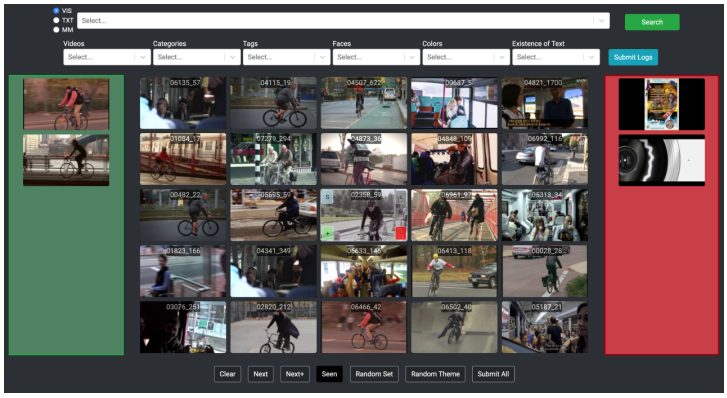
Hash-Based Indexing in Interactive Learning

Martin Aumüller (maau@itu.dk)
Riko Jakob (rikj@itu.dk)
Björn Þór Jónsson (bjth@itu.dk)

Project Outline



In interactive learning systems, such as Exquisitor, the system presents potentially relevant images to users who label them as either relevant or irrelevant. Using the labeled examples, the system then refines an interactive classifier which models the user's need. This feedback loop is repeated until the user is satisfied [1].

Exquisitor uses a cluster-based index, which allows it to return results from a collection of 100 million images in 0.3 seconds on a regular desktop computer [2]. The cluster-based index has been shown to outperform a standard hash-based indexing approach, likely because Exquisitor seeks the farthest neighbours to a plane rather than traditional nearest-neighbours to a query point. This farthest-neighbour problem has not received much attention in the hashing-based literature, so an open question is whether a hashing-based indexing approach exists which is competitive with the cluster-based approach.

The goal of this project is thus to study the application of hash-based indexing to interactive learning. The project is suitable for 1-2 well-qualified MSc students in CS. The intention is to publish the results in an international research conference.

Skills Developed

The project provides excellent training for large-scale software development, including:

- C++ / backend programming and data structure development.
- Performance evaluation and analytic skills.

References

[1] Omar Shahbaz Khan, Björn Þór Jónsson, Stevan Rudinac, Jan Zahálka, Hanna Ragnarsdóttir, Þórhildur Þorleiksdóttir, Gylfi Þór Guðmundsson, Laurent Amsaleg, Marcel Worring. Interactive Learning for Multimedia at Large. Proceedings of the European Conference on Information Retrieval (ECIR). Online, April 2020.

[2] Gylfi Þór Guðmundsson, Björn Þór Jónsson, Laurent Amsaleg. A Large-Scale Performance Study of Cluster-Based High-Dimensional Indexing. Proceedings of the Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval, Firenze, Italy, October 2010.