Using Regions of Interest for Adaptive Image Retrieval

Michael Springmann, Heiko Schuldt

University of Basel, Database and Information Systems Group
Bernoullistr 16, CH-4056, Basel, Switzerland
{michael.springmann, heiko.schuldt}@unibas.ch

The work has been supported by the Swiss National Science Foundation (SNF) in the context of the project Query by Sketching under contract No. 200021-117800 / 1
Content-Based Image Retrieval

Query by Example

![Query Image]

Feature Vector:

\[ \sum_{i=0}^{d} (a_i - b_i)^2 \]

Image Collection

Feature Vector:

![Feature Vector Chart]

unibasel
### Simplified Comparison with Text IR

<table>
<thead>
<tr>
<th></th>
<th>CBIR</th>
<th>Text IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the query?</td>
<td>Feature(s) of Image(s)</td>
<td>Set of keywords</td>
</tr>
<tr>
<td>What is desired result?</td>
<td>Ranked list of most-similar Images</td>
<td>(Ranked) list of Documents containing keywords</td>
</tr>
<tr>
<td>What contributes to distance/similarity?</td>
<td>Every difference between feature vectors (query + compared image)</td>
<td>Only keywords found in documents</td>
</tr>
<tr>
<td>How is query refined (e.g. using relevance feedback)</td>
<td>Adjust weights of features / images, add new features/images</td>
<td>Add new keywords, adjust weights</td>
</tr>
</tbody>
</table>
Adaptability in Query by Example

What contributes to distance?

Query Image

Feature Vector

\[ \sum_{i=0}^{d} (a_i - b_i)^2 \]

Feature Vector

\[
\begin{align*}
\text{Feature Vector} & : 203, 236, 172, 210, 78 \\
\text{Feature Vector} & : 180, 140, 242, 247, 78
\end{align*}
\]
Adaptability in Query by Example

What contributes to distance?

Query Image

Feature Vector

\[ \sum_{i=0}^{d} (a_i - b_i)^2 \]

Feature Vector
Overview of the Approach

- Selecting Regions of Interest (ROI)
- Matching Regions in Images
- Similarity Between Corresponding Regions

Conclusion & Outlook
Selecting Regions of Interest (ROI)

- Traditional Approaches: Rectangular Bounding Box

  Use just the bbox as new query image

  … but this matches entire images, not just interesting regions.
Selecting Regions of Interest (ROI)

• Traditional Approaches: Automatic Segmentation

Blobworld:
Selecting Regions of Interest (ROI)

- Intrinsic Bounds to Automatic Image Segmentation
Selecting Regions of Interest (ROI)

Using Interactive Paper or Tablet PC

Used for query definition and relevance feedback
Identify High-Level Concepts

• Face Detection

• ImageCLEF
  – Visual Concept Detection Task
  – Automatic Medical Annotation

• ALIPR
• User selects Region of Interest (ROI)

• Input devices more appropriate than mouse & keyboard

• System tries to find high-level concepts within ROI

But how to match selected region?
Overview of the Approach

Selecting Regions of Interest (ROI)

Matching Regions in Images

Similarity Between Corresponding Regions

Conclusion & Outlook
Matching Regions in Images: Keypoints

Generic Solution: SIFT (Scale-Invariant Feature Transform)
Matching Regions in Images: Keypoints

Generic Solution: SIFT (Scale-Invariant Feature Transform)
Matching Regions in Images: Keypoints

Generic Solution: SIFT (Scale-Invariant Feature Transform)
Keypoint Matching

1481 Keypoints
For each keypoint in query find the two most similar keypoint in reference
determine ratio: best_distance / second_distance
if better than threshold, keep as match
End

2536 Keypoints
Many!
128 dim dist.
- Slow!

Hard to define!
Clustering and Filtering

Input:
List of 50 best SIFT matches (no threshold)

Filter Outliers with RANSAC
Keypoint Matching

1481 Query keypoint: 924 ms
   Early Termination: 629 ms

684 Query keypoint: 438 ms
   Early Termination: 311 ms

900 Query keypoint: 576 ms
   Early Termination: 401 ms

536 Query keypoint: 350 ms
   Early Termination: 236 ms
Further Optimizations possible

• Did not use any index so far
• Can be performed in parallel using multiple threads / computing cluster / Grid
• Add constraints:
  – Allow only certain rotations and scaling
  – Can be evaluated based on keypoint metadata
  – Example: Allow only 10° of rotation → 152 ms for cat region

• Remember:
  Number of images to perform matching on can be filtered based on high-level metadata
  (face, visual concept, tags,...)
Feedback

- Can be traced back to originating keypoint
- Can be used to refine transformation constraints
- Can be used to refine high-level concepts
Matching Regions - Summary

• Use invariant keypoint descriptors like SIFT originating from object recognition domain

• Allow more matches, but add clustering & filtering

• Use Early Termination and constraints to reduce time

But how to determine similarity?
Overview of the Approach

- Selecting Regions of Interest (ROI)
- Matching Regions in Images
- Similarity Between Corresponding Regions

Conclusion & Outlook
Determine and Apply Affine Transformation
Similarity Between Corresponding Regions

Image Distortion Model (IDM)

\[ d_{IDM}(A, B) = \sqrt{\sum_{x=1}^{\text{width}} \sum_{y=1}^{\text{height}} (A_{i,j} - B_{i,j})^2} \]
Conclusion

Selecting Regions of Interest (ROI)

- using Interactive Paper or Tablet PCs to
  - Give maximum flexibility to the user
  - Better understand what is the need (concepts)

Matching Regions in Images

- using keypoints within query regions
  - use clustering and filtering
  - determine transformation

Similarity Between Corresponding Regions

- Based on Image Distortion Model
  - Ignore pixels outside query regions
Examples – London Eye
Examples – Tower Bridge
Examples – Big Ben
Examples – Phone Booth
Examples – Underground
Examples – Underground (2)
Outlook

• Work on Keypoints
  – SIFT is ideal for all images
  – Use additional color information in color images
  – Edge-based keypoint descriptors required for sketches

• Combinations of regions
  – Semantics and implementation of AND, OR, AVERAGE

• Evaluate approach

• Distributed computation to reduce execution time
Thanks!

Thank you for your attention.

Special thanks to Manuel Studer.