Hierarchical image browsing based on visual information

Gábor Szűcs
Inter-University Centre for Telecommunications and Informatics,
H-4028 Kassai út 26., Debrecen, Hungary;
Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics
Budapest, Hungary
E-mail: szucs@tmit.bme.hu

Abstract— **The aim of this paper was to develop an image browsing system, where only visual information is available for users and computers as well. Large image set is structured in albums and the idea was to select the most representative image from each album in the level of hierarchy, and then the next upper level in the hierarchy consists of these representative images. Selection of the most representative image is based on semantic features, which helps users in browsing process. Quantitative and qualitative analysis have been completed, where the goodness of the representative image selection is tested and comparison with other systems is accomplished respectively. Finally it can be concluded that the system developed possesses semantic and structured features.**

Keywords-image browsing, semantic, representative image, user's behavior

I. INTRODUCTION

Image search and browsing (briefly seeking) is a large, more and more important area in the world [8], because the multimedia plays a larger and larger part in everyday life. This area can be divided into two parts: retrieval and browsing. The first one is similar to information retrieval, but in this case the answer is a list of images. At image retrieval the user possesses a well-defined demand, which is constant during the retrieval process, e.g. the user would like to see an image about his/her favorite actor. In contrast to this in the browsing process the user has no constant demand; his or her interest is changing during the process. A good example for image browsing is a process of purchasing a dog, when a person would like to buy a dog, but he does not know which breed. He is looking for different breeds on the Web based on images, and he is starting with sheepdog, but during the browsing his interest changes and at the end he decides to purchase a poodle. From cognitive aspect the browsing is related to the user's changing behavior e.g. changing the mood or interest, and the retrieval is related to constant user's behavior.

The research area of this paper is in the inter-cognitive communication [15], where information transfer occurs between a human and an artificially cognitive system. In our case the latter one is an artificial browsing system, and the communication channel is visual channel.

This paper focuses on a part of image browsing, where only visual information is available for users and computers as well. So in this case there is no text information about the pictures and the metadata cannot be used for browsing. This restriction naturally influences the seeking process, especially the retrieval, because the relevant hits cannot be found so easily. However, a large advantage of this restriction is language independency, because in lack of text information the browsing process is the same in every language.

Three different circumstances of the order can be defined at this problem:

- structured: At the first one the images are in albums, the set of images in them is fixed, and the album structure is also determined in hierarchical way like subdirectories at file systems.
- shallow structured: At the second one the images can be found in albums as well, but all albums are independent, there are no hierarchical relations among them.
- unstructured: The third one is an unstructured case, where only a large set of images is defined, and there are no albums.

The aim of this work has been to develop an image browsing system in structured circumstances, where only visual information is available, and furthermore to help with user's behavior in browsing process

Navigation among the images can be web-based photo navigation [13], or an interactive search assistant [12]. An image browsing system can help with support efficient and effective (manual) annotation of image databases [11].

II. RELATED WORKS

Web image retrieval systems, such as Google or Bing image search, present search results as a relevance-ordered list. These traditional approaches use a grid layout that requires a sequential evaluation of the images. Although alternative browsing models (e.g. results as clusters or hierarchies) have been proposed and applied. One of them is Google Image Swirl

978-1-4799-1546-0/13/$31.00 ©2013 IEEE

[4], which is a widely available, hierarchical image handling system by automatically grouping the search results based on visual and semantic similarity. Another system is Visual Image Browsing and Exploration (Vibe) [3], which places visually similar images near one another, and supports dynamic zooming and panning within the image search results. Such feature like similarity-based organization of images is now available due to recent advances in image processing and computing power.

A number of image browser systems have been developed over the last few years, notably, IBM's QBIC [5], PhotoBook [6], ImageRover [7], and iMap [10]. Pathfinder network scaling [1] is a structural modeling technique originally developed for the analysis of proximity data in psychology [9]. This modeling technique has been used to simplify and visualize the strongest interrelationships in proximity data. The idea of the Pathfinder is based on the triangular inequality condition, which can be used to eliminate redundant or counter-intuitive links. Pathfinder network scaling particularly refers to this pruning process.

There is another image viewing technique, so called Rapid Serial Visual Presentation (RSVP) [2], based on an adaptive attention shifting model. This gives better user experience for browsing large images on limited and heterogeneous screen sizes of mobile devices.

III. Elaborated image browsing system

The first subtask was the creation of representation of the whole image structure for preparation of online browsing process. This preparation phase is offline, i.e. this is finished before the user starts to use the browsing system.

A. *Navigation on hierarchical album structure*

In the hierarchical album structure the navigation is based on only the thumbnails of the images. The idea was to select the most representative image from each album in the lowest hierarchy level, then the next upper level in the appropriate hierarchy consists of these representative images, finally this process should be executed recursively up to the top level. For the user the thumbnails of the representative images are shown, and each click on a thumbnail goes to the next level.

B. *Comparison of images*

In order to select the most representative picture in an album, the images are semantically compared, that is why SURF algorithm [14] is used for feature extraction.

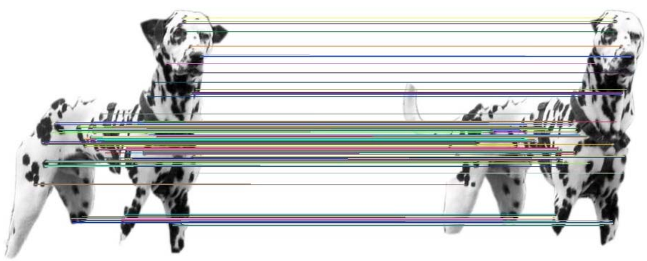

Figure 1. Comparison of two images based on SURF features

An example for comparison of two images based on SURF features (with connected similar point pairs) can be seen in Fig. 1. The number and strength of point pairs give the similarity between two images, so the distances (1/similarity - 1) among the images have been determined.

C. *Select representative images*

In order to select the most representative image, the centroid element should have been found. Clustering helps in finding similar groups, and centroid models of clustering – e.g. k-means algorithm – find these centroids well, but in our problem there is not only one set (images in an album), but many sets, furthermore the representative images should be different. The difficulty is that some images in different albums may be more similar to each other than in a common album. This may cause a confusing situation in the representative thumbnails. An idea was to filter out the confusing elements in the albums, and the calculation of the centroid element is based on only the rest elements, so called core set; and this was one of the contributions of this paper. The process of selecting the centroid image as the most representative image can be seen in Fig. 2.

IV. Implementation and Evaluation

We have implemented the system described above in Python (version 2) programming language, and a snapshot can be seen in Fig.3. The Python Imaging Library (PIL) with OpenCV has been used for the image handling, because this library contains useful functions for basic visual features. The user interface has been based on a micro web-framework, so called Bottle. Another useful tool is the NetworkX, since it possesses graph handling functions, therefore this tool has been applied at the implementation of structure of the albums. The hardware demand of the user interface and the whole application is minimal. For supporting this statement a very little hardware, a Raspberry Pi with ARM (700 MHz) processor has been used as a server, and this has been enough for the task.

In the cognitive science the analysis of user's behavior is very important, with this design the essential part of the solution, i.e. the selection of representative image has been tested. More than half thousand images have been collected, and organized in 20 albums in hierarchical structure. Three end users have selected the most representative pictures, then the second, the third, etc. ones, so they have ranked the images in each album. In the implemented solution the algorithm has also selected the most representative image in each album. These results from the algorithm have been compared with the aggregated rank of three end users. The aggregation has been a simple sum of rank values, and the integrated order has been based on this sum (in ascendant order). The machine results are not always the first in the humans' rank, but we can investigate an easier requirement. Counting the number of cases where the machine results are in the top 10 representative pictures, 14 cases have met this easier requirement out of 20 albums (which is 70% in average). We have investigated the humans' decisions as well with cross-validation, where only two end users' rank lists were considered and aggregated, then were compared with the most representative pictures of the third end user. The comparison result of cross-validation for the first person was 85%, for the second person was 75%, for the third person was 70%. These values show that the automatic solution is almost as good as humans' decisions.

```
Centroid (S)
{ calculate the average of data of S in each dimension
   let X be the point of these averages
   return X
}
Core (Xc, Sc)
{  CoreSet = empty
   do for every P e Sc
   {  NotCore=0
      for j=1 to N  if d(P, Xc) > d(P, Xj) then NotCore=1
      if NotCore=0 then CoreSet = CoreSet + {P}
   }
   return CoreSet
}

Init:
N: NumberOfClasses is given
S1, S2, ... Sc, ..., SN are given as sets of classes
for every c let Sr = Sc

do while serial Xc change
{ for every c let Xc = Centroid (Sr)
   for every c let Sr = Core (Xc, Sc)
}
```

Figure 2. Process of selection the centroid image

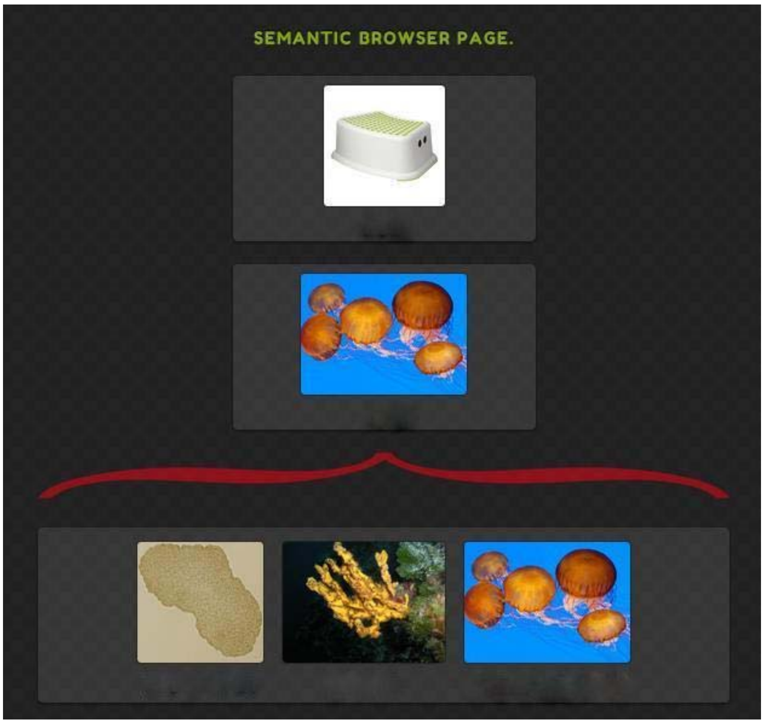

Figure 3. Snapshot of our semantic browser application

Finally the comparison of the solution described in this paper and other solutions in the literature is presented. Based on the point of view described at the introduction the image set can be structured, shallow structured or unstructured. In the next tables (Table I, II, and III) the systems are compared from the points of view of *seeking task*, *circumstances*, and *semantic information*.

TABLE I. COMPARISON OF IMAGE RETRIEVAL AND BROWSING SYSTEMS

Name	Seeking task	
	retrieval	***browsing***
ImageRover		Yes
iMap		Yes
Photobook	Yes	Yes
QBIC		Yes
RSVP		Yes
Swirl	Yes	
Vibe	Yes	
own solution		Yes

TABLE II. COMPARISON OF STRUCTURED OF THE SYSTEMS

Name	Circumstances		
	unstructured	***swallow structured***	***structured***
ImageRover	Yes		
iMap	Yes		
Photobook	Yes		
QBIC		Yes	
RSVP	Yes		
Swirl			Yes
Vibe	Yes		
own solution			Yes

TABLE III. COMPARISON OF SEMANTICS OF THE SYSTEMS

Name	Semantic information	
	not semantic	***semantic***
ImageRover	Yes	
iMap	Yes	
Photobook		Yes
QBIC	Yes	
RSVP		Yes
Swirl		Yes
Vibe	Yes	
own solution		Yes

Based on the previous tables it can be concluded that there is a system for retrieval task, which is semantic and structured: this is the Google Image Swirl. However, in the browsing area only one system possesses these features: the solution described in this paper.

V. CONCLUSION

The aim of this paper was to develop an image browsing system, where only visual information is available. The large image set is structured in albums and the idea was to select the most representative image (thumbnail) from each album, which is not only representative, but differs from other thumbnails. In the elaborated system the selection of the most representative image is based on semantic features, which helps users in browsing process.

The successful navigation depends on the goodness of the selection of the representative image, thus a quantitative analysis has been completed. In this analysis albums in hierarchical structure are tested and the implemented solution described in this paper is compared with humans' decisions. The results have presented that the automatic solution is almost as good as humans' decisions.

A qualitative analysis has been also achieved and based on the comparison of functionalities it can be seen that our solution possesses semantic and structured features in the image browsing area.

VI. ACKNOWLEDGMENTS

The publication was supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 (FIRST) project. The project has been supported by the European Union, co-financed by the European Social Fund.

REFERENCES

[1] Chen, C., Gagaudakis, G., & Rosin, P. "Similarity-based image browsing". In Int. conference on intelligent information processing, 2000, pp. 206-213.

[2] Fan, X., Xie, X., Ma, W. Y., Zhang, H. J., & Zhou, H. Q. "Visual attention based image browsing on mobile devices". In Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on (Vol. 1, pp. I-53). IEEE.

[3] Strong, G., Hoeber, O., & Gong, M. "Visual image browsing and exploration (Vibe): User evaluations of image search tasks". In Active Media Technology. Springer Berlin Heidelberg. 2010, pp. 424-435.

[4] Jing, Y., Rowley, H., Wang, J., Tsai, D., Rosenberg, C., & Covell, M. "Google image swirl: a large-scale content-based image visualization system". In Proceedings of the 21st international conference companion on World Wide Web, ACM, 2012, pp. 539-540.

[5] Flickner, M., H. Sawhney, W. Niblack, J. Sahley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, 1995, pp. 23-32.

[6] Pentland, A., R. W. Picard, and S. Sclaroff, "Photobook: Tools for content-base manipulation of image databases," presented at SPIE Conference on Storage and Retrieval of Image and Video Databases II, San Jose, CA, 1994.

[7] Sclaroff, S., L. Taycher, and M. LaCascia, "ImageRover: A content-based image browser for the World Wide Web," presented at IEEE Content-Based Access of Image and Video Libraries, 1997.

[8] Smith, J. R. and S.-F. Chang, "Searching for images and video on the World Wide Web," Multimedia Systems, vol. 3, 1995, pp. 3-14.

[9] Schvaneveldt, R. W., F. T. Durso, and D. W. Dearholt, "Network structures in proximity data," in The Psychology of Learning and Motivation, 24, G. Bower, Ed.: Academic Press, 1989, pp. 249-284.

[10] Wang, C., Reese, J. P., Zhang, H., Tao, J., & Nemiroff, R. J. "iMap: A stable layout for navigating large image collections with embedded search". In IS&T/SPIE Electronic Imaging (pp. 86540K-86540K). International Society for Optics and Photonics, 2013.

[11] Schaefer, G., & Stuttard, M. "Intuitive Image Database Browsing to Support Image Annotation". Automatika□Journal for Control, Measurement, Electronics, Computing and Communications, 53(4), 2012.

[12] Hoque, E., Hoeber, O., Strong, G., & Gong, M. "Combining conceptual query expansion and visual search results exploration for Web image retrieval". Journal of Ambient Intelligence and Humanized Computing, 1-12., 2013.

[13] Chiarandini, L., & Jaimes, A. "Browsing-based content discovery". In Proceedings of the Designing Interactive Systems Conference. ACM. 2012,

[14] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, 2008, pp. 346-359.

[15] Baranyi P. and Csapo A. "Definition and Synergies of Cognitive Infocommunications". Acta Polytechnica Hungarica, 2012, vol. 9, no. 1, pp. 67-83.