

# Towards Large-Scale Multimedia Exploration

Christian Beecks<sup>•</sup>, Tomáš Skopal<sup>◦</sup>, Klaus Schöffmann<sup>\*</sup>, Thomas Seidl<sup>•</sup>

<sup>•</sup>Data Management and Data Exploration Group, RWTH Aachen University, Germany

<sup>◦</sup>SIRET Research Group, Charles University in Prague, Czech Republic

<sup>\*</sup>Distributed Multimedia Systems Group, Klagenfurt University, Austria

<sup>•</sup>{beecks,seidl}@cs.rwth-aachen.de, <sup>◦</sup>skopal@ksi.mff.cuni.cz, <sup>\*</sup>ks@itec.uni-klu.ac.at

## ABSTRACT

With the advent of the information age and the increasing size and complexity of multimedia databases, the question of how to support users in getting access and insight into those large databases has become immensely important. While traditional content-based retrieval approaches provide query-driven access under the assumption that the users' information needs are clearly specified, modern content-based exploration approaches support users in browsing and navigating through multimedia databases in the case of imprecise or even unknown information needs. By means of interactive graphical user interfaces, exploration approaches offer a convenient and intuitive access to unknown multimedia databases which becomes even more important with the arrival of powerful mobile devices. In this paper, we formulate challenges of user-centric multimedia exploration with a particular focus on large-scale multimedia databases. We claim that adaptability and scalability should be researched on both conceptual as well as technical level in order to model multimedia exploration approaches which are able to cope with millions of multimedia objects in near-realtime.

## 1. INTRODUCTION

In the presence of ever-increasing multimedia databases, supporting users in getting access and insight into such large databases remains a challenging issue. For this purpose, *content-based retrieval approaches* [9, 12, 15] support users in searching and finding particular objects or topics in a content-based way by providing them with a static database's subset of potentially relevant objects with respect to their information needs. In this way, those approaches can only be successful when the users' information needs are clearly specified. When these information needs are imprecise or even unknown, *relevance feedback* techniques [27, 19] may assist the user to refine the search process and help him or her to find the desired content. However, refining the search process and, thus, sharpen the user's information need by means of relevance feedback techniques is less useful when he



Figure 1: Example multimedia exploration systems

or she doesn't know what to search for and how the desired content looks like.

In order to counteract the general problem of unknown information needs and to support users to browse and navigate through multimedia databases, numerous *content-based multimedia browsing and exploration systems* [13, 22, 17, 21, 10] have been proposed in the last decade. They overcome the limitations of conventional content-based retrieval approaches, i.e., their low degree of flexibility and interactivity, and frequently support the exploration process by means of attractive interactive graphical user interfaces, such as those depicted in Figure 1 for image and video browsing systems.

In general, all of these multimedia exploration systems follow the same general structure of an *exploration process* [5], which is illustrated in Figure 2. The exploration process is initialized by mapping a meaningful subset of database objects into a low-dimensional visualization topology, of which the *similarity-based layout* [17] is a generic and frequently encountered approach. In doing so, the exploration systems should adhere to the three requirements proposed by Nguyen and Worring [17]: the visual topology should reflect the similarity relationships among the multimedia objects, the user interface should offer the possibility to get an overview of the entire multimedia database, and the displayed objects should be arranged in a non-overlapping way. As can be seen in the Figure 1, most of the proposed systems satisfy these three requirements to some extent.

After having generated an initial visualization topology, users can start browsing and navigating through the underlying database content by interacting with the exploration system and particularly with the displayed objects. They can zoom in or out, they can shift or rotate the visualization topology, or they can issue a concise information need to extend/reduce the set of displayed objects. In particular, adding new objects to the visualization can be implemented, for instance, by simply navigating to some point of interest and just clicking it with the mouse, as studied by Fan et al. [11], or by making use of hot-keys. In a more mobile scenario where users are browsing multimedia databases with their handheld devices, the aforementioned operations, to which

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. This article was presented at:

*Fifth International Workshop on Ranking in Databases (DBRank 2011).*

Copyright 2011.

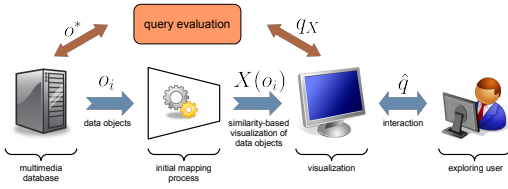


Figure 2: Structure of an exploration process

we refer to as *exploration queries*, can also be implemented by haptic interactions, for instance by touching the screen.

## 1.1 Motivation

Although a lot of research has been conducted on providing users interactive and intuitive browsing interfaces and systems, there is still a lack of processing exploration queries over large multimedia databases efficiently and even in near-realtime. In particular, none of the approaches mentioned above is able to cope with millions of multimedia objects in near-realtime. Our previous attempts [5, 1], which show how to improve the content-based exploration process by means of optimal multistep architectures [23] or exploration-adapted index structures, such as the M-tree [8], are neither sufficient to cope with large-scale multimedia databases nor with user-centric process adaptations.

## 1.2 Paper contributions

In this vision paper, we suggest to fundamentally research the user-centric process of content-based multimedia exploration with a particular focus on large-scale databases. We emphasize the roles of adaptability and scalability of content-based multimedia exploration on both conceptual as well as technical level in order to develop a user-centric multimedia exploration approach which is able to cope with millions of multimedia objects in near-realtime. To this end, we focus at defining the process of anytime exploration and the adaptation of that process to continuously changing user preferences. Furthermore, by investigating novel adaptive indexing methods supporting near-realtime query processing, we consider to establish anytime browsing of large multimedia databases with customized similarity measures.

## 2. EXPLORATION PROCESS

The design of multimedia exploration systems is based on a higher-level architecture for which the elementary operations are simple similarity queries provided by the underlying multimedia database. Because interactivity of the multimedia exploration process turns in massive triggering of similarity queries, their efficient implementation is one of the key requirements. Although similarity queries are easy to define and generate in a multimedia exploration framework, their efficient database implementation is nontrivial. Adaptive distance-based similarity measures for multimedia data [3], for instance the Earth Mover’s Distance [18] or the Signature Quadratic Form Distance [4], are often expected to be computationally expensive. Consequently, processing exploration queries through a naive sequential scan of the underlying multimedia database, as shown in our previous work [5] in the context of content-based multimedia exploration, is not feasible for large-scale multimedia databases comprising millions of data objects.

At the multimedia database level, there have been proposed approaches to efficient evaluation of simple similarity

queries, such as *range* or *k-Nearest Neighbor* (kNN) queries. Historically, the database research considered just one specific model of similarity measures – the metric space model. As metric distances obey the metric postulates, they can be indexed by use of *metric access methods* (MAMs), also called metric indexes [6, 26, 20]. MAMs have been designed to provide efficient processing of similarity queries. However, most of the MAMs assume a static metric distance [25]. While this is not an obstacle for content-based applications where range and kNN queries are already the desired means of retrieval, it becomes a problem when the MAMs should be a part of a more complex retrieval framework. Recently, we have begun to focus our research towards efficient MAM-based indexing of the Signature Quadratic Form Distance which outperforms other state-of-the-art distance-based similarity measures [3]. We have investigated the distance’s *indexability* [2] and shown that this distance is a *ptolemaic metric* [16] which is indexable by ptolemaic pivot tables [14]. However, all of these approaches were developed to process simple range or kNN similarity queries. Thus, in their initial form, they are not suitable to be directly coupled with flexible user-centric multimedia exploration systems supporting anytime browsing at large scale.

In the following section, we summarize the challenges related to content-based multimedia exploration and argue for the necessity of treating the research in this area as a complex problem that cannot be solved by a simple combination of existing low-level techniques.

## 3. CHALLENGES

**Unifying theory.** Although numerous content-based exploration systems have been proposed, there is still a lack of a *unifying theory* which generally models the exploration process and particularly the arising exploration queries. A theoretical model would allow to evaluate and compare existing as well as prospectively developed exploration approaches. Furthermore, it would allow to transform exploration queries arising during the exploration process into a set or stream of simple similarity queries which can be efficiently processed by generic database indexing structures.

**User-centric adaptation.** It is immensely important to develop and research effective and efficient database indexing methods which can cope with massive amounts of continuously changing similarity queries generated during the exploration process. In order to allow the user the highest possible flexibility, i.e., interactive browsing and similarity adaptation at any time within the content-based exploration process, methods that support continuous adaptation to individual and dynamically changing information needs during the exploration process must be investigated. In fact, none of the approaches mentioned above supports the continuous adaptation of the exploration process’ underlying similarity model during the browsing process.

**Near-realtime index support.** So far, database indexing structures designed for efficient similarity search were optimized for simple range or kNN queries. In contrast, the research in multimedia retrieval dealing within a complex architecture, such as multimedia exploration, did not pay attention to the database point of view, implicitly assuming the sequential scan of the database as sufficient. However, when it comes to a real-world application of content-based multimedia exploration that generates massive amounts of

similarity queries at real time, it is important to take into account the database performance and to develop techniques for near-real time data access at a large scale. Thus, an optimization of the interplay between user-centric exploration queries and data-centric indexing methods is needed.

**Adaptive index support.** Existing database indexing methods assume a strongly constrained similarity model, where the similarity measure cannot be freely adjusted during the index life time. In most cases, the similarity measure is restricted to a metric distance that is static (cannot change at all). This becomes a problem for multimedia exploration, as the similarity model should be adapted to individual user requirements, resulting in a dynamic (and possibly non-metric) distance function. Although there have been proposed general approaches to efficient search using non-metric similarity measures [24], their usage in the multimedia exploration scenario is not feasible due to the remaining assumption on a static distance. There have also been proposed approaches supporting dynamic metric functions [7], these methods, however, are limited to a very narrow class of metric functions (and their efficiency is guaranteed just for slight changes in the function).

## 4. CONCLUSIONS

In this vision paper, we formulated challenges of user-centric multimedia exploration with a particular focus on large-scale multimedia databases. We stated that adaptability and scalability should be investigated and researched on both conceptual as well as technical level in order to model and develop content-based multimedia exploration approaches which are able to cope with millions of multimedia objects in near-realtime.

## Acknowledgments

This research has been supported by the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 686 and by Czech Science Foundation project GACR 202/11/0968.

## 5. REFERENCES

- [1] C. Beecks, P. Driessen, and T. Seidl. Index support for content-based multimedia exploration. In *Proc. ACM Multimedia*, pages 999–1002, 2010.
- [2] C. Beecks, J. Lokoč, T. Seidl, and T. Skopal. Indexing the signature quadratic form distance for efficient content-based multimedia retrieval. In *Proc. ACM ICMR*, 2011.
- [3] C. Beecks, M. S. Uysal, and T. Seidl. A comparative study of similarity measures for content-based multimedia retrieval. In *Proc. IEEE ICME*, pages 1552–1557, 2010.
- [4] C. Beecks, M. S. Uysal, and T. Seidl. Signature quadratic form distance. In *Proc. ACM CIVR*, pages 438–445, 2010.
- [5] C. Beecks, S. Wiedenfeld, and T. Seidl. Improving the efficiency of content-based multimedia exploration. In *Proc. ICPR*, pages 3163–3166, 2010.
- [6] E. Chávez, G. Navarro, R. Baeza-Yates, and J. L. Marroquín. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, 2001.
- [7] P. Ciaccia and M. Patella. Searching in metric spaces with user-defined and approximate distances. *ACM TODS*, 27(4):398–437, 2002.
- [8] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proc. VLDB*, pages 426–435, 1997.
- [9] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40:5:1–5:60, 2008.
- [10] O. de Rooij and M. Worring. Browsing video along multiple threads. *IEEE Transactions on Multimedia*, 12(2):121–130, 2010.
- [11] J. Fan, D. A. Keim, Y. Gao, H. Luo, and Z. Li. Justclick: personalized image recommendation via exploratory search from large-scale flickr images. *IEEE Trans. Cir. and Sys. for Video Technol.*, 19:273–288, 2009.
- [12] P. Geetha and V. Narayanan. A survey of content-based video retrieval. *Journal of Computer Science*, 4(6):474–486, 2008.
- [13] D. Heesch. A survey of browsing models for content based image retrieval. *MTAP*, 40:261–284, 2008.
- [14] M. L. Hetland. Ptolemaic indexing. [arXiv:0911.4384 \[cs.DS\]](https://arxiv.org/abs/0911.4384), 2009.
- [15] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *TOMCCAP*, 2:1–19, 2006.
- [16] J. Lokoč, M. L. Hetland, T. Skopal, and C. Beecks. Ptolemaic indexing of the signature quadratic form distance. In *Proc. SISAP*, pages 9–16, 2011.
- [17] G. P. Nguyen and M. Worring. Interactive access to large image collections using similarity-based visualization. *J. Vis. Lang. Comp.*, 19:203–224, 2008.
- [18] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover’s Distance as a Metric for Image Retrieval. *IJCV*, 40(2):99–121, 2000.
- [19] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. Cir. and Sys. for Video Technol.*, 8(5):644–655, 1998.
- [20] H. Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.
- [21] G. Schaefer. A next generation browsing environment for large image repositories. *MTAP*, 47:105–120, 2010.
- [22] K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and J. M. Jose. Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1):018004, 2010.
- [23] T. Seidl and H.-P. Kriegel. Optimal multi-step k-nearest neighbor search. In *Proc. ACM SIGMOD*, pages 154–165, 1998.
- [24] T. Skopal. Unified framework for fast exact and approximate search in dissimilarity spaces. *ACM TODS*, 32(4):1–46, 2007.
- [25] T. Skopal. Where are you heading, metric access methods? a provocative survey. In *Proc. SISAP*, pages 13–21, 2010.
- [26] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search: The Metric Space Approach*. Springer, 2005.
- [27] X. Zhou and T. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems*, 8(6):536–544, 2003.