SAFIRE: Towards Standardized Semantic Rich Image Annotation

Christian Hentschel, Andreas Nürnberg
Ingo Schmitt, Sebastian Stober
Faculty of Computer Science
Otto-von-Guericke University Magdeburg
Contents

- Motivation
- MPEG 7
- SAFIRE
- Query Processing
- Conclusions
Image Retrieval

the vision:
- user can provide a natural language query, in which the desired content is described

This requires semantics!

approaches to bridge the “Semantic Gap”
focus on:
- extraction of more relevant and descriptive features from images
- segmentation of images into descriptive regions
- ways to allow (manual) semantic annotations
- methods to appropriately analyze and process user queries

In the following: Two simplified approaches…
Image Segmentation

- Idea: Extract Objects
- Example: Blobworld (UC Berkeley)
  - Automatically segment images into regions which roughly correspond to objects or parts of objects
  - Extract features from object:
    - Shape, color, and texture features
    - Location of object in image
  - Use object features for searching (CBIR – content based image retrieval)
Image Segmentation

Similar approach: STRICT (LIP6, Paris)

Open Problems

- Assign labels to segments
- Manual annotation not possible for all images
- Research issues:
  - Methods to automatically label images based on collections of (partially) annotated images
- Problems:
  - Only a few annotated collections available, that can be used to develop and evaluate learners
  - Most use different ways to store information about segments and annotation
  - Missing standardization!
MPEG-7: An Appropriate Standard?

- ISO/IEC standard developed by MPEG (Moving Picture Experts Group)
- Formally named “Multimedia Content Description Interface”
- Standard for describing multimedia content data
- Uses XML (defined by XML Schema)
- Supports some degree of interpretation of the information meaning
- Not aimed at any application in particular
- Elements standardized in MPEG-7 support a broad range of applications
MPEG-7: Description Schemes

Overview of the MPEG-7 Multimedia Description Schemes (DSs)

Source: MPEG-7 Overview (version 10) http://www.chiariglione.org/mpeg
MPEG-7: Content Description

Description of the content structural aspects

- Temporal segment (Video Segment, Audio Segments)
  - Segment composed of one connected component
  - Segment composed of three connected components

- Spatial segment (Still Region)
  - Segment composed of one connected component
  - Segment composed of three connected components
MPEG-7: Content Description

Examples of image description with Still Regions
MPEG-7: Summary

- Standard is very complex and covers several aspects that are required for user and context adaptive IR systems.
- Most likely the best foundation right now for the development of algorithms and systems for indexing ALL digital media dealing with image sound, or both.
- It is possible to use only specific DS, e.g. for storing information about image segments and their semantic meaning!
SAFIRE (Semantic Annotation Framework for Image RETrieval)

- Framework to edit, segment and annotate images based on MPEG-7 standard
SAFIRE (Semantic Annotation Framework for Image REtrieval)

- Supports
  - flat and hierarchical segmentation
  - user specific annotation (collaborative use possible!)
Image annotation “My brother at the bank of the Thames.”

Terms “brother” and “bank” are linked to SynSets in the WordNet and EuroWordNet InterLingual Index (ILI)

Thus, it is possible to avoid ambiguities! (*Not part of standard*)

```xml
<Sentence xml:lang="en" id="annot_1">
  <Phrase id="annot_1.1">
    My
    <Phrase semantics="WordNet:SynSetID=2471824
    EuroWordNet:ILISynSetID=8542395">
      brother
    </Phrase>
  </Phrase>
  at
  <Phrase id="annot_1.2">
    the
    <Phrase semantics="WordNet:SynSetID=86786241
    EuroWordNet:ILISynSetID=1332468">
      bank
    </Phrase>
    of the Thames
  </Phrase>
</Sentence>
```
SAFIRE (MPEG-7 StillRegion Description Scheme)

- Linking semantic annotations to image segments

```xml
<image>
    <SemanticRef idref="annot_1" /> <!--sentence annotating whole image-->
    ...
    <SpatialDecomposition overlap="true" gap="true"
        criteria="flat decomposition of the image"
        id="flatDecomposition">
        <StillRegion id="region_1">
            <SemanticRef idref="annot_1.1" /> <!--annotation: my brother-->
            <SpatialLocator>...</SpatialLocator>
        </StillRegion>
        <StillRegion id="region_2">
            <SemanticRef idref="annot_1.2" /> <!--annotation: the bank of the Thames-->
            <SpatialLocator>...</SpatialLocator>
        </StillRegion>
    </SpatialDecomposition>
</image>
```
Remarks on the use of Ontologies

- Segment “Ayers Rock” belongs to the general concept “mountain”
- No need to store this general concept if domain ontology for mountains is available (and term is linked to it!)
- Search for mountain could return all images (or videos) that contain images of mountains
- Automatically, also an access structure can be provided, e.g.

<table>
<thead>
<tr>
<th>Mountain (total hits: 176)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayers Rock, Uluru (10)</td>
</tr>
<tr>
<td>Alps (145)</td>
</tr>
<tr>
<td>Zugspitze (32)</td>
</tr>
<tr>
<td>Materhorn (49)</td>
</tr>
<tr>
<td>Rocky Mountains (32)</td>
</tr>
<tr>
<td>Sawback Range (14)</td>
</tr>
</tbody>
</table>
Query Processing

- Challenge: combining various search paradigms
  - Textual descriptions → text retrieval
  - Interconnected image collection → navigation
  - low-level Features → content-based retrieval
  - Spatial / temporal relations → database query
  - User-defined data → database query

- Main problems:
  - Reconciliation of evaluation results: discrete Boolean values ↔ similarity scores from \([0,1]\)
  - Logic on imprecise similarity scores
Query Processing: Fuzzy Logic

Solution approach: **Fuzzy logic**

- Similarity scores and Boolean values as membership values of a fuzzy set
- T-norm / T-conorm for conjunction / disjunction
- Languages: e.g. same, WS-QBE, SDC, SA

Problems with Fuzzy-logic

- Min/max are **dominating**: returns one input value regardless of the other one
- Non-dominating t-norm / t-conorm: **no idempotence**
- Weighting search terms (Fagin) violates **distributivity**
Query Processing: Comparability

missing comparability of score functions

Query processing
fuzzy logic evaluation

evaluation of a database condition
evaluation of retrieval condition 1
evaluation of retrieval condition 2

score

perceived similarity
Conclusions

- Standardized Annotation can help to
  - Create benchmark collections
  - Develop learning methods to enable (semi-)automatic labeling
  - Develop and evaluate refined query languages

- SAFIRE
  - An initial prototype that provides a platform to integrate methods for
    - Segmentation
    - Annotation
    - Querying