UNIVERSITÄT KLAGENFURT

ASSOC.PROF. DI DR. KLAUS SCHÖFFMANN

ALPEN-ADRIA-UNIVERSITÄT KLAGENFURT
Institut für Informationstechnologie
Universitätsstraße 65-67
9020 Klagenfurt
Austria
T +43463 2700 - 3620
E Klaus.Schoeffmann@aau.at
W www.KlausSchoeffmann.com

To whom it may concern.

Klagenfurt, 22.08.2019

Review of the Doctoral Thesis 'Methods of Multi-Modal Data Exploration', by Tomáš Grošup

The ubiquitous use of image and video capturing devices yields to a sheer amount of multimedia data in our current world. Unfortunately, this data is typically recorded without any meta-data that would describe the content sufficiently enough to allow for precise content-based search. This necessitates automatic content analysis and methods as well as ways that enable the interactive exploration of multimedia data. The latter is especially important for such unstructured data, since exact queries are often not appropriate or even impossible for particular (and often fuzzy/unclear) search tasks of users. Users often want to look around and inspect the data and use browsing and navigation (and other interaction means) as alternative to common textual queries.

Tomáš Grošup has dedicated his doctoral thesis to this challenging field and investigated different methods for multi-modal data exploration. In his cumulative doctoral thesis he starts with a good motivation for the topic and introduces different aspects and methods for similarity search. He also explains different types of queries as well as the characteristics of the metric space approach, before he introduces the central topic of his thesis: multimedia exploration. He discusses effectiveness and efficiency as well as different indexing schemes and already presents results from early research studies he performed in order to evaluate multimedia exploration systems. He continues with an overview of multi-modal search and content-based retrieval, discusses shortcomings of meta-data annotations, and gives a detailed summary about latent visual attributes. Over the entire introduction, the author already presents links to the different research works he has contributed and thereby shows to which part of the overall topic they relate. This gives the whole thesis a good structure and the reader a very overview of the content.

After the general introduction, Tomáš Grošup presents a selection of research papers he has published in the course of his doctoral thesis. In total, he has contributed to 15 different research papers, from which he selected eight to be included into his doctoral thesis. According to the overview in the beginning of his thesis (Page 4; where he makes very clear to what extent and which parts he has contributed) these are the eight papers where he either made the main contribution or at least significant contributions. Still, this is an exceptionally high number of research papers for a doctoral thesis. The papers contain fundamental contributions to several different aspects of multi-modal data exploration, and I will present a short summary of them in the remaining part of this review.

The first paper (Chapter 2) is about 'Methods of Multi-Modal Data Exploration' and has been published at the ACM International Conference on Multimedia Retrieval (ICMR) in 2019. It presents discussions and early experimental results about multi-modal multimedia exploration, applied to the domain of e-commerce product search and recommendation by using visual pattern search. The work summarizes how visual attributes are extracted, indexed, and integrated into the product search of an e-shop with a relational database.

The second paper (Chapter 3) is entitled 'Image Exploration Using Online Feature Extraction and Reranking' and has been presented as a demo at the ACM International Conference on Multimedia (ICMR) in 2012. In this work, an image search system is demonstrated, which uses a particle physics model for arrangement of the content. The system is based on the query-by-example approach and uses images from a web search system (e.g., Google Images or Bing Images) to start the search.

The third paper (Chapter 4) addresses 'Continuous Hierarchical Exploration of Multimedia Collections' and has been published at the IEEE International Workshop on Content-Based Multimedia Indexing (CBMI) in 2015. This work proposes a combination of iterative querying and iterative browsing for more efficient exploration, an approach that is entitled as 'hierarchical querying'. This technique starts with a zero page but limits the search space for subsequent queries. Additionally, this work also proposes a method for preserving the user context over several queries for more continuous exploration.

The fourth paper (Chapter 5) is about 'MLES: Multilayer Exploration Structure for Multimedia Exploration' and has been published at the 19th East-European Conference on Advances in Databases and Information Systems (ADBIS) in 2015. It addresses the problem where a user may not have a clear search intent but rather would like to browse and look around in the collection. For this scenario, a multilayer multimedia exploration structure is presented, which is a further refinement of the work presented in Chapter 4 and natively supports zoom and pan operations.

The fifth paper (Chapter 6) presents 'A Web Portal for Effective Multi-Modal Exploration' and has been published at the 21st International Conference on MultiMedia Modeling (MMM) in 2015. This web portal allows searching in a multimedia collection with a game-like manner, where at each iteration objects are rearranged, which are similar to the previous objects of interest. It is based on a novel distance weighting and result mixing scheme and provides an effective way of content browsing and exploration.

The sixth paper (Chapter 7) targets 'Product Exploration Based on Latent Visual Attributes' and has been published at the 26th ACM International Conference on Information and Knowledge Management (CIKM) in 2017. It is a demo of a web-based exploration system to be used in an e-shop for fashion search and uses multiple examples as a query and CNN weights (from different layers of a convolutional neural network architecture, such as AlexNet) as feature descriptor. The novelty of the system, however, is that the features are used to detect latent visual attributes that enable search for products with similar surface texture. The evaluation of the system with 3613 products and 33 visual attributes shows that the best results are achieved with central segments from the CNN (i.e., FC6, Conv4, and FC7 in AlexNet) and that the optimal query length is about 2-3 objects.

The seventh paper (Chapter 8) is entitled 'Towards Augmented Database Schemes by Discovery of Latent Visual Attributes' and has been published at the 22nd International Conference on Extending Database Technology (EDBT) in 2019. It is a visionary paper that presents subsequent research to the idea already used in the paper above: using CNN weights from different layers as feature descriptors for automatic detection of latent visual attributes (for the fashion domain in this work). The work presents a roadmap for the full vision of descriptor extraction, similarity search

(for image patches), additional processing steps (e.g., noise removal), and integration with relational database schemes (for database augmentation).

Finally, the eighth and last paper (Chapter 9) is about 'Augmenting Database Schemes by Latent Visual Attributes' and has been submitted to the Journal of Information Systems. It proposes a new model for extracting latent visual attributes for a large e-commerce system (with fashion products) and presents a system architecture for multi-modal search that can take advantage of such visual attributes/patterns of image patches for relational database queries (with SQL for relational entities). As a significant extension to the work summarized above, it also discusses the extraction of frequent patterns and the utilization of collected feedback. The system is evaluated with real users and nearly 20k objects and about 200 categories as well as about 700 tags. They found most of the attributes relevant (218 accepted attributes vs. 163 rejected ones) and considered the overall approach as very promising for database augmentation/enrichment.

Summary:
In his doctoral thesis, Tomáš Grošup has investigated several different aspects of multi-modal data exploration, which is a highly relevant research problem. The investigated aspects range from content analysis with local features as well as deep convolutional neural networks, over interface design for interactive content exploration, until database integration and practical usage with real systems. He has successfully published several conference papers at solid international venues and prepared a journal paper that is currently under review in a prestigious international journal. Overall, he has contributed to 15 research papers that address important problems in the multimedia community. From the many works he contributed to, he has selected eight to be included in his doctoral thesis. In these selected contributions, he used state-of-the-art methods for content analysis, data modelling, interface design and system evaluations. He has shown that he is able to do both design and implement novel demo applications (which are considered as important contributions in the multimedia domain as well [1]) and perform fundamental scientific studies with sound and extensive evaluations. For example, the investigation of database augmentation with latent visual attributes in Chapter 9 has been performed with a very large dataset and many different sub-tasks in a practical environment, so that different aspects of the proposed method are investigated, and the practical usefulness is demonstrated.

This shows that Mr. Grošup has profound knowledge of his research field, including the state-of-the-art in the literature, open issues and unsolved problems, appropriate research methods that can be used to perform evaluations. He is able to systematically investigate a series of scientific research questions in a challenging field of computer science.

The presented work makes several novel contributions that are of clear interest to the research community and provide the basis for additional investigations. It also gives clear evidence that the author is able to compactly but precisely describe detailed technical knowledge in great detail and perform substantial scientific tasks on his own, in a methodically correct way.

Therefore, I suggest the **best score**, without restriction.

Assoc.Prof. DI Dr. Klaus Schöffmann

[1] Worring, M., Sajda, P., Santini, S., Shamma, D. A., Smeaton, A. F., & Yang, Q. (2012). Where is the user in multimedia retrieval? IEEE MultiMedia, 19(4), 6-10.