to facilitate the photo browsing, search and selection process,

not to automate it, the results of the user studies are promising. The photo selection 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 similarity information to help the user select photos for emailing, uploading, or making a book. Surprisingly, the visual similarity was considered helpful and as the scores of our trial show the participants were quite happy with the tool and the selected photographs. The evaluation outcomes can be summarised as follows:

- The selected photographs reflected the events very well (4.3/5)
- The tool was considered helpful (3.7/5), and better or as good as their existing ones (3.9/5)
- The participants selected on average 10-23 photographs, and spent from 2-6 minutes in selecting the photographs.

What we learned from our trial was that our tool seems to work well with personal collections: the participants knew their own photographs which helped them to feel in control. This became especially clear with one participant who had in her collection also photographs taken by someone else. This caused confusion and a feeling of being lost. 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 useful and helpful.

The results from the interactive search experiments with video collections demonstrate that through intuitive interaction users can find very specific content with very high precision, yet having a pleasant and fun user experience. The tool in its simplicity has potential as a general user interface for selecting media from a large collection.

In future research, we are adding other similarity measurements to the user interface: location, people, tags, and time. We are also planning to add controls for the user to change the importance of a parameter at any time (e.g., location similarity is more important than visual similarity).

References

1. M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. In Proceedings of the eleventh ACM international conference on Multimedia, pages 364-373. ACM Press, 2003.
2. A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries, 2002.
3. A.Loui and A. E. Savakis, Automatic image event segmentation and quality screening for albuming applications, In IEEE International Conference on Multimedia and Expo, 2000.
4. Toyama, K., Logan, R., and Roseway, A. Geographic Location Tags on Digital Images. In Proc. of 11th Annual ACM International Conference on Multimedia (MM2003) (Berkeley, CA, November 2-8, 2003). ACM Press, New York, NY, 2003, 156-166.
5. Apple Ltd., iPhoto09. http://www.apple.com/ilife/iphoto/
6. Huynh, D. F., Drucker, S. M., Baudisch, P., and Wong, C. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. In CHI '05 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, 1937-1940. 2005.
7. Bederson, B. PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps, Proceedings of the 14th annual ACM symposium on User interface software and technology, pp. 71-80, 2001.
8. Jung, Carl G. Psychological Types. Princeton, New Jersey: Princeton University Press, 1971.

9. Di Zhong, HongJiang Zhang, and Shih-Fu Chang, Clustering methods for video browsing and annotation, Proc. SPIE Int. Soc. Opt. Eng. 2670, 239 (1996), DOI:10.1117/12.234800
10. Yueting Zhuang; Yong Rui; Huang, T.S.; Mehrotra, S., "Adaptive key frame extraction using unsupervised clustering," Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on , vol.1, no., pp.866-870 vol.1, 4-7 Oct 1998
11. Ahmet M. Ferman and A. Murat Tekalp, Multiscale content extraction and representation for video indexing, Proc. SPIE Int. Soc. Opt. Eng. 3229, 23 (1997), DOI:10.1117/12.290352
12. J. Calic, D. P. Gibson, and N. W. Campbell, "Efficient layout of comic-like video summaries", IEEE Trans. on Circ. and Sys. for Video Tech., Vol. 17, Iss. 7, pp. 931 - 936, July 2007.
13. N. Chong-Wah, M. Yu-Fei, Z. Hong-Jiang, Video summarization and scene detection by graph modelling, IEEE Transactions on Circuits and Systems for Video Technology, 15 (2) (2005) 296305.
14. J. Calic and N. W. Campbell, "Compact Visualisation of Video Summaries," EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 19496, 2007.
15. P. F. Felzenszwalb and D. P. Huttenlocher, Efficient Graph-Based Image Segmentation, International Journal of Computer Vision, Vol. 59, No. 2, September 2004.
16. K. Ren and J. Calic. FreeEye - Interactive Intuitive Interface for Large-scale Image Browsing. Proc. of ACM Multimedia (2009)
17. Kan Ren and Risto Sarvas and Janko Calic. FreeEye - Intuitive Summarisation of Photo Collections. Proc. of ACM Multimedia Multimedia Grand Challenge (2009)