Semantic and Interactive Content-based Image Retrieval

Björn Barz
Computer Vision Group, Friedrich Schiller University Jena, Germany

**AUTOMATIC QUERY IMAGE DISAMBIGUATION (AID)**

Query images are typically ambiguous. Retrieval results based solely on feature comparisons will hence mix multiple related concepts. We enable the users to refine the results with just a single click towards their intended concept.

1. Retrieve top m nearest neighbors. Usually miss results from multiple senses (here: babies, hands).
2. Means clustering of top results to discover senses. Determined automatically based on largest expunging.
3. User selects relevant cluster.
4. Adjust effective distances of all database items.
5. Fetch refined results.

**HIERARCHY-BASED SEMANTIC EMBEDDINGS AND THE COSINE LOSS**

Goal: learn image features whose cosine similarity resembles the semantic similarity of their classes.

Prior: semantic knowledge given as taxonomy of classes.

Train a CNN to map images onto the (semantic) space defined by the class embeddings using the Cosine Loss:

\[ L(x, y) = 1 - \frac{(\phi(x), \psi(y))}{||\phi(x)|| \cdot ||\psi(y)||} \]

Classification

- Search for nearest class in joint embedding space.
- Improves from scratch accuracy on small datasets (20110 images/class) compared with cosine + entropy (example below with sub sampled CNN)
- Can even work fine instead of semantic embeddings.

**INFORMATION-THEORETIC ACTIVE LEARNING (ITAL)**

Use Active Learning to increase the efficiency of relevance feedback: Instead of asking for feedback for the top-scoring results, choose images for which the feedback is expected to be most useful for improving the relevance model.

Maximization of Mutual Information Between Feedback and Relevance Model

\[ \sum \frac{(R_f, r_f) \cdot \log \frac{P(r_f | P(r_f), u)}{P(r_f | u)}} \]

References