Effective analysis, searching and browsing throughout arbitrary multimedia collections is still a challenging task. To perform a search among multimedia objects, first, a similarity model has to be defined. Such a model establishes methods describing how the content of individual objects is processed and how key features and descriptors, that are used for modeling similarity between objects, are formed. This task is not trivial since there can be many ways of determining how to comprehend the content of multimedia data. Furthermore, with the growing size of contemporary database collections, multimedia retrieval and exploration are extremely computationally intensive. Hence, researchers investigate support indexing structures that can evaluate similarity queries and can respond to user's queries in almost real-time even on datasets counting billions of objects. Another very important aspect of a retrieval system is the user interface for defining queries as well as presenting retrieved results. A multimedia system should offer various inputs for formulating user's queries, especially for situations in which a user cannot provide an ideal query example. Finally, a well-arranged and easy to read interface for visualization of retrieved results is essential for the success of a multimedia exploration and retrieval framework.

In this thesis, we showcase many aspects of content-based retrieval and multimedia exploration in specific scenarios in multiple domains (e.g., images, video, network traffic data). On top of that, we investigate state-of-the-art retrieval prototypes and applications and discuss their advantages and limitations identified by automatic and user experimental evaluations.

To deal with scalability issues, we profoundly study similarity joins for evaluating queries in metric spaces implemented in a distributed MapReduce environment adopting Hadoop and Spark platforms. We propose several variants of similarity joins offering a wide range of algorithms with different speed/precision (accuracy) trade-offs. Specifically, we study exact, approximate, and $epsilon$-approximate joins based on different approaches to data processing parallelization. Moreover, we have published java source codes of presented similarity joins for the Spark platform on the GitHub.com server.