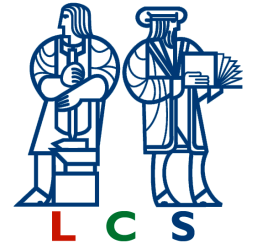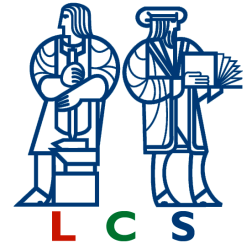# Project overview

- Geometric pattern matching is a fundamental problem, occurring in

  - Computer Vision

  - Computational Drug Design

  - Computational Biology

- Need efficient algorithms for computing similarity between a pattern and

  - a target object (one-to-one matching)

  - a database of objects (one-to-many matching)

Progress Through December 2002

Implemented a prototype high-performance system for similarity search using color histograms:

• Implemented a novel version of the Locality-Sensitive Hashing (LSH) Algorithm working in Manhattan (L1) norm

• Integrated with the algorithm for embedding Earth-Mover Distance (EMD) into Manhattan norm

• Evaluated error vs. speed tradeoff on a data set consisting of 20,000 Corel-Draw images

  • LSH algorithm runs 10-20 times faster than linear scan, and most often returna the actual (not approximate) nearest neighbor in the L1 norm

  • This results in 10-15% error with respect to the EMD metric
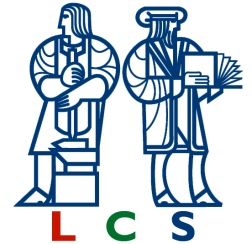
Research Plan for the Next Six Months

• Simplify the user interface, to make the system accessible to non-specialists

• Perform user experiments evaluating the quality of returned answers