CLAN Photo Presenter: Multi-modal Summarization Tool for Image Collections

Michal Batko, Petra Budikova, Petr Elias and Pavel Zezula
Faculty of Informatics, Masaryk University, Brno, Czech Republic
{batko,budikova}@fi.muni.cz, e1i@mail.muni.cz, zezula@fi.muni.cz

ABSTRACT

Effective management of multimedia data is becoming vital for success in the modern era of omnipresent data. Summarization tools, which allow users to quickly get the gist of a given data collection and have proven their usefulness in text domain, are now gaining popularity also in multimedia processing. However, existing algorithms provide visual-only summaries for image collections, which are difficult to index and search. This paper introduces a prototype software tool that automatically creates multi-modal summaries of personal image collections by enriching the visual collage with keyword annotation. The result is presented as a web page that allows users to browse and share the summarized data.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

Keywords

image collections, multi-modal summarization, content-based data management, clustering, automatic annotation

1. INTRODUCTION

Due to the low costs and high popularity of data capturing devices, modern man is surrounded by digital data. This situation provides great opportunities for data mining and learning, but finding a specific piece of information becomes more and more difficult. With hundreds of images in a single holiday photo collection, browsing image-by-image is tiresome. Therefore, visual summaries strive to offer a quick overview of the contents of a given collection, which is preferably both highly informative and visually appealing. Such summary can be used for personal recollection or attractive presentation of one's photos in a web gallery.

Existing works in the area of image data summarization focus mainly on identifying related images and organizing selected photos on the canvas. Some approaches attempt to visualize the whole collection [2], others perform clustering by visual, spacial or temporal similarity and then display only cluster representatives [5, 6]. In all these works, the summarization process results in some sort of image collage that can help users understand the contents of the dataset; however, it cannot be used to locate the collection on the disk or on the web. In spite of the development of content-based retrieval, most data management tools still rely on text searching. Thus, to ensure that a given image collection is findable, it is necessary to associate it with text metadata.

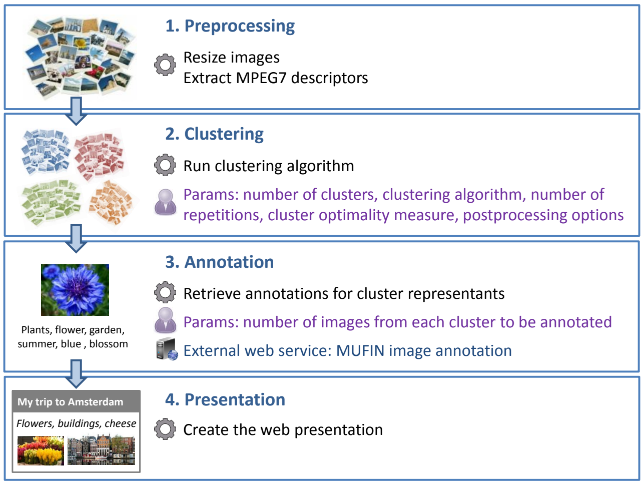
1. Preprocessing
Resize images
Extract MPEG7 descriptors
2. Clustering
Run clustering algorithm
Params: number of clusters, clustering algorithm, number of repetitions, cluster optimality measure, postprocessing options
3. Annotation
Plants, flower, garden, summer, blue , blossom
Retrieve annotations for cluster representants
Params: number of images from each cluster to be annotated
External web service: MUFIN image annotation
4. Presentation
My trip to Amsterdam
Flowers, buildings, cheese
Create the web presentation

Figure 1: Schema of image collection processing.

To address this challenge, we present a prototype tool for creating multi-modal summaries of image data. By combining traditional clustering methods and a recent image-annotation tool, the CLAN (cluster & annotate) Photo Presenter is able to create text-and-visual representations of personal photo collections. The summarized collection is displayed as an interactive web page.

2. SYSTEM ARCHITECTURE

The CLAN Photo Presenter is designed for summarization of personal photo collections, which are assumed to be unstructured sets of images with no manually created metadata. To transform such collection into a more interesting and informative shape, we apply 4 steps depicted in Figure 1.

In the first phase, salient visual features are extracted from images to allow efficient evaluations of similarity between photos. Next, we employ a clustering algorithm to divide the objects into visually similar groups and identify the main topics of the documented event. Users can choose

ICMR '14, Apr 01-04 2014, Glasgow, United Kingdom
ACM 978-1-4503-2782-4/14/04.
http://dx.doi.org/10.1145/2578726.2582623

Collection size	Clustering method	# of clusters	Preprocessing time [s]	Clustering time [s] (25 repetitions)	Annotation time [s] (5-union)	Overall costs [min]
160/500/1000	Bisecting k-means	16/20/30	76/214/471	12/138/610	229/299/418	5.3/10.9/25.0
160/500/1000	K-means	16/20/30	76/214/471	3/18/69	229/299/418	5.1/8.9/16.0
160/500/1000	Distinct kNN	25/50/100	76/214/471	3/20/82	373/696/1428	7.5/15.5/33.0

Table 1: Processing costs of selected summarization scenarios, run on a standard desktop PC.

from several clustering techniques and adjust selected parameters, e.g. the desired number of clusters.

In the next step, keyword annotations are associated with image clusters. Image-group annotation is a novel problem that can be solved straightforwardly by applying some existing annotation technique on every image in the collection and aggregating the tags assigned to individual images. However, this approach can be quite costly, if the annotation process is expensive. In the CLAN Photo Presenter, we use search-based annotation, which is suitable for wide application domains but rather costly. Therefore, only cluster representatives are used to acquire the descriptions. To ensure reasonable annotation quality, several images from each cluster are annotated and their descriptions aggregated. The number of cluster representatives can be adjusted by users.

Finally, the annotated clusters are presented in an easy-to-use web interface that allows users to explore the collection. Each cluster is represented by a centroid photo which, when activated by mouse pointer, displays descriptive keywords of the cluster and a video-like sequence of clustered images.

2.1 Technologies

The CLAN Photo Presenter is a desktop application written in Java that utilizes our previously developed technologies for similarity searching, clustering, and annotation processing. Visual similarity of images is evaluated by five MPEG7 global visual descriptors, namely the Scalable Color, Color Structure, Color Layout, Edge Histogram, and Region Shape. The clustering phase offers the standard k-Means and the Bisecting k-Means algorithms [3] and two variants of the Distinct kNN method [4]. As all these methods are non-deterministic, the clustering phase may be performed repeatedly to increase the quality of results, using Davies Bouldin Index, Silhouette Coefficient, or Dunn Index to select the optimal clustering [3]. Furthermore, two postprocessing methods are available to deal with anomalies such as single-member clusters or outlier images. The annotation phase exploits the MUFIN Image Annotation software [1].

3. PERFORMANCE EVALUATION

Different settings of clustering and annotation techniques influence both effectiveness and efficiency of the summarization task. In this section, we provide a basic evaluation of the processing costs and result quality, which can serve as a guideline for users who want to apply non-default methods.

The time complexity of the summarization process is determined by the costs of image preprocessing, clustering, and annotation. The preprocessing and annotation costs per image are constant, whereas the clustering complexity is quadratic with respect to the collection size. For smaller collections, preprocessing costs (approximately 0.5 s per 4096x3072 px image) and annotation forming (1.5-3 s per annotated image) are prevalent. Clustering costs become significant for larger collections, especially when the Bisecting k-means algorithm is applied. The Distinct kNN method is fast, but typically produces a high number of clusters, which increases the annotation phase costs. Selected measurements on three test collections can be found in Table 1.

In contrast to efficiency, summarization quality is very difficult to measure. Clustering methods provide good results for well-clustered training collections, but on real-world data the quality varies – visually similar objects are likely to be grouped together, but the coherence of individual clusters is often low. To achieve better results, a correct choice of the number of clusters is important; with default settings, the Distinct kNN technique has the highest chance of providing coherent clusters, but the number of clusters will be high.

In annotation, the average precision of the MUFIN Image Annotation tool is about 60 % [1]. The group-annotation quality depends on the coherence of a given group of images and the number of representatives used for annotation processing. On a test collection of real-world data with well-defined topics, we obtained an average quality of 70 % relevant keywords with 7 representatives per cluster.

4. CONCLUSION

This paper presents a prototype multi-modal summarization tool for image collections, which allows users to perceive their photos in a novel way. The visual output of the CLAN Photo Presenter can be used to show one's adventures to friends, whereas the keyword summary improves the findability of the collection. Both the application and different summaries created by the CLAN Presenter can be found at `http://disa.fi.muni.cz/prototype-applications/clan/`.

Acknowledgments

This work was supported by the Czech national research project GBP103/12/G084.

5. REFERENCES

[1] M. Batko, J. Botorek, P. Budíková, and P. Zezula. Content-based annotation and classification framework: A general multi-purpose approach. In *Proceedings of IDEAS 2013*, pages 58–67. ACM, 2013.

[2] T. Liu, J. Wang, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum. Picture collage. *IEEE Transactions on Multimedia*, 11(7):1225–1239, 2009.

[3] L. Rokach. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*, pages 269–298. Springer US, 2010.

[4] T. Skopal, V. Dohnal, M. Batko, and P. Zezula. Distinct nearest neighbors queries for similarity search in very large multimedia databases. In *Proceedings of WIDM '09*, pages 11–14. ACM, 2009.

[5] L. Tan, Y. Song, S. Liu, and L. Xie. ImageHive: Interactive content-aware image summarization. *IEEE Computer Graphics and Apps.*, 32(1):46–55, 2012.

[6] L. Zhang and H. Huang. Hierarchical narrative collage for digital photo album. *Comput. Graph. Forum*, 31(7-2):2173–2181, 2012.