IMAGE ANNOTATION USING SEARCH AND MINING TECHNOLOGIES

Xin-Jing Wang, Lei Zhang, Feng Jing, Wei-Ying Ma
Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China
wangxinj@cn.ibm.com, {leizhang, fengjing, wyma}@microsoft.com

Background:
1.Image auto-annotation is a hot research topic in recent years
2.Traditional computer vision & machine learning approaches fail

Difficulties:
1.Unclear how to model the semantic concepts
2.Lack of training data to bridge the semantic gap

Motivations:
1.The huge deposit, the Web, brought solutions to many previously "unsolvable" problems
2.The search technology succeeds in many commercial systems

Basic Idea:
A data-driven approach leveraging the Web-scale image dataset and search technology to learn relevant annotations

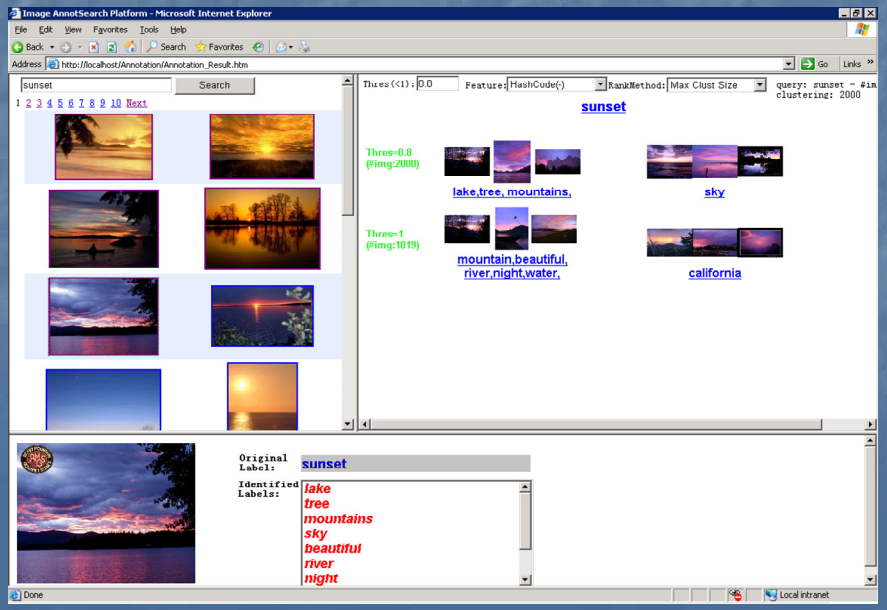


Figure 1. Interface and an example of the AnnoSearch system

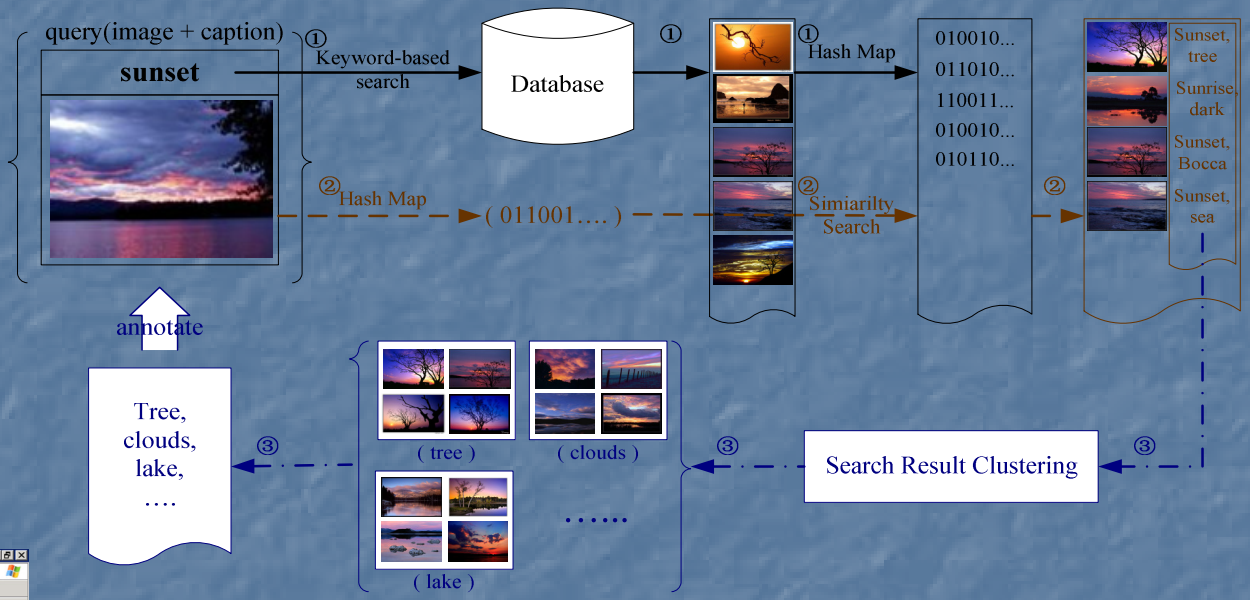


Figure 2. Framework of the AnnoSearch System

Input: a query image + an initial keyword
Output: complementary annotations

Process:

1.Text-based search: retrieve semantically similar images
2.Content-based search: retrieve visually similar images
> Hash coding algorithm to solve the efficiency problem
3.SRC clustering to mine annotations from the descriptions of the retrieved images

Performance Evaluation Results

Our Image Deposit:

2.4 million high-quality photo forum images with noisy descriptions

Testing Datasets:

1.Google image query dataset: 30 queries from categories “Apple, Beach, Beijing, Bird, Butterfly, Clouds, Clownfish, Japan, Liberty, Lighthouse, Louvre, Paris, Sunset, Tiger, Tree”

2. UW Content-based Image Retrieval dataset: categories are “Australia, Campus, Cannon beach, Cherries, Football, Geneva, Green lake, Indonesia, Iran, Italy, Japan, San juan, Spring flower, Swiss mountain, Yellowstone”. All images are used as queries.

Evaluation Criterion (Google image query set):

E = (#perfect + 0.5 x #correct - #error) / #queries

Conclusion:

1. **High effectiveness** (A much higher precision)
   >0.6 precision score on Google query set, and 0.38 on UW dataset (5 ground-truth annotations on average), while it is normally about 0.2~0.3 for previous annotation approaches
2. **High efficiency**:
   For the content-based retrieving phase, it costs 0.072s for weighted Harming distance measure. (24,000 candidate images on average, Dual Intel Pentium 4 Xeon hyper-threaded CPU, 2G memory)
3. **No supervised learning phase** and hence can handle **unlimited vocabulary**

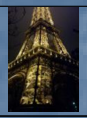
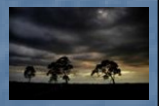
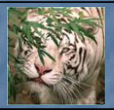
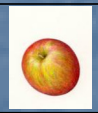
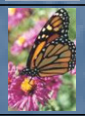
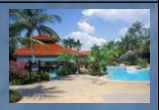
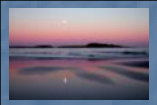
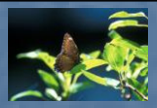
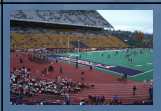
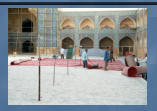
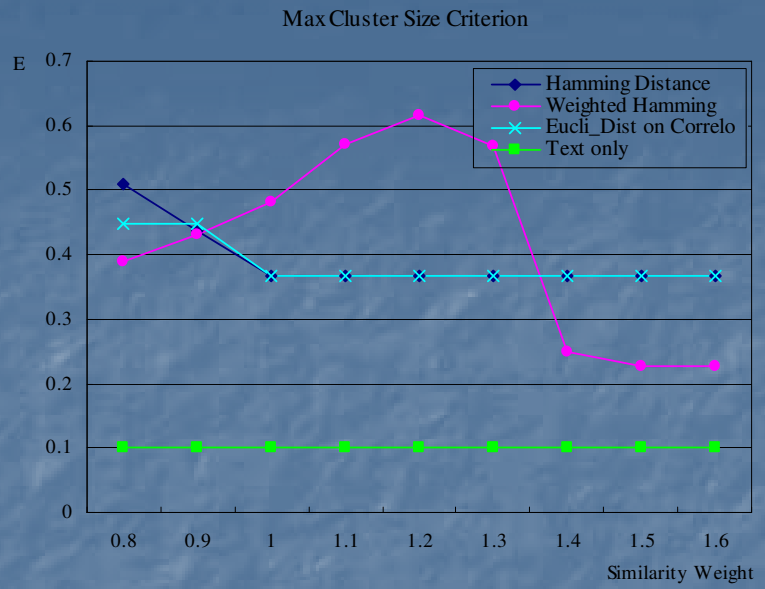
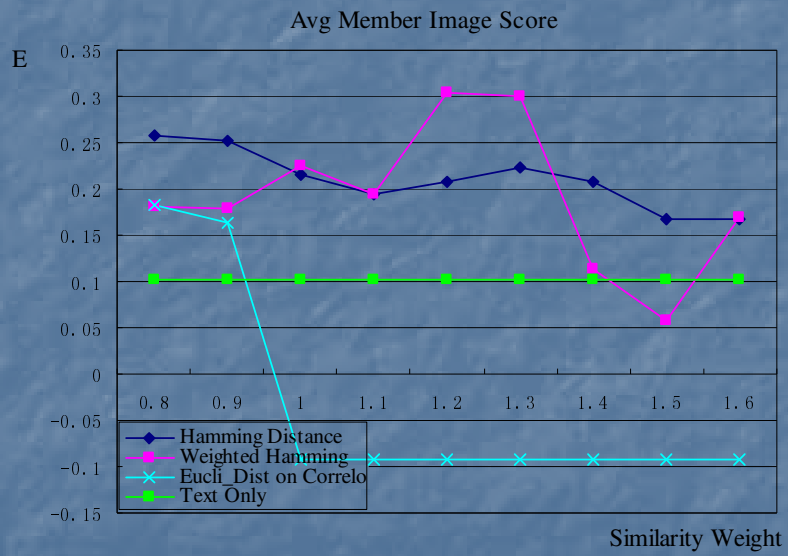
	**Paris** Las vegas, effel tower, love paris		**Paris** Sacre coeur, paris building, effel tower		**Paris** Eiffel tower, france, sky, paris nights		**Clouds** Dark clouds, sun, sky, sunrise, morn
	**Sunset** Lake, tree, mountain, sky, beautiful, water		**Tiger** Whiter tiger, usa, zoo		**Tree** House, flower, snow, sky, tree trunk		**Clouds** National park, europe, south america, blue sky
	**Apple** Studio, kitchen, fruit, color		**Apple** Fruit, apple tree		**Butterfly** Flower, butterfly house, beautiful butterfuly		**Beach** South america, beautiful beach, beach house
	**Clownfish** Anemone, reef, red sea		**Beach** Sky,island, sun beach, sunrise, beach island		**Butterfly** Yellow butterfly, swallowtail, nature		**Liberty** York, liberty statue, sun
	**Campus** college, campus life, center, tree		**Football** stadium, school football, football game, football player		**Iran** mashhad, kish island, esfahan		**Cannon Beach** haystack rock

Figure 3. Examples of annotations produced by AnnoSearch system.
The upper four rows show a few results on Google image query dataset.
The bottom row shows a few results on the UW dataset.



(a) Precision w.r.t. maximum cluster size criterion



(b) Precision w.r.t. average member image score criterion

Figure 4. Average Precision of annotation vs. image filtering threshold on the 30 Google query images