Marcel Worring
University of Amsterdam, The Netherlands

Easy categorization of large image collections by automatic analysis and information visualization

Abstract: A large part of our history as well as our daily lives is captured in visual data. Understanding visual collections requires careful categorization to reveal expected as well as hidden relations. Performing this categorization manually is a demanding and cumbersome process. On the other hand automatic methods still have limitations in performance. An optimal approach brings together the power of automatic bulk categorization with detailed and careful expert annotation. In this paper we show how advanced visualizations can aid the categorization and subsequent exploration processes.
Keywords: multimedia analytics; visual concepts; automatic categorization

1. Introduction

Image collections can carry a tremendous amount of information in diverse application domains. For art historians images of paintings form the core data source for studying history. For social scientists, visual data on social sharing sites like FaceBook, YouTube or Flickr contain a wealth of information about current issues in society and these images can steer a chain of reactions through the social network. Another field in which images play a major role is biodiversity in which images of animals and the location where these pictures are taken provide key statistics for the population. In all these applications the content of images is, however, not enough to reveal all the important characteristics.

An important step in understanding an image collection is categorizing each individual image by assignment of appropriate labels. It is the labelling that gives the images meaning, relating it to the task and context of the user. These labels can range from simple descriptors of the elements in the image, to location and time where the picture is taken, to elaborate descriptions of every detail in the image, or subjective interpretations. Finally, relations among the images in the dataset as well as connection to other datasets are ways to extend the richness of the dataset. In many cases this categorization and labelling process is manual, a cumbersome task when the collection is large.

Ideally the categorization would be automatic where the system assigns the proper labels to the content in an autonomous fashion. Automatic methods, however, have their limitations. The main cause of this is the so called semantic gap defined as:

> "The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation." (Smeulders, Worring, Santini, Gupta, Jain 2000: 1353).

This doesn't mean that automatic methods fail in all cases. In recent years a lot of progress has been made and new methods are able to provide a good indication of elements which we often encounter in an image (Snoek, Smeulders 2010). These methods are based on learning a detector from a lot of positive and negative examples. Concepts for which these methods are successful are those where examples are easy to obtain and where the visual variety in the concept is limited. Typical examples of concepts with good detection performance are sky, car, aeroplane, face, human walking, and seascape. For more complex concepts, performance is still very limited. But where the automatic labelling might not be perfect, a computer can perform the categorization with amazing speed.

In contrast to the speed of the computer human labelling is slow. Yet humans are able to understand even the most complex semantic concepts in the image. Even when they haven't encountered a concept before they infer information from the context of the concept and thus prune the set of possible interpretations quickly and will soon find a plausible interpretation. Humans are also experts in understanding abstract concepts for which the visual appearance is not consistent at all. For example consider the word democracy and think of the many different ways this might be depicted in an image. In addition humans can associate the content of the images to other information based on a variety of different clues. Humans may be slow, but their interpretations are valuable and much more subtle than what a computer can ever derive.

To move forward in the analysis of image collections the best is to bring together the best of both worlds. Doing so requires intuitive ways for the user to interact with the data and the best way to do so is to use information visualization techniques as the bridge connecting them. Such approaches in which humans and machines cooperate in understanding complex multimedia sources has recently been coined Multimedia Analytics (Chinchor, Thomas, Wong, Christel, Ribarsky 2010) and this forms a promising avenue in truly getting an understanding of large and complex visual collections.

In this paper we will elaborate on the steps we have taken towards Multimedia Analytics showcasing the work we did. Thus, we provide a clearly very biased view on the field, but hope to give some understanding of the underlying concepts and ideas. The paper is organized as follows. In section 2 we consider in more detail what a category actually is, especially in the context of approaches where automatic and manual methods come together and come to two main approaches namely clustering based and semantics based. From there in section 3 we reflect on how categorization processes are related to the understanding of the content of the collection. Then in section 4 and 5 we elaborate on two classes of systems that we have developed.

2. Categories and automatic methods

According to the Oxford dictionary a category is defined as "*a class or division of people or things regarded as having particular shared characteristics*". Putting this into the context of image collections indicates that categories in such collections can be many. Images can share similar appearance or a similar semantic interpretation, but could also simply have similar characteristics in terms of their metadata. In the context of this paper we make this precise by distinguishing three different types of interest:

- *Similar appearance*: images sharing visual characteristics independent of their semantics.
- *Semantically similar appearance:* images having the same semantic interpretation for a user where the semantic similarity is reflected in shared visual characteristics for different instances in the category.
- *Abstract similarity:* images having the same semantic interpretation for the user where the semantic similarity is NOT reflected in similar appearance.

Following the definition of the semantic gap each of these yields a different scenario of use and cooperation between the computer and the user. When images have similar appearance, the computer can automatically cluster them based on their visual characteristics. This process can in principle be performed in a completely autonomous way by the system. The user can take the results and use it in further explorations of the dataset. With abstract similarity the situation is reversed as only the user is capable of assigning images to the right category. For semi-interactive approaches the class of semantically similar appearance is most interesting. In this scenario the user provides positive as well as negative examples of the concept. Based on these examples the system can automatically derive a model of the concept by employing the similarities among the elements in one example set versus dissimilarities between the two sets. This can be a single shot process, but much better results can be obtained when the user reflects on the results of the model, providing new positive and negative examples to improve the model. With such models new unseen images can automatically be analysed for the presence of the specific concept learned.

3. Understanding as incremental categorization

Understanding a large collection of images is a difficult task. When the ultimate aim is to get insight from the collection the user has to engage in interactive sessions with the system. According to (North 2006) insight is *complex* involving several elements of the dataset in a synergetic way, *deep* building up over time, accumulating and building on itself to create depth, unexpected in the sense of being unpredictable, serendipitous, and creative, and finally is *relevant* in the particular context of the domain. True multimedia analytics solutions would thus have to cater for each of those dimensions.

Categories can play an essential role in getting insight in image collections. A user can start off with a simple set of categories and use those to describe the elements in the collection. When going deeper into the matter new subcategories may emerge or new groupings of categories into more general concepts might be found. Thus, by using categories in an incremental manner users can build up deep knowledge and create unexpected domain dependent categories.

The data-driven clustering and semantic concepts detected in the various images form the building blocks in this insight gathering process. We have made a number of steps in this direction with various image based browsing and categorization systems which we will now describe.

4. Cluster based analytics

As a first example of a cluster based analytics system we consider the GalaxyBrowser (Nguyen, Worring, 2008) as depicted in Figure 1.
The basis for this browser is the similarity among the images in the collection. By performing a similarity based clustering of the data into a hierarchy of unnamed categories, the set is decomposed into parts while as the same time is given a tree based structure. At every level of the tree, the collection is represented by a set of images which are the centres of the underlying cluster. The set of representative images at a certain level are displayed on the screen in such a way that a balance is obtained between three conflicting criteria:

1. Providing a good overview of the underlying clusters by showing a large as possible set of images at the same time.
2. Preserving as well as possible the similarities among images (which live in a high dimensional space) in the 2D display such that similar images are placed close to each other.
3. Assuring the visibility of individual images on the screen.

For images with semantically similar appearance, the visualization can speed up the categorization process considerably as images with similar appearance are shown close to each other. So several images can be categorized in one interaction step by giving them all the same label.
In (Nguyen, Worring, 2008b) active learning based extensions of the GalaxyBrowser have been described. In such a scenario the user indicates which images in the display belong to the category sought and the system then assumes that non-selected images are irrelevant. Based on this feedback the system then tries to optimize its model for categorizing images which don't have a category yet. In contract to relevance feedback based method the system does not visualize the current result of the model. Instead it shows the images in the collection which are most informative for the system i.e. examples which are on the border of the category. By getting feedback from the user on those, the system can quickly improve its model.

Figure 1: *Overview of the GalaxyBrowser. The main window provides a similarity based visualization of a subset of the collection. The top right window provides the same type of visualization, now for the whole collection using dots instead of images.*

The Chronosphere system (Worring, Engl, Smeria, 2012) takes a different approach to cluster based visualization. This browser was designed particularly for the domain of digital forensics. In such a domain the metadata of the images, in this particular case time and the name of the file and its directory are important clues for categorizing the sets. So we combine visualizations of all of these in one system. The data analysis starts with an appearance driven clustering of the data into a small (10-15) set of clusters. Instead of similarity based visualization we use a graph based representation here. To be precise we use a pathfinder based algorithm to find and visualize the most important edges between clusters (Chen, Morris, 2003). In the other visualizations for each cluster the time distribution is visualized whereas the relation between directory and number of elements of each cluster is visualized as a heatmap. Small icons are used to represent each cluster so users can easily track them between the different visualizations. By creating filters on either cluster, time, or directory the user is able to quickly find images in the same category.

Figure 2: *Overview of the ChronoSphere system. With the top-left window showing cluster distribution over time, the bottom-left holds the graph based visualization and the rightmost window the heatmap showing cluster distribution over different directories.*

5. Semantics based analytics

The final system we describe in this paper is the MediaTable (de Rooij, Worring, van Wijk, 2010) which is particularly suited for collections with semantically similar appearance among the images (see Figure 3). Central to this system is the automatic derivation of scores for the presence of a large collection of semantic concepts. From there most semantics based image retrieval systems follow the standard query-response paradigm. After selecting a concept the system returns a list of images according to this score which are then presented as either a list or grid of images. In the MediaTable every row corresponds to one image in the collection with all its scores for the various concepts. This is again visualized as a kind of heatmap showing results for a large set of images and many concepts. The three core operations in MediaTable are sorting according to a concept, filtering based on concept scores (or tags corresponding to the images), and selecting images and putting them in a category (for which the metaphor of buckets) is used. Through the heatmap correlations between the different concepts are revealed. In (deRooij, Worring 2013) we made the system more intelligent and a true multimedia analytics solution by adding unobtrusive relevance feedback. In this mode, the system continuously observes the content of the buckets and in an autonomous fashion suggests new images for each of the buckets. In the reference we showed that this significantly improved the efficiency compared to a baseline of no automatic feedback processing as well as a mode where the user explicitly had to ask for additional suggestions from the system for a specific category.

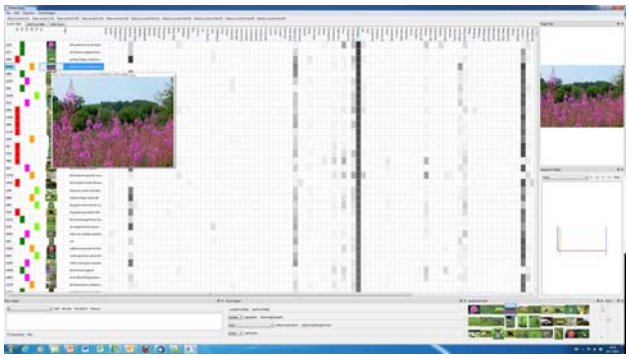

Figure 3: *Overview of the MediaTable system.*

6. Conclusion

Image collections can be an important source of information and categorization is important to incrementally gain insight in the collection. The size and characteristics of the data set dictate whether and how automatic methods and information visualization come together in an optimal way in multimedia analytics solutions.
Multimedia analytics is a relatively new field and methods are slowly appearing. In this paper we have illustrated a number of approaches that we have followed to support the categorization process in various application domains with multimedia analytics solutions. The underlying techniques are all different. As a whole, however, they cover a range of support tools for the user faced with a large collection of uncategorized images. The methods presented are first steps into this challenging field.
Only through the extensive use of tools as the ones described within various domains, will we be able to really identify the possibilities and limitations of multimedia analytics solutions. We believe that multimedia analytics has the potential to provide true domain impact.

Acknowledgment

Many people have contributed to the work described in this paper. In particular Cees Snoek, Arnold Smeulders, Giang Nguyen, Ork de Rooij, Jack van Wijk, Daan Odijk, Andreas Engl, Camelia Smeria and Dennis Koelma.

References

Chen, C., Morris,S. (2003), Visualizing evolving networks: minimum spanning trees versus pathfinder networks. In: *IEEE Symposium on Information Visualization* pp. 67-74.

Chinchor, N.A., Thomas, J.J., Wong P.C., Christel, M.G, Ribarsky, W (2010). Multimedia Analysis + Visual Analytics = Multimedia Analytics In: *IEEE Computer Graphics and Applications 2010, volume 30(5),* pp. 52-60.

Nguyen, G.P., Worring, M., (2008). Interactive access to large image collections using similarity-based visualization In: *Journal of Visual Languages & Computing 19 (2), pp. 203-224.*

Nguyen G.P., Worring M. (2008b) Optimization of Interactive Visual Similarity-Based Search. In: *ACM Transactions on Multimedia Computing, Communications and Applications, Volume 4 (1), pp. 1-23.*

North C. (2006). Towards measuring insight. In: *IEEE Computer Graphics and Applications 2006, volume 26(3),* pp. 6-9.

de Rooij O., Worring M., van Wijk J.J. (2010). MediaTable: Interactive Categorization of Multimedia Collections. In: *IEEE Computer Graphics and Applications 2010, volume 30(5),* pp. 42-51.

de Rooij O., Worring M. (2013). Active Bucket Categorization for High Recall Video Retrieval. In: *IEEE Transactions on Multimedia, Volume 15(4), pp. 898-907.*

Smeulders, A. W. M.; Worring, M., Santini S., Gupta A., Jain R. (2000). Content Based Image Retrieval at the End of the Early Years. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence 2000, volume 22(12),* pp. 1349-1380.

Snoek C.G.M., Smeulders A.W.M. (2010). *Visual-Concept Search Solved?, IEEE Computer, vol. 43(6), pp. 76-78.*

Worring M., Engl A., Smeria C. (2012), A Multimedia Analytics Framework for Browsing Image Collections in Digital Forensics. In: *ACM Conference on Multimedia, 2012*.

About authors

Marcel Worring is an associate professor in the Informatics Institute of the University of Amsterdam. His research focuses on Multimedia Analytics, the integration of Multimedia Analysis and Information Visualization into a coherent framework which yields more than its constituent components. He is one of the initiators of the show me the data conference series in Amsterdam. The series has a direct connection to a unique course in information visualization where new media students from the humanities, designers, and computer science students come together to do challenging projects.