Multimedia Systems and Applications

Real Time Google and Live Image Search Re-ranking

James Wang

Based on ACM Multimedia 2008 paper and poster by:

Cui et al., Real Time Google and Live Image Search Re-ranking, in Proceedings of ACM Multimedia 2008, pp. 729 – 732, October 26–31, 2008

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.*

2

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.**

3

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].

[8] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. J. Mach. Learn. Res., 4:933–969, 2003.

4

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 $s_i^m(j)$, which takes value in the range of [0; 1].**
- **A vector α_i is defined for each image i to express its specific "point of view" towards different features. The larger α_{im} is, the more important the mth feature will be for image i.**
- **Assume $\alpha \succeq 0$ and $\|\alpha\|_1 = 1$, the adaptive similarity measurement at image i is a linear combination of its similarities on different features weighted by α_i: $s_i(\cdot) = \sum_{m=1}^{F} \alpha_{im} s_i^m(\cdot)$**
  - adjust the weight α adaptively for each query image i.

5

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".**

6

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, Section3). The bigger the Kurtosis is, the stronger the image shows directionality.
- **Color Spatial Homogeneousness:**
  - Variance of values in different blocks of Color Spatialet (CSpa, Section3), 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

7

Intention Specific Feature Fusion

Algorithm 1 Learning Feature Weight Inside Intention Category

1. **Input:** initial weight D_i for all possible query images i in the current intention category Q, similarity matrix on each feature $s_i^m(\cdot)$ for all i and m;
2. Initialize: Set $D_i^1 = D_i$ for any i. Set step $t = 1$;

while Algorithm not converged **do**

  for each $i \in Q$ **do**

3. Select best feature and corresponding similarity $s_i^t(\cdot)$ for current re-ranking problem under weight D_i^t;
4. Calculate real value α_i^t according to Equation 1;
5. Adjust weight $D_i^{t+1}(j,k) \propto D_i^t(j,k) \exp\left\{\alpha_i^t\left[s_i(j) - s_i(k)\right]\right\}$;
6. Normalize D_i^{t+1} with factor Z_i^t so that D_i^{t+1} will be a distribution;
7. t++;

  end for

end while

8. **Output:** Final optimal similarity measure for current intention category: $s(\cdot) = \sum_{i,t} \alpha_i^t s_i^t(\cdot)$.

8

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 Spatialet.**
  - 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 Spatialet (CSpa), a vector of n2 color values.

9

Feature Design (cont.)

- **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-dimension 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.

10

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.

11

Feature Design (cont.)

- **Histogram of Gradient (HoG).**
  - HoG [4] is the histogram of gradients within image blocks divided by a regular grid.
  - HoG reflects the distribution of edges over different parts of an image, and is especially effective for images with strong long edges.
- **Facial Feature.**
  - Face existence and their appearances give clear semantic interpretations of the image.
  - Face detection algorithm [17] is applied to each image, and to obtain the number of faces, face size and position as the facial feature (Face).
  - The distance between two images is calculated as the summation of differences of face number, average face size, and average face position.

12

Experimental Data

- **a collection of 451,352 images associated with 483 keywords crawled from Google Image Search and Microsoft Live Image Search.**
  - 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.**

13

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.

14

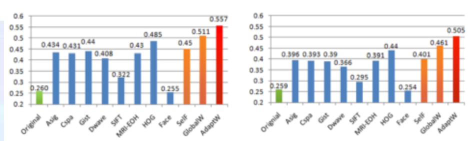
Training and Testing Results

- **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.

15

Performance Evaluation

Keywords	Precisions of each feature								Selected Feature	Global similarity	Adaptive similarity
	ASig	CSpa	Gist	DWave	SIFT	MRI-EOH	HoG	Face			
airplanes	0.333	0.320	0.327	**0.353**	0.165	0.305	0.348	0.152	**0.350**	**0.363**	**0.424**
beach	**0.633**	0.607	0.608	0.565	0.475	0.578	0.625	0.482	**0.620**	**0.687**	**0.727**
car	**0.430**	0.406	0.372	0.362	0.132	0.349	0.405	0.214	**0.378**	**0.406**	**0.492**
dolphin	**0.620**	0.583	0.491	0.573	0.319	0.446	0.467	0.272	**0.520**	**0.591**	**0.646**
guitar	0.207	0.213	0.446	0.400	0.265	0.339	**0.485**	0.103	**0.446**	**0.449**	**0.513**
paris	0.312	0.340	0.357	0.319	0.286	0.328	**0.373**	0.333	**0.3252**	**0.461**	**0.502**
rice	0.511	0.489	0.478	0.438	0.384	0.464	**0.579**	0.284	**0.499**	**0.591**	**0.656**

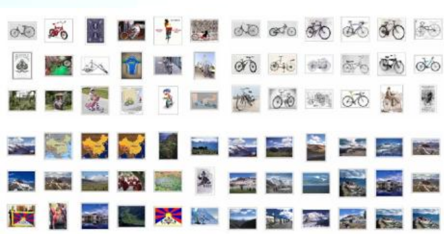


(a) Precision in Top 20 (b) Precision in Top 40

16

Some Search Results

- **Results of searching keywords "bicycle" and "Tibet" respectively:**



(a) (b)

17

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.

18