Large Scale Image Retrieval

Ondřej Chum and Jiří Matas

Center for Machine Perception

Czech Technical University in Prague

Features

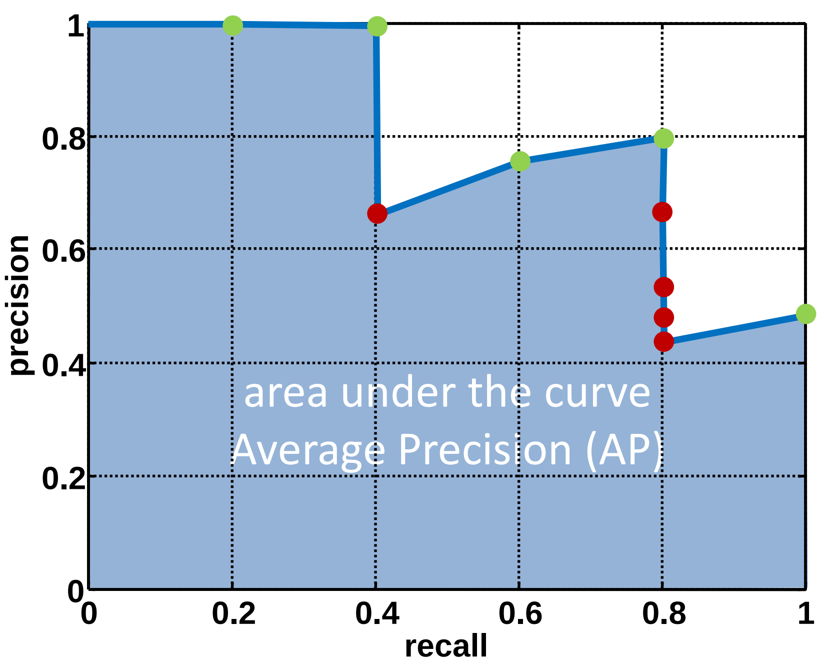
- Affine invariant features
- Efficient descriptors
- Corresponding regions in images have similar descriptors – measured by some distance in the features space
- Images of the same object have many correspondences in common

Retrieval Quality

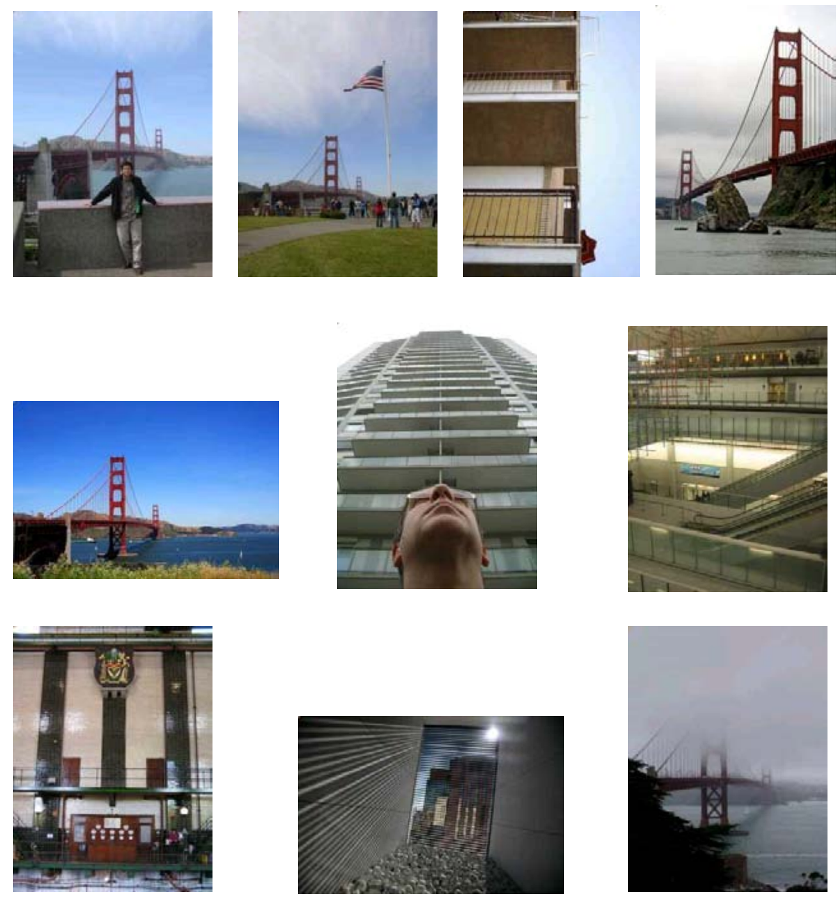
Query

Database size: 10 images
Relevant (total): 5 images

precision = #relevant / #returned
recall = #relevant / #total relevant



Results (ordered):

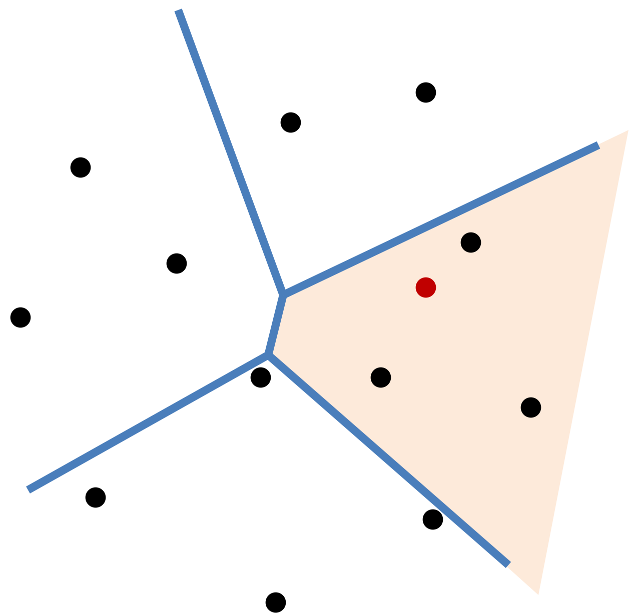


Video Google

- Feature detection and description
- Vector quantization
- Bag of Words representation
- Scoring
- Verification

Sivic & Zisserman – ICCV 2003
Video Google: A Text Retrieval Approach to Object Matching in Videos

Feature Distance Approximation

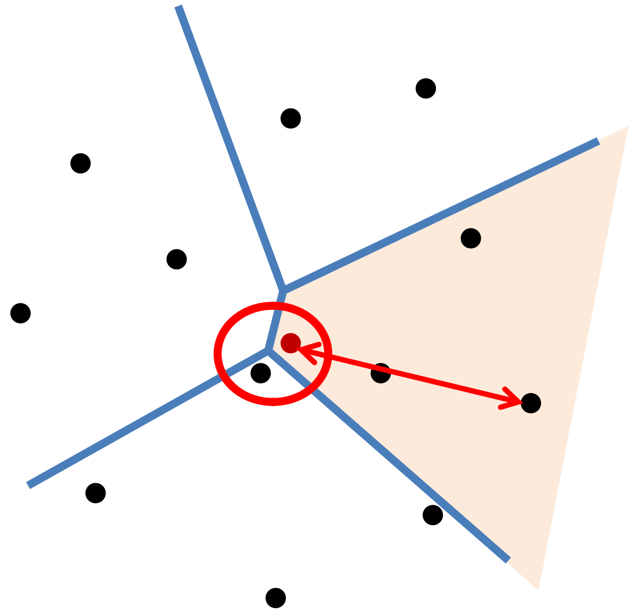


Partition the feature space
(k – means clustering)

Feature distance
0 : features in the same cell
∞ : features in different cells

+ most of the features are not considered (infinitely distant)
+ near-by descriptors accessible instantly – storing a list of features for each cell

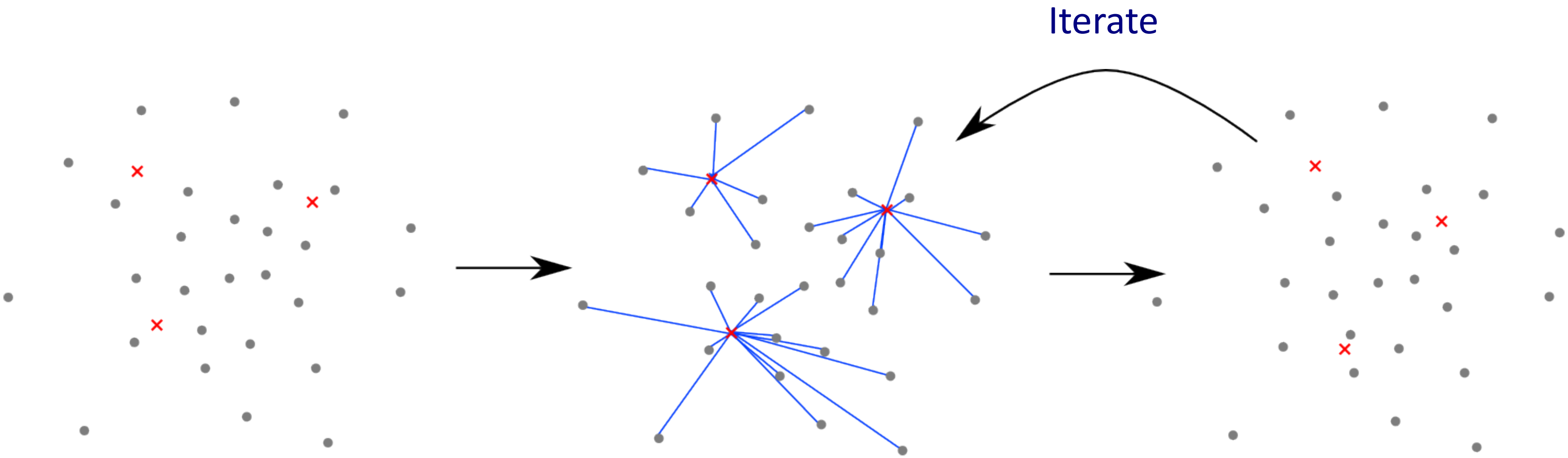
Feature Distance Approximation



Feature distance
0 : features in the same cell
∞ : features in different cells

- quantization effects
- large (even unbounded) cells

Vector Quantization via k-Means

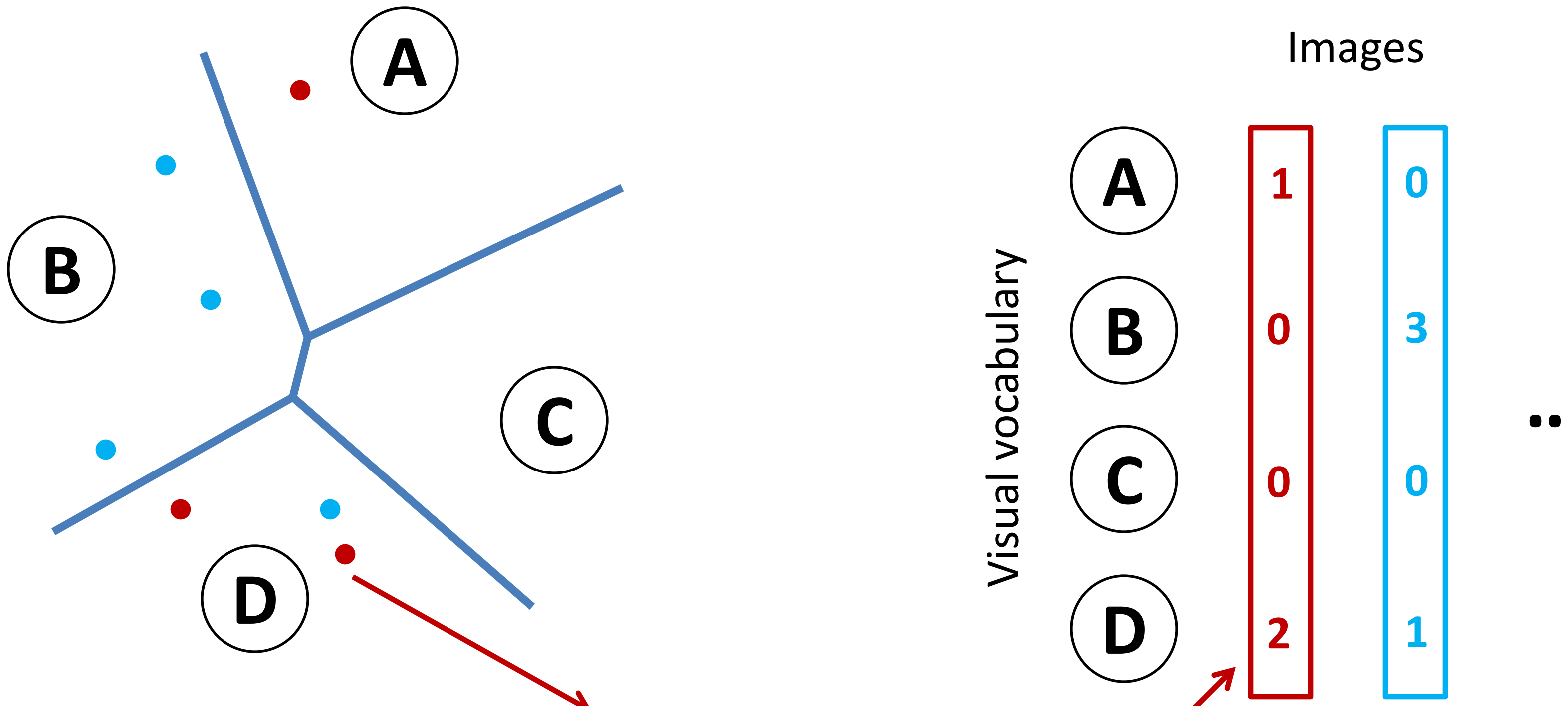


Initialize cluster centres

Find nearest cluster to each datapoint (**slow**) O(N k)

Re-compute cluster centres as centroids

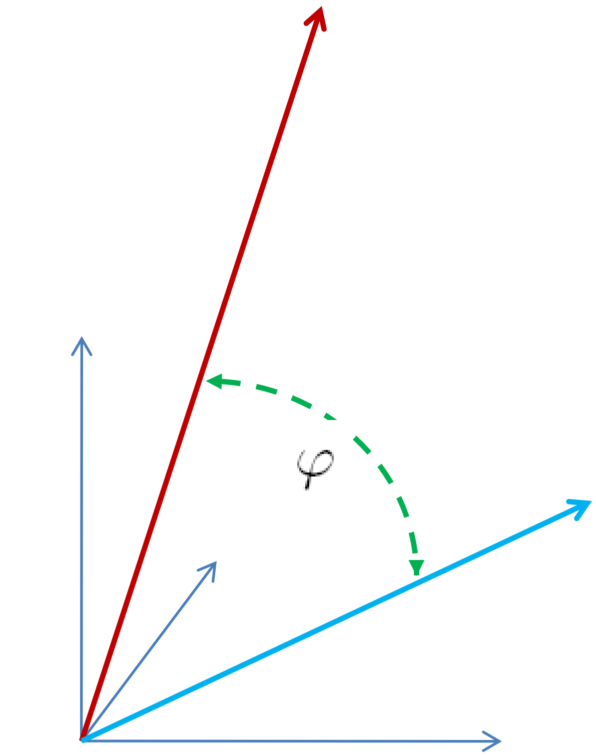
Bags of Words



Term-frequency (tf) – visual word D is twice in the image

Images are represented by sparse vector / histogram of visual words present in them

Efficient Scoring



bag of words representation
(up to 1,000,000 D)

$$\cos\varphi = \frac{\mathbf{x}\cdot\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||} = \frac{1}{||\mathbf{x}||\,||\mathbf{y}||}\sum_{i=1}^{N} x_i y_i$$

$$\sum_{x_i \neq 0,\, y_i \neq 0} x_i y_i$$

Database

	Ⓐ	Ⓑ	Ⓒ	Ⓓ
α_1	(1	0	0	2)
α_2	(0	2	0	1)
α_3	(1	0	0	0)
		⋮		

•

Query

Ⓐ	0
Ⓑ	3
Ⓒ	0
Ⓓ	1

=

Score

s_1 / α_q
s_2 / α_q
s_3 / α_q
⋮

Inverted files

$$
\begin{array}{cccc}
\text{Database} & & \text{Query} & & \text{Score} \\
\begin{array}{c|cccc}
 & A & B & C & D \\
\hline
\alpha_1 & 1 & 0 & 0 & 2 \\
\alpha_2 & 0 & 2 & 0 & 1 \\
\alpha_3 & 1 & 0 & 0 & 0 \\
 & \vdots & & &
\end{array}
& \bullet &
\begin{array}{c} 0 \\ 3 \\ 0 \\ 1 \end{array}
& = &
\begin{array}{c} s_1 / \alpha_q \\ s_2 / \alpha_q \\ s_3 / \alpha_q \\ \vdots \end{array}
\end{array}
$$

$$\sum_{x_i \neq 0, y_i \neq 0} x_i y_i$$

A	B	C	D
1 (1)	**2** (2)	$\perp$	**1** (2)
3 (1)	$\perp$		**2** (1)
$\perp$			$\perp$

Inverted file (posting list)
list of documents containing
certain visual word

Word Weighting

Words (in text) common to many documents are less informative
- ‘the’, ‘and’, ‘or’, ‘in’, …

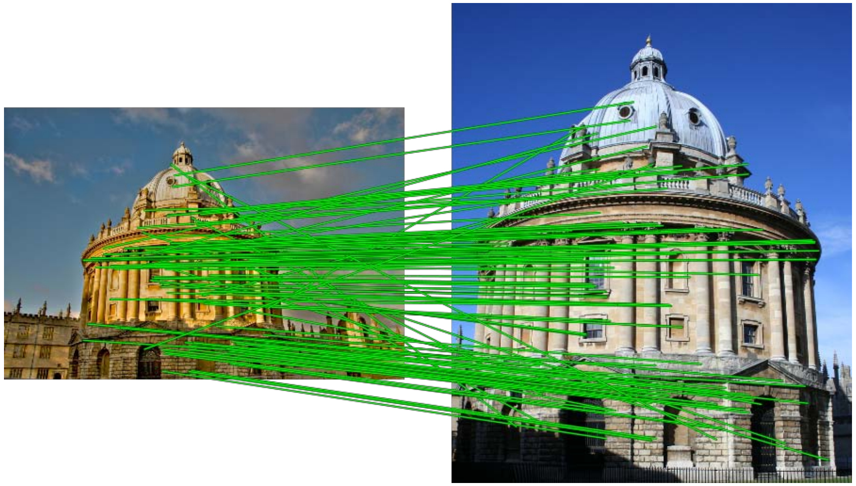
$$idf_X = \log \frac{\#\text{ documents}}{\#\text{ docs containing } X}$$

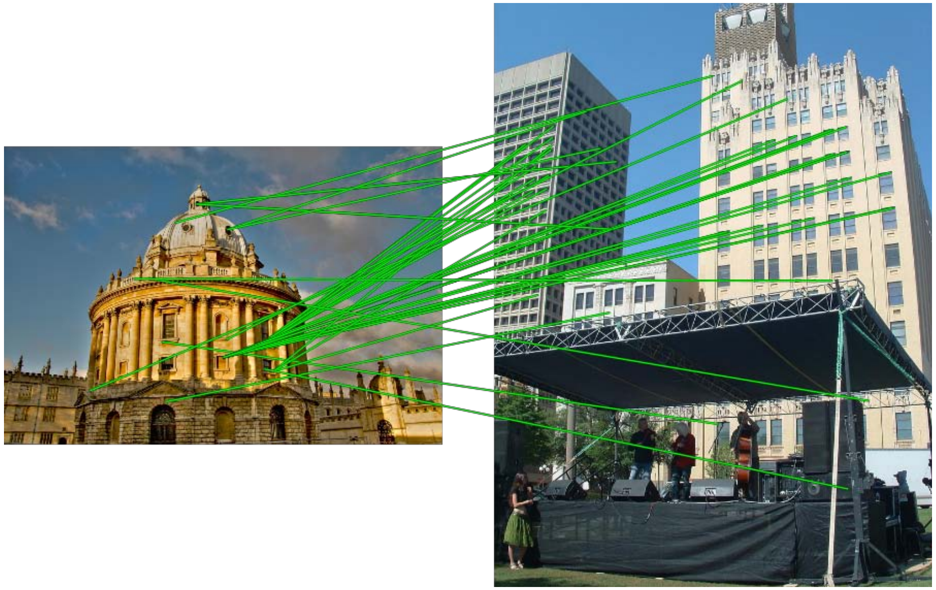
Images are represented by weighted histograms $tf_X \, idf_X$
(rather than just a histogram of tf_X)

Words that are too frequent (virtually in every document) can be put on a stop list
(ignored as if they were not in the document)

Baeza-Yates, Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999.

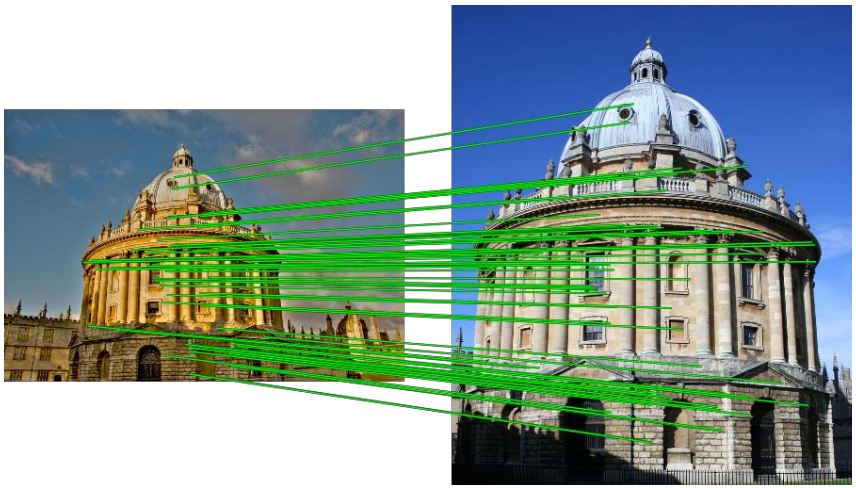
Spatial Verification





Both image pairs have many visual words in common
Look at the position and shape of the features

Spatial Verification

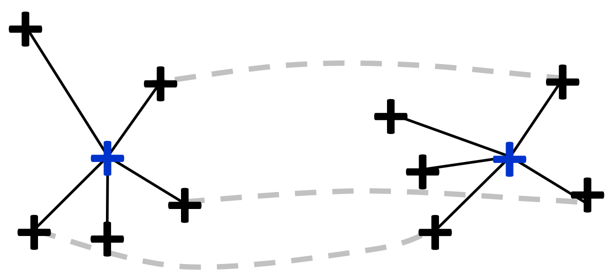


Only some of the correspondences are mutually consistent

(View Point Invariant) Spatial Verification

Weak geometric constraints

neighbourhoods of matching points must match

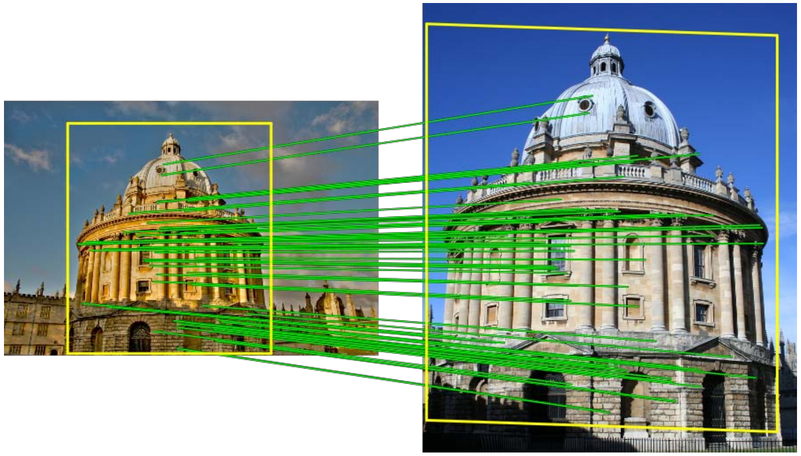


can be computed locally

Schmid and Mohr - PAMI 1997
Local Greyvalue Invariants for Image Retrieval

RANSAC – like estimation:

hypothesize transformation
verify consensus



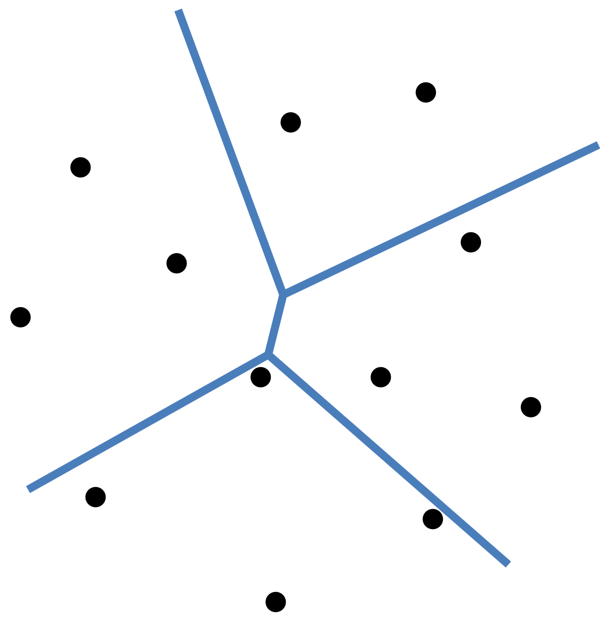
provides localization

Chum, Matas, and Obdržálek - ACCV 2004
Enhancing RANSAC by Generalized Model Optimization

Vector Quantization

- k-means
- Fixed quantization [Tuytelaars and Schmid ICCV 2007]
- Agglomerative [Leibe, Mikolajczyk and Schiele BMVC 2006]
- Hierarchical k-means
- Approximate k-means

Visual Vocabulary



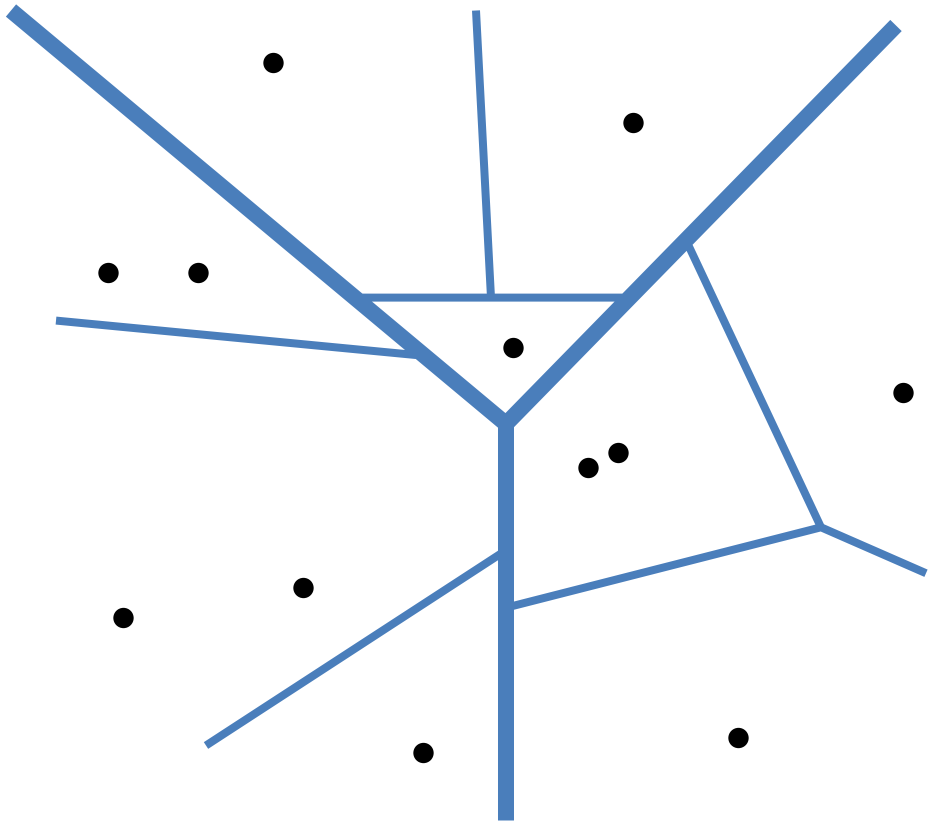
How many clusters in k-means?

- O (*k* N) – slow for large *k*
- The larger *k* the fewer tentative matches
- Experimentally – higher *k* better retrieval

Which data to cluster?

- Features from the database to be searched
  - better performance
- Some other fixed training set
- Universal vocabulary???

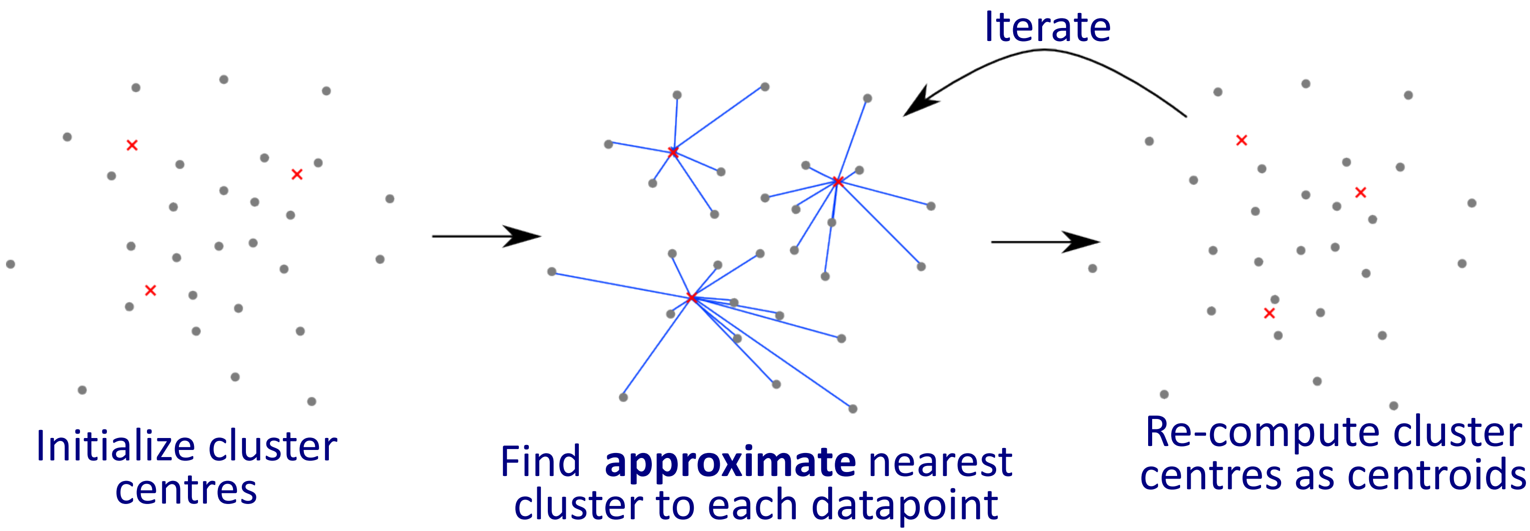
Hierarchical k-means



+ fast O(N log k)
+ incremental construction
- not so good quantization

Nistér & Stewénius: Scalable recognition with a vocabulary tree. CVPR 2006

Approximate k-means



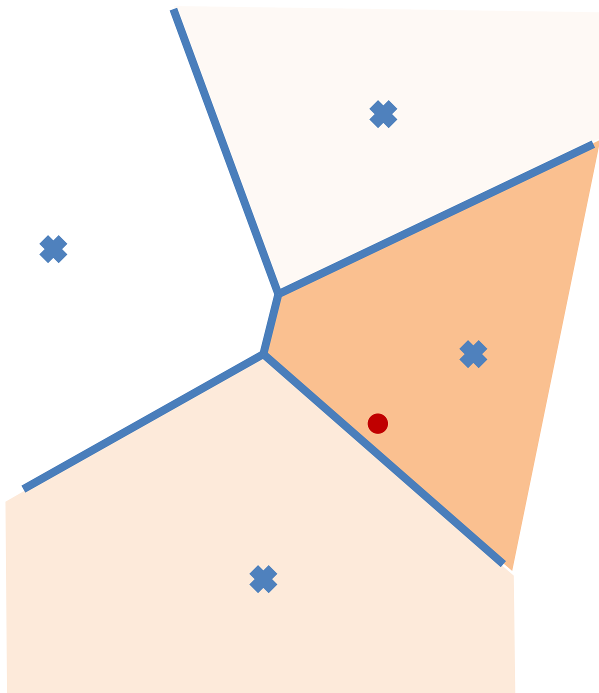
- \+ fast O(N log k)
- \+ reasonable quantization
- \- Can be inconsistent when ANN fails

Philbin, Chum, Isard, Sivic, and Zisserman – CVPR 2007
Object retrieval with large vocabularies and fast spatial matching

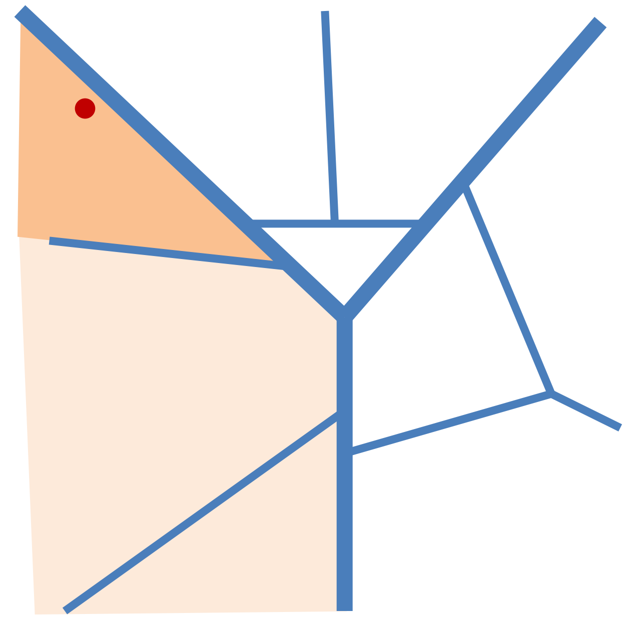
Approximate Nearest Neighbour
kd forest

D. Lowe. Distinctive image features from scale-invariant keypoints. IJCV 2004

Soft Assignment



(Approximate) k-means
- database side
- query side



Hierarchical k-means

Philbin, Chum, Isard, Sivic, and Zisserman – CVPR 2008
Lost in Quantization

Nistér & Stewénius – CVPR 2006 Scalable
recognition with a vocabulary tree

Query Expansion

Automatic Relevance Feedback

Using Results to Improve the Query

Query: ***golf green***

Results:

- How can the grass on the ***greens*** at a ***golf*** course be so perfect?
- For example, a skilled ***golf***er expects to reach the ***green*** on a par-four hole in **...**
- Manufactures and sells synthetic ***golf*** putting ***green****s* and mats.

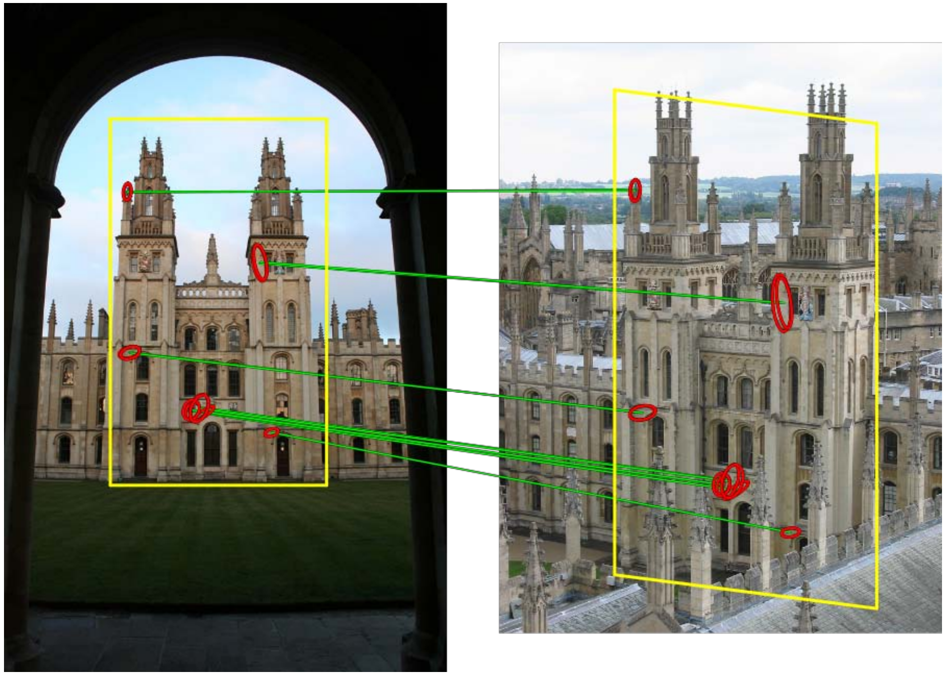
Irrelevant result can cause a `topic drift':

- Volkswagen ***Golf***, 1999, ***Green***, 2000cc, petrol, manual, , hatchback, 94000miles, 2.0 GTi, 2 Registered Keepers, HPI Checked, Air-Conditioning, Front and Rear Parking Sensors, ABS, Alarm, Alloy

Query Expansion

Chum, Philbin, Sivic, Isard, Zisserman: Total Recall..., ICCV 2007

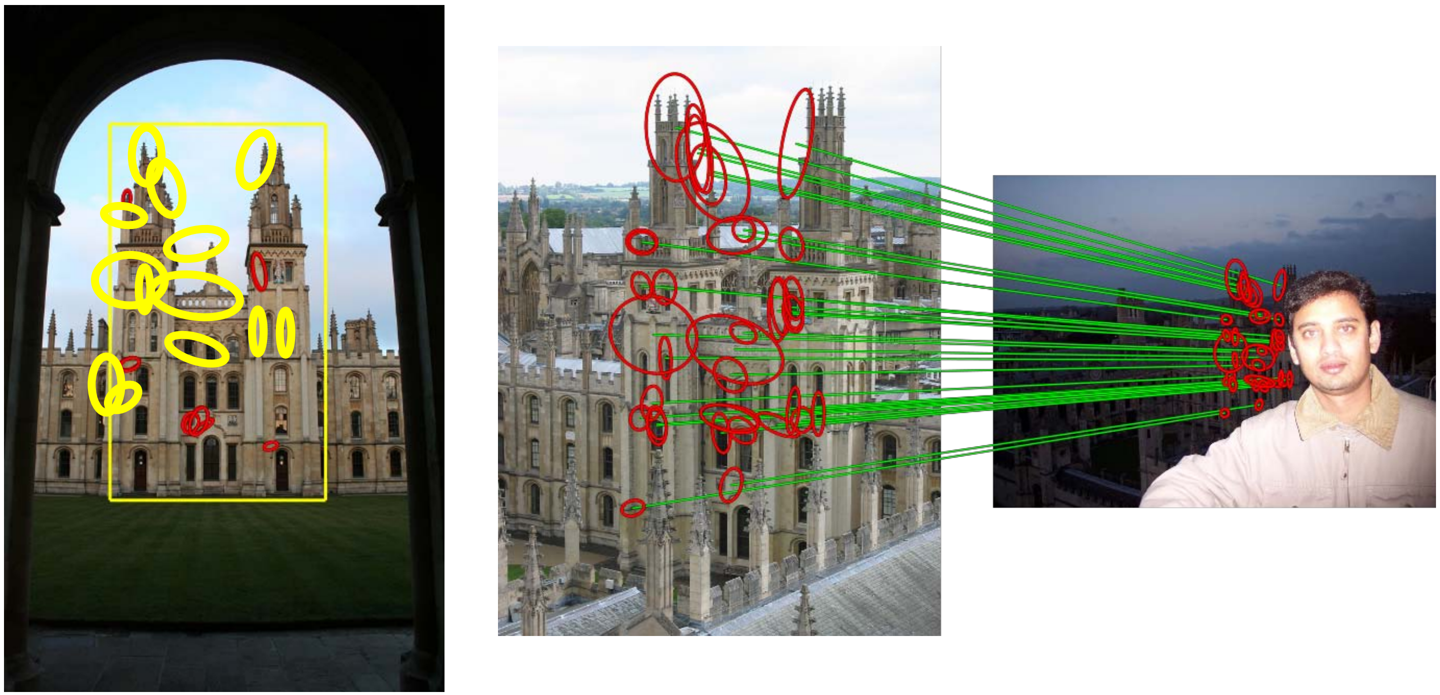
Query Expansion Step by Step



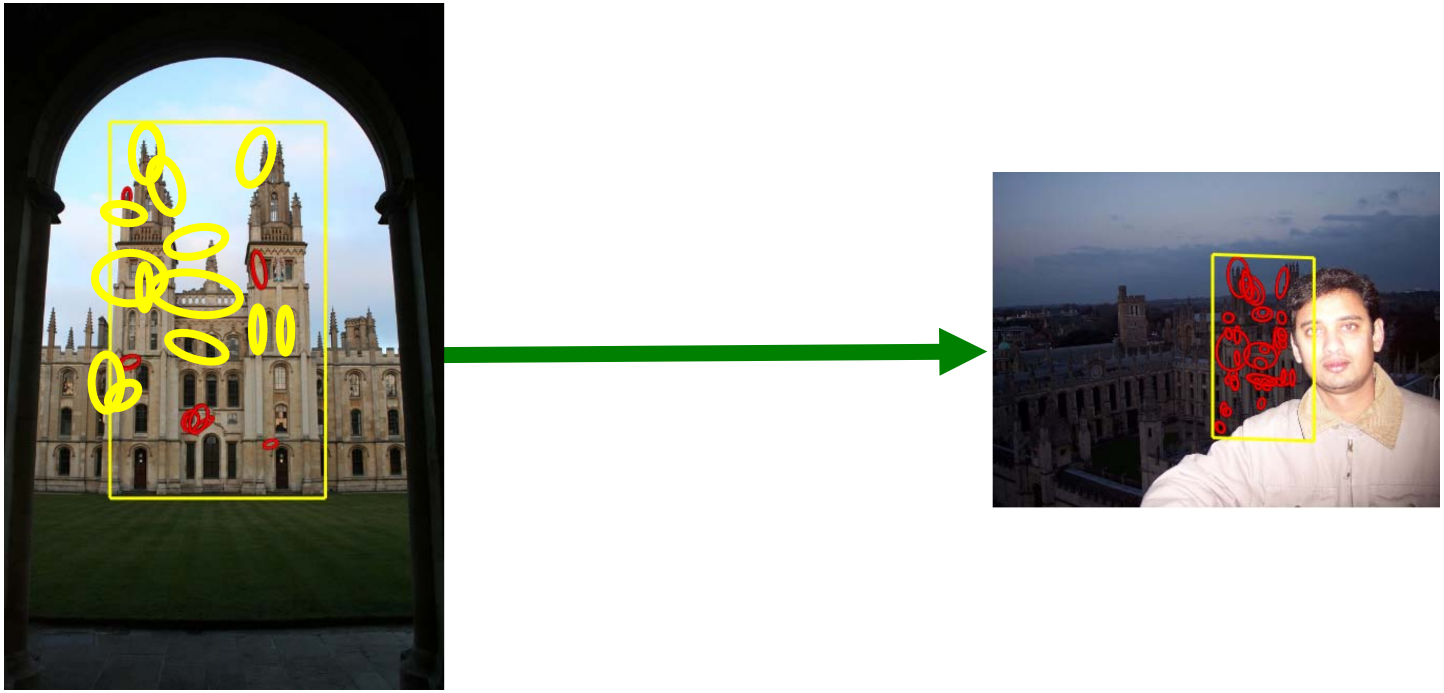
Query Image Retrieved image

Originally not retrieved

Query Expansion Step by Step



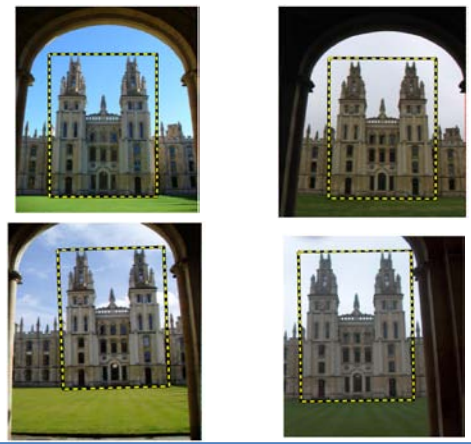
Query Expansion Step by Step

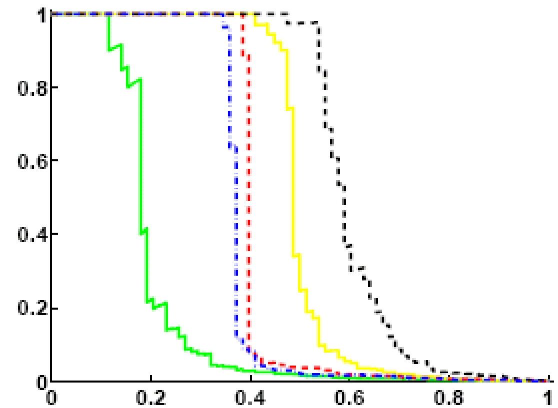


Query Expansion Results

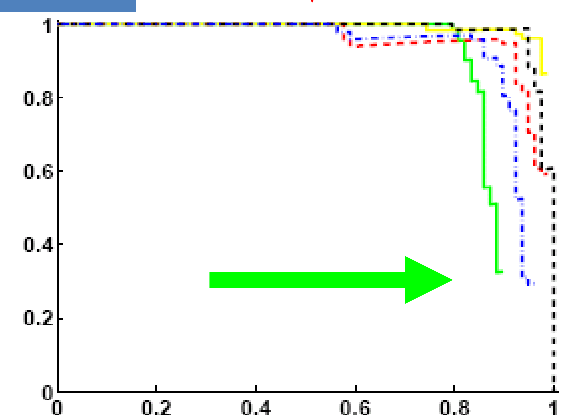
Query image

Original results (good)





Expanded results (better)



Conclusion

- Basic image retrieval is easy
  - Visual vocabulary be vector quantization to approximate distance between features
  - Bag of words representation
  - Efficient scoring function
  - Re-ranking via spatial verification
- Automatic query expansion
  - Geometry prevents thetopic drift