1. MUFIN: GENERAL APPROACH

Practically any information can currently be in digital form. Searching in future Internet will be complicated because of: (1) the diversity of data types and ways in which data can be sorted, compared, or classified, and (2) the quickly increasing amount of digital data. Accordingly, a successful search engine should address problems of extensibility and scalability. We present and demonstrate capabilities of MUFIN (Multi-Feature Indexing Network). From a general point of view, the search problem has three dimensions: (1) data and query types, (2) index structures and search algorithms, and (3) infrastructure to run the system on. MUFIN adopts the metric space as a very general model the similarity [3]. Its indexing and searching mechanisms are based on the concept of structured Peer-to-Peer (P2P) networks which makes the approach highly scalable and independent of the specific hardware infrastructure. This approach is schematically depicted in the following figure.

2. MUFIN: SIMILARITY IMAGE SEARCH

We demonstrate an “instance of MUFIN” designed for content-based search on large databases of general digital images. The dataset consists of 100 million images taken from CoPhIR Database\(^1\). Each image is represented by five global MPEG-7 descriptors [1] aggregated into a single metric space and the system retrieves \(k\) images which are the most similar to a given query image (according to the aggregated metric). The data is indexed by a P2P-based data structure M-Chord [2] with 2000 logical peers. Peers organize their data locally in an M-Tree. The system physically runs on six IBM servers (two quad-core CPUs, 16G RAM, six disks with RAID 5). The search is demonstrated via a Web-based interface that is available online at http://mufin.fi.muni.cz/imgsearch/. The following figure shows an example of a MUFIN query result.

3. CURRENT CBIR SYSTEMS

Several systems for Content-based Image Retrieval (CBIR) are currently available:

- **ALIPR** searches a set of images according to automatically generated annotations [http://www.alipr.com],
- **ImBrowse** allows to search about 750,000 images by color, texture, shapes (and combinations) employing five independent engines [http://media-vibrance.itn.liu.se],
- **idée** searches a commercial database of 2.8 million images according to image signatures [http://labs.ideeinc.com],
- **GazoPa** is a private service by Hitachi searching 50 million images by color and shape [http://www.gazopa.com].

All current CMIR systems, except for the very recent project GazoPa, search databases two orders of magnitude smaller than MUFIN. Moreover, approaches based on signatures usually work well only for near-duplicates. The mentioned systems are designed only for searching digital images by a specific method, which is in contrast with highly versatile MUFIN approach.

4. REFERENCES


Technical Requirements

The image search engine runs continually on our hardware infrastructure and we demonstrate it via a standard Web-based interface at http://mufin.fi.muni.cz/imgsearch/ using our own notebook. For the demonstration, we require:

- power access, ideally with a European power adapter,
- Internet connection, preferably cable (because of the bandwidth and latency),
- a projector and a projecting screen or a large display, if available,
- a board for a poster with the description of our approach (not necessary).