Query Specific Fusion for Image Retrieval

Shaoting Zhang[2], Ming Yang[1], Timothee Cour[1], Kai Yu[1], Dimitris N. Metaxas[2]

[1]NEC Laboratories America, Inc., Cupertino, CA, USA

[2]Dept. of Computer Science, Rutgers University, Piscataway, NJ, USA

Motivations

- Our goal: develop scalable algorithms for large-scale image retrieval.
- Current state-of-the-art methods:
  - Quantized local invariant features indexed by a large vocabulary tree.
  - Holistic features indexed by compact binary hashing codes.
- Pros and Cons:

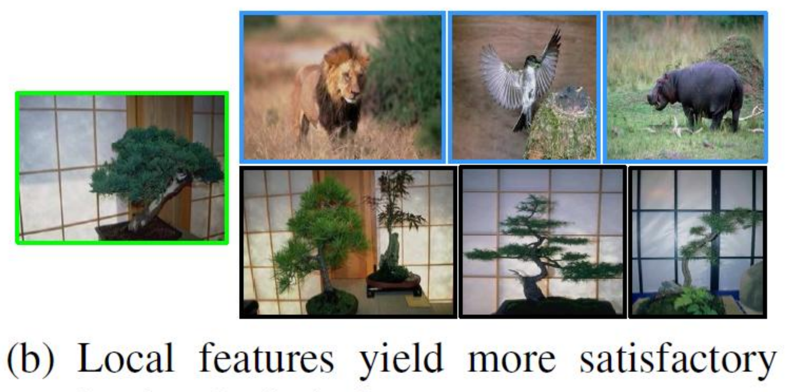
	Retrieval speed	Memory usage	Retrieval precision	applications	Image properties to attend to
Local feat.	fast	high	high	near duplicate	local patterns
Global feat.	faster	low	low	general images	global statistics

- Can we unite the strengths of two lines of approaches adaptively?

Challenges

- The features and algorithms are dramatically different.
  - Hard for the feature-level fusion
  - Hard for the rank aggregation
- The fusion shall be query specific and database dependent.
  - Hard to learn how to combine across different datasets
- No supervision and relevance feedback!
  - Hard to evaluate the retrieval quality *online*

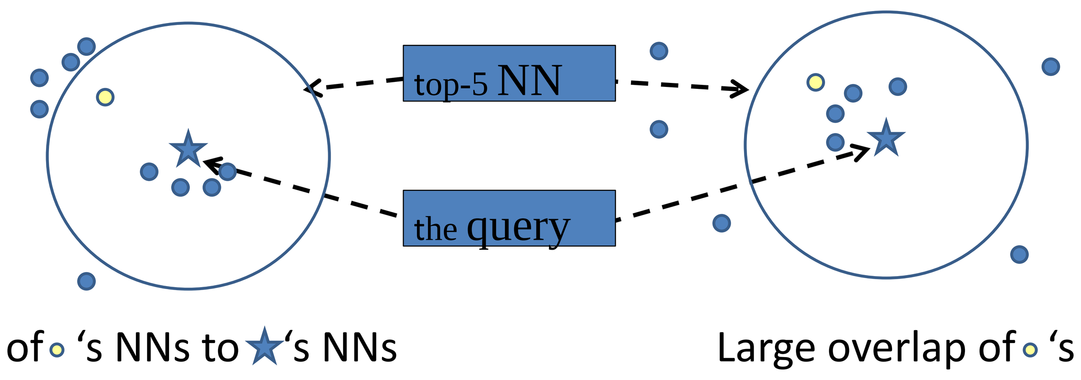
(a) Holistic features yield more satisfactory results than local features.



(b) Local features yield more satisfactory results than holistic feature.

Our method (1)

- How to evaluate *online* the quality of retrieve results from methods using local or holistic features?
- Assumption: The consensus degree among top candidate images reveals the retrieval quality.
- The consistency of top candidates' nearest neighborhoods.
- Intuition: which case, *i.e.,* the yellow dot ◦, is more preferable?

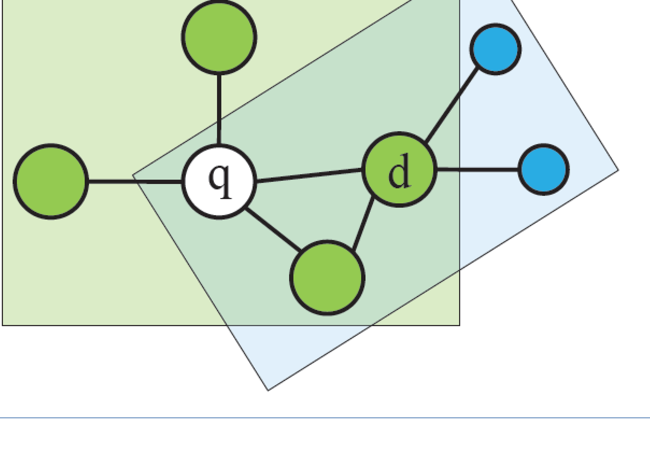


Little overlap of ◦ 's NNs to ★'s NNs   Large overlap of ◦ 's NNs to ★'s NNs

- A graph-based approach to fusing and re-ranking retrieval results of different methods.
- **Step 1:** Construct a weighted undirected graph to represent a set of retrieval results of a query image q.
  - Edge: the *reciprocal* neighbor relation
  - Edge weight: the Jaccard similarity between neighborhoods with a decaying parameter α *w.r.t.* to the number of hoops to the query q:

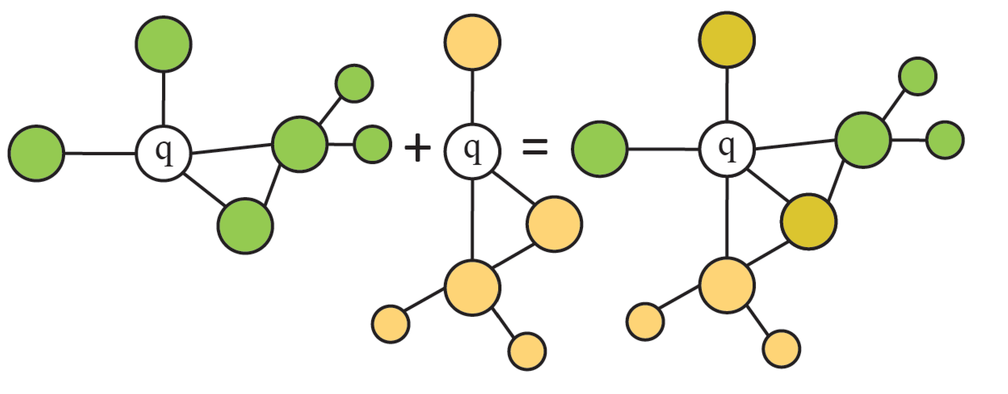
$$J(i,i') = \frac{|N_k(i) \cap N_k(i')|}{|N_k(i) \cup N_k(i')|}$$

$$w(i,i') = \alpha(q,i,i')J(i,i'),$$



Our method (2)

- **Step 2:** Fuse multiple graphs to one graph: union of the nodes/edges and sum of the weights



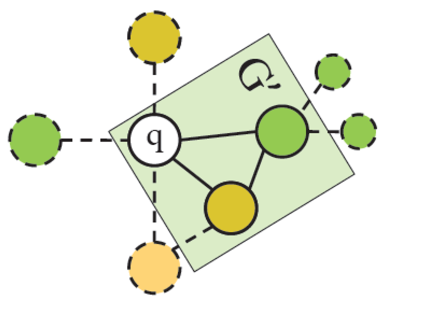
- **Step 3:** Re-rank according to this fused graph:
  - Ranking by a local Page Rank. Perform a link analysis on G, and rank the nodes by their connectivity in G by the intelligent surfer model.

$|V|\times|V|$ transition matrix $\mathbf{P}$ as $P_{ii'} = w(i,i')/\deg(i)$

$$p^{t+1} = (1-\beta)\pi + \beta \mathbf{P}^T p^t.$$

  - Ranking by maximizing the weighted density

$$G' = \underset{G'=(V',E',w)\subset G:\ q\in V'}{\operatorname{argmax}} \frac{\sum_{(i,i')\in E'} w(i,i')}{|V'|}$$



Experimental results

- **UKBench:** 2550 × 4=10200 images. N-S Score, top-4 recall rate, max=4

Jégou *et al.* [11]	Qin *et al.* [12]	*HSV*	*VOC* [28]	*HSV graph*	*VOC graph*	*Rank aggregation*	*SVM fusion*	*Graph PageRank*	*Graph density*
3.68	3.67	3.17	3.54	3.28	3.67	3.47	3.56	3.76	**3.77**

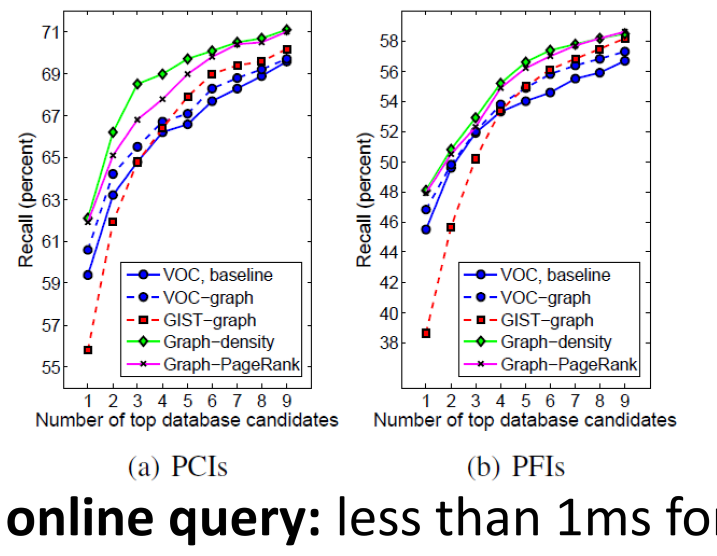
- **Corel-5K:** 50 × 100 = 5000 images. Top-1 precision

VOC	*GIST*	*VOC-graph*	*GIST-graph*	*SVM-fusion*	*Graph-PageRank*	*Graph-density*
46.66	46.16	51.50	50.72	51.34	51.76	**54.62**

- **Holidays:** 1491 images in 500 groups. mAP (%)

Jégou *et al.* [17]	Jégou *et al.* [31]	*HSV*	*VOC* [28]	*Rank aggregation*	*SVM fusion*	*Graph PageRank*	*Graph density*
81.3	83.9	62.60	77.50	78.62	79.04	84.56	**84.64**

- **SF Landmark :** 803 queries in 1.06M PCI and 638K PFI images



(a) PCIs   (b) PFIs

Query images: from smart phones
Database images:
PCI: *perspective central images*
PFI: *perspective frontal images*
Obtained from panoramic images

- **Efficient online query:** less than 1ms for the fusion

Dataset	# of images	*VOC*	*HSV/GIST*	*Graph-fusion*	t_r *(ms)*
UKbench	10200	85	1	< 1	87
Corel-5K	4999	76	< 1	< 1	78
Holidays	1490	72	< 1	< 1	73
PCI-SFLandmark	1,062,468	645	103	< 1	749
PFI-SFLandmark	638,090	467	64	< 1	532

- **Memory cost:** 340MB extra storage for the top-50 nearest neighbors for 1.7M images in the SF Landmark, a small fraction of the inverted indexes.

Take home messages

- Significantly improved the accuracy over individual retrieval methods.
- Required a minimal computational overhead and acceptable storage.
- No supervision, few parameters, easy to implement and reproduce.

Contact email: shaoting@cs.rutgers.edu and myang@nec-labs.com

Project Page: http://tinyurl.com/image-retrieval