Multimedia Systems and Applications

Real Time Google and Live Image Search Re-ranking
James Wang

Based on ACM Multimedia 2008 paper and poster by:

Problem, Idea and Existing Approaches

Problems:
- Search engines use only text information to find images.
- Many of returned results are noisy, disorganized, or irrelevant.

Idea
- Using visual information to re-rank and improve text based image search results

Existing Approaches:
- Assume that there is one dominant cluster of images inside each image set returned by a keyword query, and treat images inside this cluster as “good” ones.
- Require online training, so cannot be used for realtime online image search.
- Cannot handle ambiguity inside a keyword query.

Contribution
- A framework is proposed to build a system to re-rank text based image search results in an interactive manner.
- After query by keyword, user can click on one image, indicating this is the query image.
- Then all the returned images are re-ranked according to their similarities with the query.
- A fast and effective online image search re-ranking algorithm based on one query image only without online training.
- Search by adaptive similarity based on the user intention.

Method
- The query image is firstly categorized into one of several predefined categories.
- Inside each category, a specific weight schema, which combines the features adaptive to this kind of images, is used to measure the user’s intention when using this image to query.
- This weighting schema is based on minimizing the rank loss for all query images on a training set through the proposed method modified from RankBoost [8].

Search by Adaptive Similarity
- Given F different features of an image, the normalized similarity between image i and j on feature m is denoted as $s^m_i(j)$, which takes value in the range of [0; 1].
- A vector $\alpha_i$ is defined for each image i to express its specific “point of view” towards different features. The larger $\alpha_{im}$ is, the more important the mth feature will be for image i.
- Assume $\alpha \geq 0$ and $||\alpha||_1 = 1$, the adaptive similarity measurement at image i is a linear combination of its similarities on different features weighted by $\alpha_i = \sum_m \alpha_{im} s^m_i$. Adjust the weight $\alpha$ adaptively for each query image i.

Intention Categories
- Images containing close-ups of general objects;
- Object with Simple Background;
- Scenery images;
- Images containing portrait of a single person;
- Images with general people inside, and are not “Portrait”.

Attributes for intention categorization

- **Face existence:** Whether the image contains faces.
- **Face number:** Number of faces occurred in the image.
- **Face size:** The percentage of the image frame taken up by the face region.
- **Face position:** Coordinate of the face center relative to the center of the image.
- **Directionality:** Kurtosis of Edge Orientation Histogram (EOH, Section 3). The larger the Kurtosis is, the stronger the image shows directionality.
- **Color Spatial Homogeneity:** Variance of values in different blocks of Color Spatiolet (CSpa, Section 3), describing whether color in the image is distributed spatially homogeneously.
- **Edge Energy:** Total energy of edge map obtained from Canny Operator on the image.
- **Edge Spatial Distribution:** First divide the image into 3 by 3 regular blocks, then calculate the variance of Edge Energy in the 9 blocks. Describe whether edge energy is mainly distributed at the image center.

Intention Specific Feature Fusion

**Algorithm 1: Learning Feature Weight Inside Intention Category**

1. **Input:** initial weight $D^0$ for all possible query images $s$ in the current intention category $Q$, similarity matrix on each feature $s^T_i$ (1 for all i and inc).
2. **Initialize:** Set $D^t = D$, for any $i$. Set step $t = 1$.
3. **While:** Algorithm not converged do for each $s \in Q$ do
4. **Select best feature and corresponding similarity** $s^T_i$ for current re-ranking problem under weight $D^t$.
5. **Adjust weight** $D^{t+1}$ $(j,k) \propto D^t(j,k) \exp [\alpha(t) (s_i(j) - s_i(k))]$.
6. **Normalize** $D^{t+1}$ with factor $Z_t$ so that $D^{t+1}$ will be a distribution.
7. **End for**
8. **End while**
9. **Output:** Final optimal similarity measure for current intention category: $s^* = \sum_{i \in Q} \alpha_i s_i$.

Feature Design

- **Attention Guided Color Signature:** A color signature that accounts for varying importances of different parts of an image. An attention detector [10] is used to compute a saliency map for the image, then perform k-Means clustering weighted by this map.

- **Color Spatiolet:** An image is first divided into n £ n patches by a regular grid. Within each patch, we calculate its main color as the largest cluster after k-Means clustering. The image is finally characterized by Color Spatiolet (CSpa), a vector of n² color values.

Feature Design (cont.)

- **Multi-Layer Rotation Invariant EOH:** Edge Orientation Histogram (EOH) [7], which describes histogram of edge orientations, has long been used in vision applications.
- **Rotation invariance is used when comparing two EOHs, resulting in a Multi-Layer Rotation Invariant EOH (MRI-EOH).**
- **To calculate the distance between two MRI-EOHs, one of them is rotated to best match the other, and take this distance as the distance between the two.**

Feature Design (cont.)

- **Histogram of Gradient (HoG).**
- **Gist:**
  - Gist is proposed in [15] to characterize the holistic appearance of an image, and is proven to work well for scenery images.

- **Daubechies Wavelet:**
  - The 2nd order moments of wavelet coefficients in various frequency bands (DWave) are used to characterize texture properties in the image [16].

- **SIFT:**
  - A 128-dimensional SIFT [11] is used to describe regions around Harris interest points. A codebook of 450 words is obtained by hierarchical k-Means on a set of 1.5 million SIFT descriptors extracted from the training set. Descriptors are then quantized by this codebook.
Experimental Data

- Contains many concepts including object, scenery, people name, place name, etc., and covers a large range of keywords (http://mmlab.ie.cuhk.edu.hk/intentsearch).
- 26,908 images from 30 keywords labeled, among which 46% are labeled as good, consisting of 128 sub classes and 70 main classes.

Experimental Method

- Evaluation Criteria.
  - P20 and P40 (Proportion of images within the same sub class in top 20 and 40 returned ones) are used to evaluate performance.
- Evaluation baselines.
  - Select the feature which have the largest variance of similarity scores, based on the assumption that an effective feature should give good image much larger score than a mediocre one.
  - Use global weights to combine features.

Training and Testing Results

- Divide the labeled benchmark database into two sets.
  - One includes 9017 images from 10 keywords, and is used for SIFT codebook training (Sec. 3), intention classifier training, and feature combination weights training.
  - Another includes 17,891 images from other 20 keywords, and is used for testing.
- For each intention category, 200 images are manually labeled from the training set and a C4.5 decision tree is trained.
  - Use this decision tree to classify all “good” images (3021 images) of the training set into 5 intention categories.
  - Then optimal weight for each intention category is learnt using the algorithm described in Section 2.2.

Performance Evaluation

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Alg</th>
<th>CHP</th>
<th>Citer</th>
<th>DMB</th>
<th>SIFT</th>
<th>MRR(CCH)</th>
<th>Pat</th>
<th>Lambda</th>
<th>Global</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>elephant</td>
<td>0.913</td>
<td>0.904</td>
<td>0.859</td>
<td>0.856</td>
<td>0.846</td>
<td>0.850</td>
<td>0.850</td>
<td>0.850</td>
<td>0.850</td>
<td>0.850</td>
</tr>
<tr>
<td>bench</td>
<td>0.863</td>
<td>0.857</td>
<td>0.856</td>
<td>0.855</td>
<td>0.854</td>
<td>0.854</td>
<td>0.854</td>
<td>0.854</td>
<td>0.854</td>
<td>0.854</td>
</tr>
<tr>
<td>car</td>
<td>0.933</td>
<td>0.925</td>
<td>0.882</td>
<td>0.881</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
</tr>
<tr>
<td>puppy</td>
<td>0.923</td>
<td>0.916</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
</tr>
<tr>
<td>river</td>
<td>0.933</td>
<td>0.925</td>
<td>0.882</td>
<td>0.881</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
<td>0.882</td>
</tr>
<tr>
<td>cow</td>
<td>0.911</td>
<td>0.901</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
<td>0.860</td>
</tr>
</tbody>
</table>

(a) Precision in Top 20  (b) Precision in Top 40

Some Search Results

- Results of searching keywords “bicycle” and “Tibet” respectively:

Conclusion

- A realtime re-ranking algorithm is proposed to enhance the performance of Google Image Search and Microsoft Live Image Search, by letting user select a query image from text search results.
  - An intention categorization model to integrate a set of complementary features adaptive to the query image.
  - A large labeled database is built to share with the community.
  - A realtime online image search engine, combining text and IntentSearch, is implemented.