

Similarity Search in Multimedia Databases

Daniel A. Keim Benjamin Bustos

Department of Computer and Information Science, University of Konstanz
{keim,bustos}@informatik.uni-konstanz.de

The research on multimedia databases involves different areas in Computer Science, such as computer graphics, databases, and information retrieval. There are many practical applications that benefit from this research, e.g., molecular biology, medicine, CAD/CAM, and geography. An important characteristic of these applications is the variety of data that should be supported, e.g., text, images (both still and moving), and audio. This implies that the development of a multimedia information system is considerably more complex than a traditional information system. An important research issue in the field of multimedia databases is the content-based retrieval of similar objects. Given a multimedia query object, the search for an exact match in a database is not meaningful in most applications, because the probability that two multimedia objects are identical is negligible (unless they are digital copies from the same source). For this reason, the development of efficient and effective similarity search techniques has become an important topic in the multimedia database research community.

The goal of this advanced technology seminar is to provide an overview of the similarity search problem and to present the state-of-art techniques for performing efficient and effective similarity queries in multimedia databases. The seminar begins with an introduction and a motivation of multimedia databases. The two main approaches for describing multimedia objects (as elements in a metric space or in a vector space) are introduced, as well as a description of the "Multimedia Content Description Interface" (MPEG)-7 standard. The efficiency issue is addressed for both metric and vector space approaches, describing the data structures and algorithms used to answer similarity queries. For the effectiveness issue, the seminar introduces some widely used retrieval performance measures. Several examples of techniques for particular multimedia applications (text, image, CAD, 3D objects, audio and video) are presented. The seminar outline is as follows:

1. *Introduction*: Motivation; Multimedia data; Multimedia Content Descriptor Interface (MPEG-7); Content-based retrieval in multimedia databases; Similarity queries; Modeling multimedia data: Metric and vector spaces.

2. *Efficiency*: Efficiency considerations; Metric indices; Spatial access methods; Generic multimedia object indexing; Approximate and probabilistic approaches.
3. *Effectiveness*: Retrieval evaluation in multimedia databases; Effectiveness measures; User-oriented effectiveness measures; Reference collections.
4. *Applications*: Text; Images; CAD; 3D object databases; Audio; Video.
5. *Summary and conclusions*.

Speakers

Daniel A. Keim is working in the area of multimedia databases, image similarity search and high-dimensional indexing structures. He has published extensively on multimedia databases and data mining, and has given tutorials on related issues at several conferences including SIGMOD, VLDB, ICDE and KDD. He has been program co-chair of the KDD conference in 2002 and of the IEEE Information Visualization Symposia in 1999 and 2000, and is editor of IEEE Trans. on Knowledge and Data Engineering, IEEE Trans. on Visualization and Computer Graphics, and Palgrave's Information Visualization Journal. He received his diploma (equivalent to an MSc degree) in Computer Science from the University of Dortmund in 1990, and his PhD in Computer Science from the University of Munich in 1994. He has been assistant professor in the CS department of the University of Munich, associate professor in the CS department of the Martin-Luther-University Halle, and is currently full professor and head of the database and visualization group in the CS department of the University of Konstanz, Germany.

Benjamin Bustos is working in the area of multimedia databases and similarity search in high-dimensional spaces. He received a bachelor degree in Computer Science in 1999, a professional title in Computer Engineering in 2001 and a MSc degree in Computer Science in 2002, all from the University of Chile. Currently, he is PhD student at the Department of Computer and Information Science of the University of Konstanz, and his advisor is Prof. Daniel A. Keim.