

A DATABASE SYSTEM FOR MULTIMEDIA ANALYTICS AND ANALYSIS

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Ralph Marc Philipp Gasser

Basel, 2022

Zusammenfassung

Abstract

Acknowledgements

Good luck.

This work was partly supported by the Swiss National Science Foundation, which is also thankfully acknowledged.

Contents

Zusammenfassung	iii
Abstract	v
Acknowledgements	vii
List of Figures	xi
List of Tables	xiii
I Introduction	1
1 Introduction	3
1.1 Working with Multimedia Data	4
1.1.1 Multimedia Retrieval	4
1.1.2 Multimedia Analysis	5
1.1.3 Multimedia Analytics	5
1.2 Research Gap and Objective	5
1.2.1 Motivation	6
1.2.2 Research questions	6
1.3 Contribution	6
2 Applications and Use Cases	9
2.1 Use case 1: Multimedia Retrieval System	9
2.2 Use case 2: Analysis of Social Media Streams	9
2.3 Use case 3: Magnetic Resonance Fingerprinting (MRF)	9
3 Related Work	11
II Foundations	13
4 On Multimedia Analysis and Retrieval	15
4.1 Multimedia Data and Multimedia Collections	15
4.2 Similarity and the Vector Space Model	15
4.3 Approximate Neares Neighbor Search	15

4.4	Online Multimedia Analysis Pipeline	15
4.5	Multimedia Retrieval Pipeline	15
4.6	Beyond Similarity	16
5	On The Design of a Database Management System	17
III	Dynamic Multimedia Data Management	19
6	Modelling a Database for Dynamic Multimedia Data	21
6.1	Architecture	21
6.2	Extended Model for Similarity Search	21
6.3	Management of Secondary Indexes	21
6.4	A Cost Model for Retrieval Accuracy	21
6.5	Further considerations	21
7	Cottontail DB	23
IV	Discussion	25
8	Evaluation	27
9	Conclusion & Future Work	29
	Appendix	31
	Bibliography	33
	Curriculum Vitae	35
	Declaration on Scientific Integrity	37

List of Figures

List of Tables

1.1	List of research questions resulting from challenging assumptions one, two and three.	7
-----	--	---

PART I

Introduction

1

Introduction

The term *multimedia* describes the combination of different forms of content – also called *modalities* – into a single, sensory experience that carries a higher level semantic. Those modalities include but are not limited to textual, visual (e.g., images or videos) or aural (e.g., music, sound effects, speech) input. In addition, signals produced by various sensors and devices can also be seen as media modalities, even though direct experience by a human consumer may require pre-processing through specialized hard- and software.

People interact with multimedia on a daily basis when watching videos on Netflix or YouTube, when listening to music on Spotify or when browsing a private image collection on their laptop. With continued digitization in all areas and aspects of our lives, one consider most of this mutlimedia content to be digital content. Multimedia data is therefore ubiquitous and touches almost every aspect of our social and economic lives.

Multimedia content makes up a large part of today's Internet and its mere omnipresence makes it a major driving force behind data growth in the Internet as a whole, as both volume and variety increases steadily. A large contributing factor here are social media platforms, where users act as consumer and producer of content. Current estimates ¹ suggest, that there are roughly 4.66 billion active Internet users worldwide, of which 4.2 million can be considered active social media users. Facebook alone contributed to 144000 uploaded images per minute in 2020. And many more of these platforms, such as *Instagram* org *Twitter* serve billions of users with mixed, self-made content involving text, images, videos or a combination thereof. A similar study found ², that by 2025 we will produce a yearly amount of 175 Zettabytes (i.e, 10^{21} bytes).

¹ Source: Statista.com, "Social media usage worldwide", January 2021

² Source: Statista.com, "Big Data", January 2021

Given these numbers, the need for efficient and effective tools for *managing, manipulating, searching, exploring* and *analysing* multimedia data corpora comes at no surprise. Unlike traditional data collections that contain only text and numbers, however, this multimedia data is unstructured and more complex to handle, which gives rise to need for special tools that allow such analysis to take place.

1.1 Working with Multimedia Data

1.1.1 Multimedia Retrieval

Retrieval, i.e., the act of searching and finding an item of interest in a large collection, is very different for multimedia data than it is for structured data in a relational database [BVB⁺07]. The structure of a relational database is a given and using languages like SQL, a user can specify exactly what elements from the database should be selected using predicates that either match or don't match.

Retrieving multimedia data, on the other hand, comes with many challenges. Firstly, we usually consider different representations of a multimedia item when doing multimedia retrieval, each of which comes with a certain loss of information, a problem usually referred to as the *semantic gap*. Examples for such representations could be a textual description or a feature vector. Secondly, different types of media exhibit different relationship with time. Media items can thus be classified as either being *static*, i.e. independent of time (e.g., in image), or *dynamic* (e.g, a video). And finally, some types of media are multimodal by nature such as, for example, videos. Those media types exhibit both visual as well as aural content, both of which can in theory carry different and sometimes even conflicting information.

In addition, all of the mentioned problems are actually two-fold, since they usually apply both when deriving storable representations from a media item as well as when expressing an information need in form of a queries.

1.1.2 Multimedia Analysis

1.1.3 Multimedia Analytics

1.2 Research Gap and Objective

The starting point for the research described in this thesis is the current state-of-the-art for data management in multimedia retrieval and analysis as briefly touched upon in the previous sections. Starting from the models and solutions proposed in [GS16; Gia18] and motivated by the “Ten Research Questions for Scalable Multimedia Analytics” [JWZ⁺16], this thesis challenges three basic assumptions currently employed and operated upon and explores the ramifications of doing so, with the higher level goal of furthering convergence between research conducted in *multimedia retrieval*, *multimedia analysis* and *multimedia analytics* on the one hand, and classical *database systems* on the other. The assumptions that are being challenged are namely:

Assumption 1: Staticity of data collections Most multimedia retrieval systems today make a clear distinction between an *offline* phase during which media items are analysed, features are generated and derived data is ingested into a data management system, and an *online* phase, during which queries of the database take place. Usually, no changes to the data collection are being made during the online phase. This formal model is advertised by both [Gia18] and [Ros18] and to the best of our knowledge, most existing systems explicitly or implicitly function in this manner. This simplification allows for time consuming processes related to feature extraction and indexing to take place separated from any concurrent query activities and eases requirements on transaction isolation. While this is convenient from a perspective of system design, such a mode of operation is limiting when facing data that changes frequently, as is the case, for example, when doing real-time analysis or when having an application with CRUD support.

Assumption 2: Similarity search is nearest neighbor search Multimedia retrieval relies strongly on a notion of similarity search that is usually expressed as finding the k nearest neighboring feature vectors $\vec{v}_{i \in [1, k]} \in C$ to a query vector $\vec{q} \in \mathbb{R}^d$ in a collection $C \subset \mathbb{R}^d$ given a certain distance function. Very often, metrics such as the Euclidean or the Manhattan distance are employed for this comparison. This is also referred to as the *vector space model*. While this model is very concise and rather simple, it merely allows for ranking of

potential results and finding the relevant or desired item(s). This model is, however, limited when looking at tasks such as summarization or structuring of data, as required for multimedia analysis and analytics. Furthermore, there are cases in which features are not real-valued vectors and similarity is not directly related to proximity.

Assumption 3: User defines execution Database management systems usually evaluate and select the execution plan for an incoming query during a step that is referred to as *query planning*. The underlying assumption here is that the database system has all the information required to determine the most effective execution plan in terms of cost parameters such as required I/O, CPU and memory usage. In similarity search, this is not the case since, for example, index selection relies on a lot of different aspects that, to some extent, can be parametrized by the client issuing a query or that may be subject to change. Therefore, the index used for executing a query is usually selected manually by the user issuing the query. This limits the amount of optimization that can be applied by the multimedia database system especially in the face of non-static data collections.

1.2.1 Motivation

Teh a

1.2.2 Research questions

Challenging the aforementioned assumptions raises very specific questions that fundamentally impact the design of a *multimedia database*. These questions are briefly summarized in Table 1.1.

Obviously, the list of questions in Table 1.1 is not exhaustive and many more implications can be derived from the assumptions challenged so far. However, the questions are the one tackled by the research presented in this thesis.

1.3 Contribution

Table 1.1 List of research questions resulting from challenging assumptions one, two and three.

RQ	Question	Related to
1	Which commonly used, secondary index structures for NNS (e.g., VA [WSB98], LSH [IM98], PQ [JDS11] based indexes) can cope with dynamic data collections and to what extent?	Assumption 1
2	Can we quantify and estimate how much the retrieval quality of index structures from RQ1 deterioration as changes are being made to the underlying data collections?	Assumption 1
2	How can we handle secondary index structures from RQ1 for which to expect deteriorated retrieval quality during query planning and execution?	Assumption 1
4	How can user knowledge about the the retrieval task at hand be factored into query planning without forcing the user the make explicit choices about how a query should be executed?	Assumption 1 & 3
5	88	

2

Applications and Use Cases

2.1 Use case 1: Multimedia Retrieval System

2.2 Use case 2: Analysis of Social Media Streams

**2.3 Use case 3: Magnetic Resonance
Fingerprinting (MRF)**

3

Related Work

PART II

Foundations

4

On Multimedia Analysis and Retrieval

4.1 Multimedia Data and Multimedia Collections

Formalisation of what multimedia data is and how it looks like.

4.2 Similarity and the Vector Space Model

Introduction of the vector space model and how it is related to similarity.

4.3 Approximate Neares Neighbor Search

Introduction into approximate nearest neighbor search.

4.4 Online Multimedia Analysis Pipeline

Introducing an online analysis pipeline (e.g., Pythia / Delphi).

4.5 Multimedia Retrieval Pipeline

Introducing a retrieval pipeline (e.g., from Cineast).

4.6 Beyond Similarity

Techniques that go beyond simple similarity search (e.g. SOM),

5

On The Design of a Database Management System

Digression into design considerations of a database management system (storage, locking, query planning, execution model etc.)

PART III

Dynamic Multimedia Data Management

6

Modelling a Database for Dynamic Multimedia Data

6.1 Architecture

6.2 Extended Model for Similarity Search

6.3 Management of Secondary Indexes

6.4 A Cost Model for Retrieval Accuracy

6.5 Further considerations

7

Cottontail DB

Implementation chapter for Cottontail DB

PART IV

Discussion

8

Evaluation

9

Conclusion & Future Work

Appendix

Bibliography

- [BVB⁺07] Henk M Blanken, Arjen P de Vries, Henk Ernst Blok, and Ling Feng. *Multimedia Retrieval*. Springer, 2007.
- [Gia18] Ivan Giangreco. *Database Support For Large-Scale Multimedia Retrieval*. PhD thesis, University of Basel, Switzerland, August 2018.
- [GS16] Ivan Giangreco and Heiko Schuldt. ADAMpro: Database Support for Big Multimedia Retrieval. *Datenbank-Spektrum*, 16(1):17–26, 2016. doi: 10.1007/s13222-015-0209-y.
- [IM98] Piotr Indyk and Rajeev Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 604–613, 1998.
- [JWZ⁺16] Björn Þór Jónsson, Marcel Worring, Jan Zahálka, Stevan Rudinac, and Laurent Amsaleg. Ten Research Questions for Scalable Multimedia Analytics. In *International Conference on Multimedia Modeling*, pages 290–302. Springer, 2016.
- [JDS11] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. Product Quantization for Nearest Neighbor Search. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(1):117–128, January 2011. doi: 10.1109/TPAMI.2010.57.
- [Ros18] Luca Rossetto. *Multi-Modal Video Retrieval*. PhD thesis, University of Basel, Switzerland, 2018.
- [WSB98] Roger Weber, Hans-Jörg Schek, and Stephen Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In *VLDB*, volume 98, pages 194–205, New York City, NY, USA. Morgan Kaufmann, 1998.

Curriculum Vitae

Name	Ralph Marc Philipp Gasser Brunnenweg 10, 4632 Trimbach
Date of Birth	28.03.1987
Birthplace	Riehen BS, Switzerland
Citizenship	Switzerland

Education

since Jan. 2000	Ph. D. in Computer Science under the supervision of Prof. Dr. Heiko Schuldt, Databases and Information Systems research group, University of Basel, Switzerland
Sept. 1997 – Aug. 1999	M.Sc. in Computer Science, University of Basel, Switzerland
Sept. 1994 – Aug. 1997	B.Sc. in Computer Science, University of Basel, Switzerland

Employment

since Jan. 2000	Research and teaching assistant, Databases and Information Systems research group, University of Basel, Switzerland
-----------------	---

Publications

1937

– turing:1937.

Hand-in in thesis and separately

Declaration on Scientific Integrity

includes Declaration on Plagiarism and Fraud

Author

Ralph Marc Philipp Gasser

Matriculation Number

2007-050-131

Title of Work

A Database System for Multimedia Analytics and Analysis

PhD Subject

Computer Sciences

Declaration

I hereby declare that this doctoral dissertation "*A Database System for Multimedia Analytics and Analysis*" has been completed only with the assistance mentioned herein and that it has not been submitted for award to any other university nor to any other faculty at the University of Basel.

Basel, DD.MM.YYYY

Signature

Hand-in separately