

Creative Computing II

Christophe Rhodes
c.rhodes@gold.ac.uk

Autumn 2009, Tuesdays, 10:00–15:00
Winter 2010, Tuesdays, 13:30–17:30

Multimedia Information Retrieval

MIR Systems

The overall architecture:

- ▶ a collection of items;
- ▶ some set of queries that will be supported;
- ▶ some user interface for specifying queries and retrieving results;

Multimedia Information Retrieval

MIR Systems: Precision and Recall

Two kinds of error in an MIR system:

- ▶ false **positive**;
- ▶ false **negative**.

	relevant	irrelevant
retrieved	TP	FP
rejected	FN	TN

Multimedia Information Retrieval

MIR Systems: Precision and Recall

Two kinds of error in an MIR system:

- ▶ false **positive**;
- ▶ false **negative**.

	relevant	irrelevant
retrieved	TP	FP
rejected	FN	TN

Measures of performance:

- ▶ **precision:**
 - ▶ what proportion of retrieved results are relevant?
 - ▶ $\frac{TP}{TP+FP}$

Multimedia Information Retrieval

MIR Systems: Precision and Recall

Two kinds of error in an MIR system:

- ▶ false **positive**;
- ▶ false **negative**.

	relevant	irrelevant
retrieved	TP	FP
rejected	FN	TN

Measures of performance:

- ▶ **precision:**
 - ▶ what proportion of retrieved results are relevant?
 - ▶ $\frac{TP}{TP+FP}$
- ▶ **recall:**
 - ▶ what proportion of relevant items did I find?
 - ▶ $\frac{TP}{TP+FN}$

Multimedia Information Retrieval

MIR Systems: Small Collections

Given an existing database, a query item, and some kind of query specification:

- ▶ compute some feature of the query item;
- ▶ for each database item
 - ▶ compute the same feature for the item;
 - ▶ compare the query and database item features with an appropriate distance measure;
 - ▶ if the distance is sufficiently small, retrieve the item, otherwise reject it;
- ▶ return the collection of retrieved items.

Multimedia Information Retrieval

MIR Systems: Small Collections

Given an existing database, a query item, and some kind of query specification:

- ▶ compute some feature of the query item;
- ▶ for each database item
 - ▶ compute the same feature for the item;
 - ▶ compare the query and database item features with an appropriate distance measure;
 - ▶ if the distance is sufficiently small, retrieve the item, otherwise reject it;
- ▶ return the collection of retrieved items.

Refinements:

- ▶ limit the number of retrieved items (by removing larger-distance items from the retrieved set);
- ▶ cluster the retrieved set in some way for visualisation;
- ▶ infinite possibilities...

Multimedia Information Retrieval

MIR Systems: Large Collections

The algorithm in the previous slide is $O(N)$ in the size of the database: the feature for each database item is computed and compared.

- ▶ precompute database features (easy)

Multimedia Information Retrieval

MIR Systems: Large Collections

The algorithm in the previous slide is $O(N)$ in the size of the database: the feature for each database item is computed and compared.

- ▶ precompute database features (easy)
- ▶ precompare features with query (**index**, usually hard)

Multimedia Information Retrieval

MIR Systems: Large Collections

The algorithm in the previous slide is $O(N)$ in the size of the database: the feature for each database item is computed and compared.

- ▶ precompute database features (easy)
- ▶ precompare features with query (**index**, usually hard)

For a scalar feature:

- ▶ precomparison is in fact easy: binary tree (for example);
- ▶ finding the nearest neighbour is $O(\log N)$.

Multimedia Information Retrieval

MIR Systems: Large Collections

The algorithm in the previous slide is $O(N)$ in the size of the database: the feature for each database item is computed and compared.

- ▶ precompute database features (easy)
- ▶ precompare features with query (**index**, usually hard)

For a scalar feature:

- ▶ precomparison is in fact easy: binary tree (for example);
- ▶ finding the nearest neighbour is $O(\log N)$.

For a vector feature:

- ▶ precomparison is difficult in general;
- ▶ spatial trees, locality-sensitive hashing, probabilistic algorithms;
- ▶ (beyond the scope of this course: current research!)