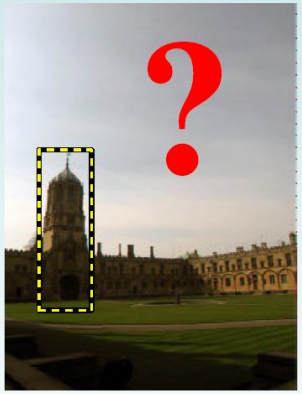
Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval

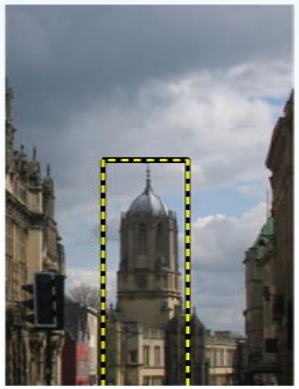
O. Chum, et al.

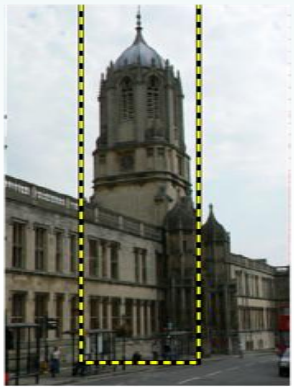
Presented by Brandon Smith
Computer Vision – Fall 2007

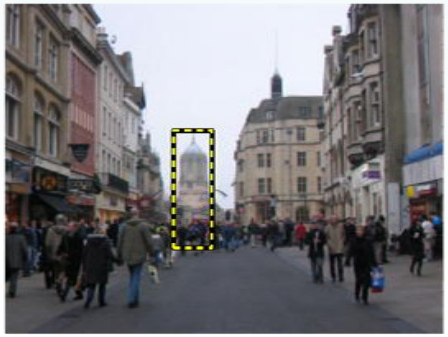
Objective

- Given a query image of an object, retrieve all instances of that object in a large (1M+) image database









General Approach

- Use bag-of-visual-words architecture
- Use query expansion to improve performance
- Improve query expansion by…
  - …(1) using spatial constraints to suppress false positives, and
  - …(2) constructing a latent feature model of the query object

Procedure
(Probabilistic Interpretation)

- Extract a *generative model* of an object from the query region
- Form the *response set* from images in the corpus

Generative Model

- A spatial configuration of visual words extracted from the query region, plus…
- A “background” distribution of words that encodes the overall frequency statistics of the corpus

Response Set

- Images from the corpus that contain a large number of visual words in common with the query region
- My subsequently be ranked using spatial information to ensure that…
  - …(1) response and query contain similar features, and
  - …(2) features occur in compatible spatial configurations

Objective

- Given a query image of an object, retrieve all instances of that object in a large (1M+) image database
- Do this by improving retrieval performance…

Objective

- Given a query image of an object, retrieve all instances of that object in a large (1M+) image database
- Do this by improving retrieval performance… by deriving better (generative) object models given the query region

A Better Model

- Keep the form fixed (still a configuration of visual words)
- Enrich the object model with additional information from the corpus
- Refer to this as a *latent model*

Latent Model

- Generalization of *query expansion*

Query Expansion

- Well-known technique in text-based information retrieval

(1) Generate a query
(2) Obtain an original response set
(3) Use a number of high ranked documents to generate a new query
(4) Repeat from (2)…

Latent Model

- Generalization of query expansion
- Well suited to this problem domain:

  (1) Spatial structure can be used to avoid false positives

  (2) The baseline image search *without* query expansion suffers more from false negatives than most text retrieval systems

Real-time Object Retrieval

- A visual vocabulary of 1M words is generated using an approximated K-means clustering method

Approximate K-means

- Comparable performance to exact K-means (< 1% performance degradation) at a fraction of the computational cost
- New data points assigned to cluster based on approximate nearest neighbor
- We go from $O(NK)$ to $O(N\ log(K))$

  N = number of features being clustered from

  K = number of clusters

Spatial Verification

- It is vital for query-expansion that we do not…

  …expand using false positives, or

  …use features which occur in the result image, but not in the object of interest
- Use hypothesize and verify procedure to estimate homography between query and target
- > 20 inliers = spatially verified result

Methods for Computing Latent Object Models

- Query expansion baseline
- Transitive closure expansion
- Average query expansion
- Recursive average query expansion
- Multiple image resolution expansion

Query expansion baseline

1. Find top 5 (unverified) results from original query
2. Average the term-frequency vectors
3. Requery once
4. Append these results

Transitive closure expansion

1. Create a priority queue based on number of inliers
2. Get top image in the queue
3. Find region corresponding to original query
4. Use this region to issue a new query
5. Add new *verified* results to queue
6. Repeat until queue is empty

Average query expansion

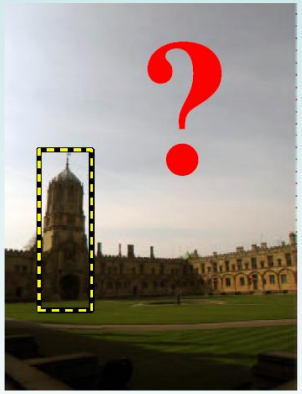
1. Obtain top (m < 50) *verified* results of original query
2. Construct new query using average of these results
3. Requery once
4. Append these results

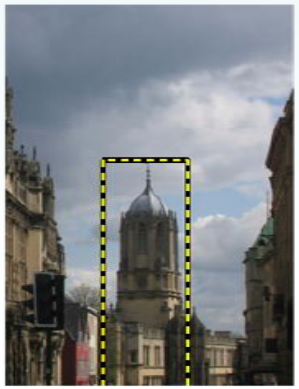
Recursive average query expansion

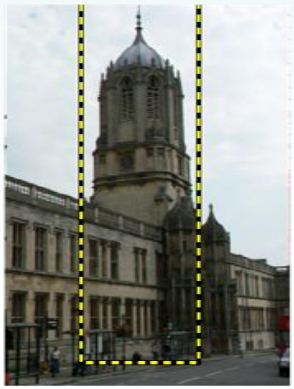
1. Improvement of average query expansion method
2. Recursively generate queries from all verified results returned so far
3. Stop once 30 verified images are found or once no new images can be positively verified

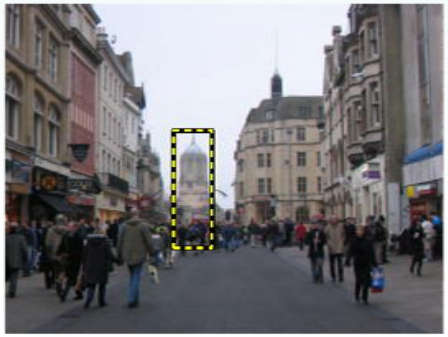
Multiple image resolution expansion

1. For each verified result of original query…
2. Calculate relative change in resolution required to project verified region onto query region









Multiple image resolution expansion

3. Place results in 3 bands:

   (0, 4/5) (2/3, 3/2) (5/4, infinity)

4. Construct average query for each band
5. Execute independent queries
6. Merge results
   1. Verified images from first query first, then…
   2. Expanded queries in order of number of inliers

Datasets

- Oxford dataset
- Flickr1 dataset
- Flickr2 dataset

Oxford Dataset

- ~5,000 high res. (1024x768) images
- Crawled from Flickr using 11 landmarks as queries
- Ground truth labels
  - *Good* – a nice, clear picture of landmark
  - *OK* – more than 25% of object visible
  - *Bad* – object not present
  - *Junk* – less than 25% of object visible

Flickr1 Dataset

- ~100k high resolution images
- Crawled from Flickr using 145 most popular tags

Flickr2 Dataset

- ~1M medium res. (500 x 333) images
- Crawled from Flickr using 450 most popular tags

Datasets

Dataset	Number of images	Number of features
Oxford	5,062	16,334,970
Flickr1	99,782	277,770,833
Flickr2	1,040,801	1,186,469,709
Total	1,145,645	1,480,575,512

Table 1. The number of descriptors for each dataset.

Evaluation Procedure

- Compute Average Precision (AP) score for each of the 5 queries for a landmark
- Average these to obtain a Mean Average Precision (MAP) for the landmark
- Use two databases
  - D1: Oxford + Flickr1 datasets (~100k)
  - D2: D1 + Flickr1 datasets (~1M)

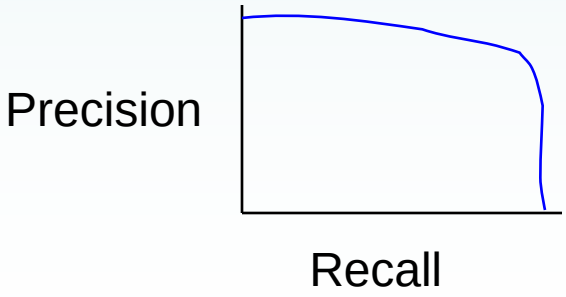
Average Precision

- Area under the precision-recall curve
- Precision = RPI / TNIR
- Recall = RPI / TNPC

RPI = retrieved positive images

TNIR = total number of images retrieved

TNPC = total number of positives in the corpus



Retrieval Performance

- D1
  - 0.1s for typical query
  - 1GB for index
- D2
  - 15s – 35s for typical query
  - 4.3GB for index (use offline version)

Retrieval Performance

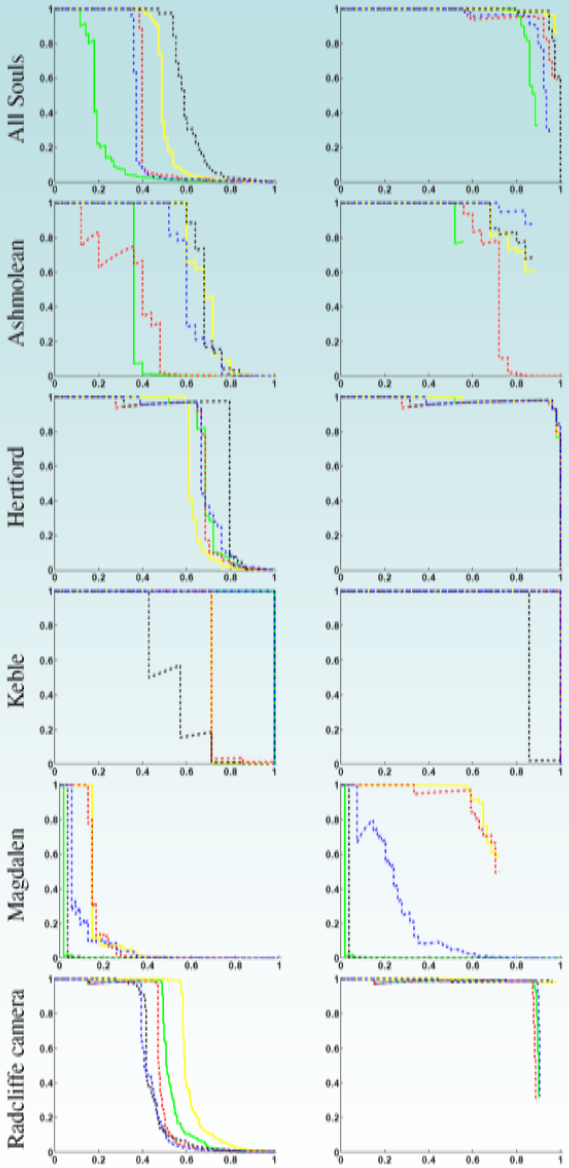


Figure 3. Precision recall curves before (left) and after (right) query expansion on experiment D1. These results are for resolution expansion, our best method. In each case the five curves correspond to the five queries for that landmark.

Retrieval Performance

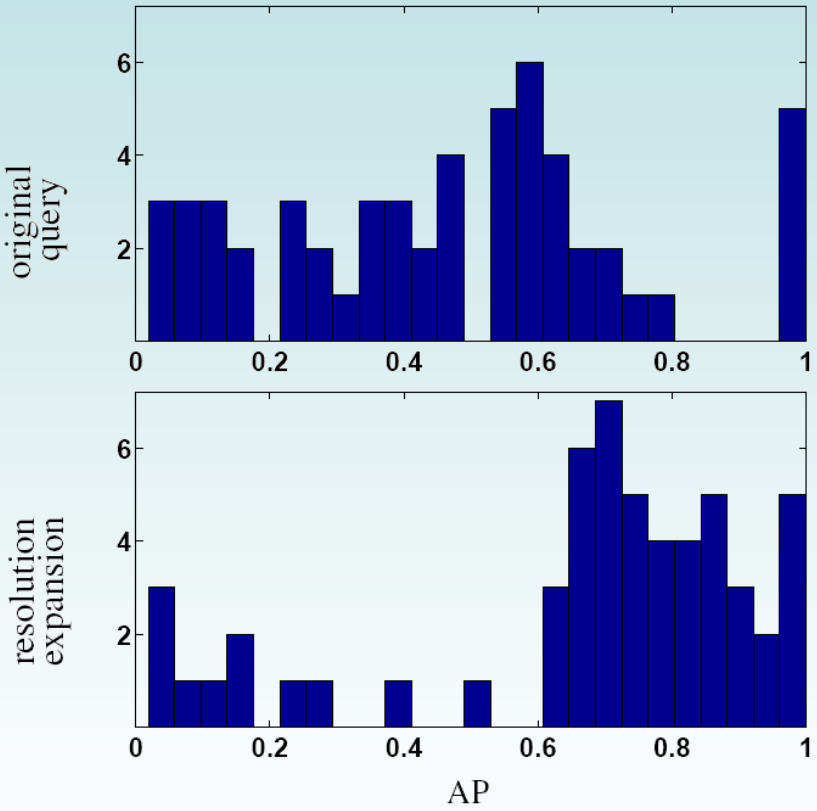


Figure 4. Histograms of the average precision for all 55 queries in experiment D2. Note, that the query expansion moves the mass of the histogram towards the right-hand side, *i.e.* towards total recall.

Retrieval Performance

Figure 5. Some false positive images for Magdalen Tower query. The tower shown is actually part of Merton College chapel.

Retrieval Performance

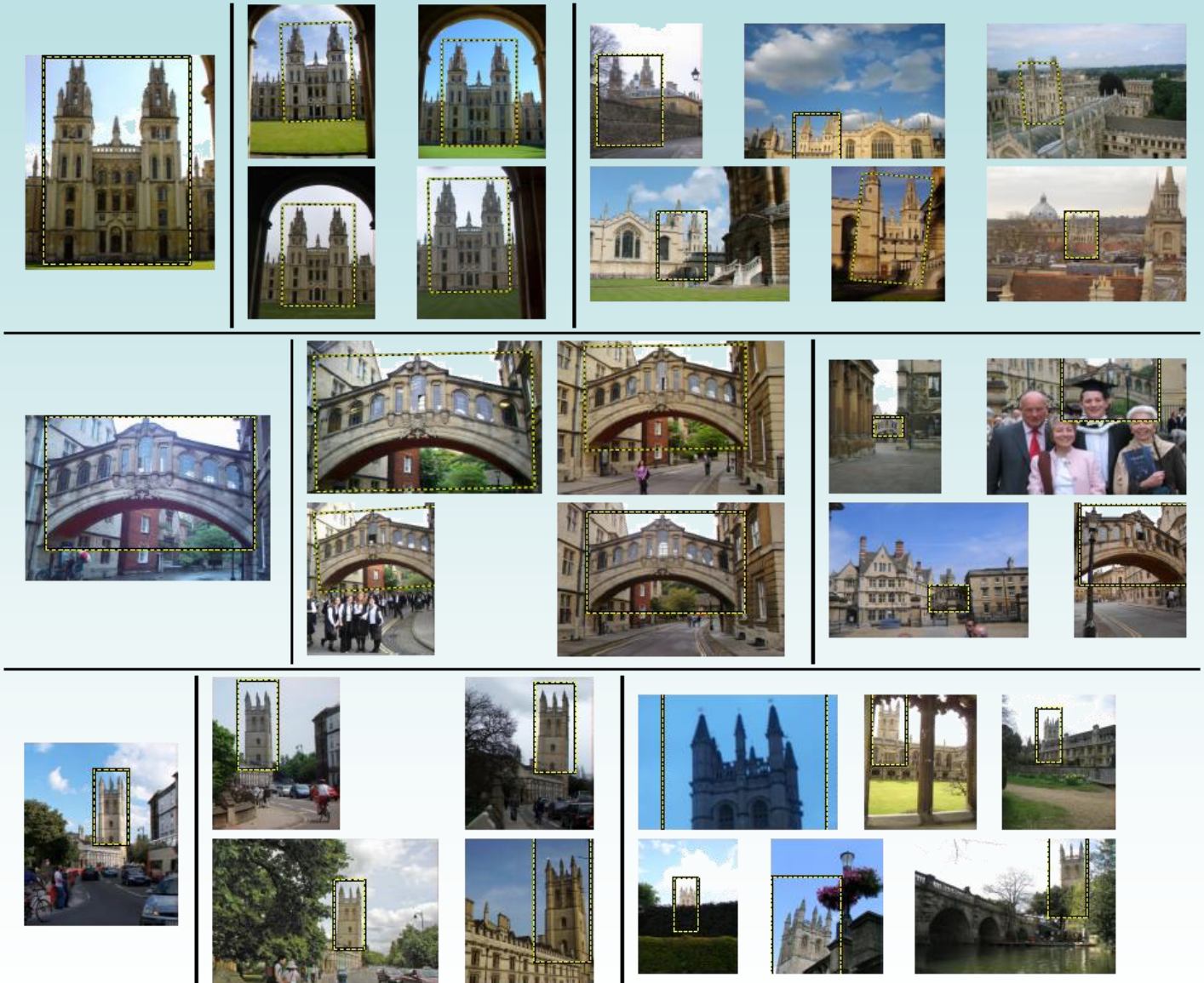


Figure 6. Demonstrating the performance of the method on a number of different queries. The image to the left shows the original query image. The four images in the middle show the first four results returned by the original query before query expansion. The images to the right show true positive images returned after query expansion which were not found from the bag-of-words method.

Method and Dataset Comparison

	Ground truth		*Oxford* + *Flickr1* dataset						*Oxford* + *Flickr1* + *Flickr2* dataset					
	OK	Junk	ori	qeb	trc	avg	rec	sca	ori	qeb	trc	avg	rec	sca
All Souls	78	111	41.9	49.7	85.0	76.1	85.9	**94.1**	32.8	36.9	80.5	66.3	73.9	**84.9**
Ashmolean	25	31	53.8	35.4	51.4	66.4	74.6	**75.7**	41.8	25.9	45.4	57.6	**68.2**	65.5
Balliol	12	18	50.4	52.4	44.2	63.9	**74.5**	71.2	40.1	39.4	39.6	55.5	**67.6**	60.0
Bodleian	24	30	42.3	47.4	49.3	**57.6**	48.6	53.3	32.3	36.9	43.5	**46.8**	43.8	44.9
Christ Church	78	133	53.7	36.3	56.2	63.1	**63.3**	63.1	52.6	18.9	55.2	**61.0**	57.4	57.7
Cornmarket	9	13	54.1	60.4	58.2	74.7	74.9	**83.1**	42.2	53.4	56.0	65.2	68.1	**74.9**
Hertford	24	31	69.8	74.4	77.4	89.9	90.3	**97.9**	64.7	70.7	75.8	87.7	87.7	**94.9**
Keble	7	11	79.3	59.6	64.1	90.2	**100**	97.2	55.0	15.6	57.3	**67.4**	65.8	65.0
Magdalen	54	103	9.5	6.9	25.2	28.3	**41.5**	33.2	5.4	0.2	16.9	15.7	**31.3**	26.1
Pitt Rivers	7	9	**100**	**100**	**100**	**100**	**100**	**100**	**100**	90.2	**100**	**100**	**100**	**100**
Radcliffe Cam.	221	348	50.5	59.7	88.0	71.3	73.4	**91.9**	44.2	56.8	86.8	70.5	72.5	**91.3**
Total	539	838	55.0	52.9	63.5	71.1	75.2	**78.2**	46.5	40.5	59.7	63.1	67.0	**69.6**

Table 2. Summary of ground truth, and the relative performance of the different expansion methods.

ori = original query
qeb = query expansion baseline
trc = transitive closure expansion
avg = average query expansion
rec = recursive average query expansion
sca = multiple image resolution expansion

Questions