A Classification and Survey of Multi-Dimensional Query Types

David Broneske
Student Conference on Software Engineering and Database Systems
July 6, 2012
Motivation

• Importance of multimedia databases increases
• Introduction of multi- or high-dimensional index structures
Motivation

• Importance of multimedia databases increases
• Introduction of multi- or high-dimensional index structures
  • Evaluation of index structures and comparisons between them
  • **But:** Consideration limited to one or two query types when comparing index structures
Motivation

- Importance of multimedia databases increases
- Introduction of multi- or high-dimensional index structures
  - Evaluation of index structures and comparisons between them
  - **But:** Consideration limited to one or two query types when comparing index structures
- QuEval Framework
- Implementing many query types for performance comparisons
Query Types – Chosen Query Types

- Exact match query
- Range query
- Partial match query
- Similarity range query
- Nearest neighbor query
- Similarity join query
- K-closest pairs query
- All-nearest-neighbor query
Query Types – Chosen Query Types

- Exact match query
- Range query
- Partial match query
- Similarity range query
- Nearest neighbor query
- Similarity join query
- K-closest pairs query
- All-nearest-neighbor query
Query Types – Example Database

• Query types in a feature-vector space
• E.g., Image database
• Features: average color, texture

![Feature-vector space diagram with letters A to Z representing different types of texture and average color.

Text:]

average color

red
green
blue

irregular

regular

slightly irregular

highly regular

A B D C E F G H I J K M G N O P Q R S T U V W X Y Z

0 100 0 50 75 50 25 75

average color

texture

highly regular

regular

slightly irregular

irregular

0 50 100
Query Types – Similarity Join Query

- Similarity range query for a set of query images
- Result: Pairs of images
Query Types – K-Closest Pairs Query

- Combination of $k$-nearest neighbor query and similarity join
- Retrieves $k$ pairs with the smallest distance
Query Types – All-Nearest-Neighbor Query

- Retrieves the nearest neighbor for every image in the query set
- Multiple usage of an image in the queried set possible
Classification

- Query types differ in:
  - selection of results
  - amount of retrieved points
  - complexity

- Criteria influence the usage of queries in specific use case
Classification – Selection of Results

- Different selection criteria for different query types
- Retrieval of resulting points according to:
  - their values in selected dimensions → **Boolean queries**
  - their similarity/distance to the query → **Similarity queries**
Classification – Selection of Results

- Different selection criteria for different query types
- Retrieval of resulting points according to:
  - their values in selected dimensions → **Boolean queries**
  - their similarity/distance to the query → **Similarity queries**

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Similarity Query</th>
<th>Boolean Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact Match Query</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Range Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Partial Match Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Similarity Range Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Nearest Neighbor Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Similarity Join Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>K-Closest Pairs Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>All-Nearest-Neighbor Query</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Classification – Amount of Retrieved Points

- Query types with fixed and unfixed size of result set
Classification – Amount of Retrieved Points

- Query types with **fixed** and **unfixed** size of result set

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Fixed Size of Result Set</th>
<th>Unfixed Size of Result Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact Match Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Range Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Partial Match Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Similarity Range Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Nearest Neighbor Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Similarity Join Query</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>K-Closest Pairs Query</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>All-Nearest-Neighbor Query</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Classification – Complexity

- Complexity depends on dimensionality ($n$), amount of queried points ($m$), amount of query points ($l$)
Classification – Complexity

- Complexity depends on dimensionality \((n)\), amount of queried points \((m)\), amount of query points \((l)\)

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact Match Query</td>
<td>(O(n \times m))</td>
</tr>
<tr>
<td>Range Query</td>
<td>(O(n \times m))</td>
</tr>
<tr>
<td>Partial Match Query</td>
<td>(O(n \times m))</td>
</tr>
<tr>
<td>Similarity Range Query</td>
<td>(O(n \times m))</td>
</tr>
<tr>
<td>Nearest Neighbor Query</td>
<td>(O(n \times m))</td>
</tr>
<tr>
<td>Similarity Join Query</td>
<td>(O(n \times m \times l))</td>
</tr>
<tr>
<td>K-Closest Pairs Query</td>
<td>(O(n \times m \times l))</td>
</tr>
<tr>
<td>All-Nearest-Neighbor Query</td>
<td>(O(n \times m \times l))</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

• Different query types reviewed
• Majority of queries are similarity queries  
  → Multimedia retrieval
• Fixed or unfixed size of result set  
  → Impact on Pastprocessing
• Different complexities  
  → Factors influencing query performance
• Next step: implementing more query types to QuEval
Thank you for your attention!


Saake, Gunter; Sattler, Kai-Uwe; Heuer, Andreas: *Datenbanken - Implementierungskonzepte*. 3. Auflage. MITP, Bonn, 2011

Zhang, Rui; Ooi, Beng C.; Tan, Kian-Lee: Making the Pyramid Technique Robust to Query Types and Workloads. In: *Proc. Int’l. Conf. on Data Engineering (ICDE)*, IEEE Computer Society, 2004, S. 313–324
Background – Distance Functions

- Requirement of similarity functions
- Replaced by distance functions
- Minkowski distance functions

\[ L_p(x, y) = \sqrt[p]{\sum_{i=1}^{n} (||x_i - y_i||^p)} \]

- Manhattan: \( p = 1 \)
- Euclidean: \( p = 2 \)
- Supremum: \( p = \infty \)

adapted from [SSH11]
Query Types – Exact Match Query

- Retrieves images with same feature vector as query
  \[ \rightarrow \text{Retrieves image } x \text{ for query } q \text{ if } \forall i \in [1, n] : q_i = x_i \]
Query Types – Range Query

- Define an lower ($l$) and upper ($u$) bound
- Retrieve image $x$ if $\forall i \in [1, n] : l_i \leq x_i \leq u_i$
Query Types – Partial Match Query

- Special kind of range query
- Either \( l_i = u_i \) or \( l_i = 0 \land u_i = \infty \)
Query Types – Similarity Range Query

- Define a distance threshold $\epsilon$
- Retrieve images $x$ if $d(x, y) \leq \epsilon$
Query Types – Nearest Neighbor Query

- Result: Nearest point to the query point
- Image $x$ is nearest point to $q$ in database $DB$ if $\forall z \in DB, z \neq x : d(q, x) \leq d(q, z)$
- Special case: $K$-nearest neighbor query