Min-Hashing and Geometric min-Hashing

Ondřej Chum, Michal Perdoch, and Jiří Matas

Center for Machine Perception
Czech Technical University
Prague

Outline

1. Looking for representation of images that:
   - is compact (useful for very large datasets)
   - is fast (linear in the size of the output)
   - supports local (image and object) recognition/retrieval
   - is accurate (very low false positive rate)

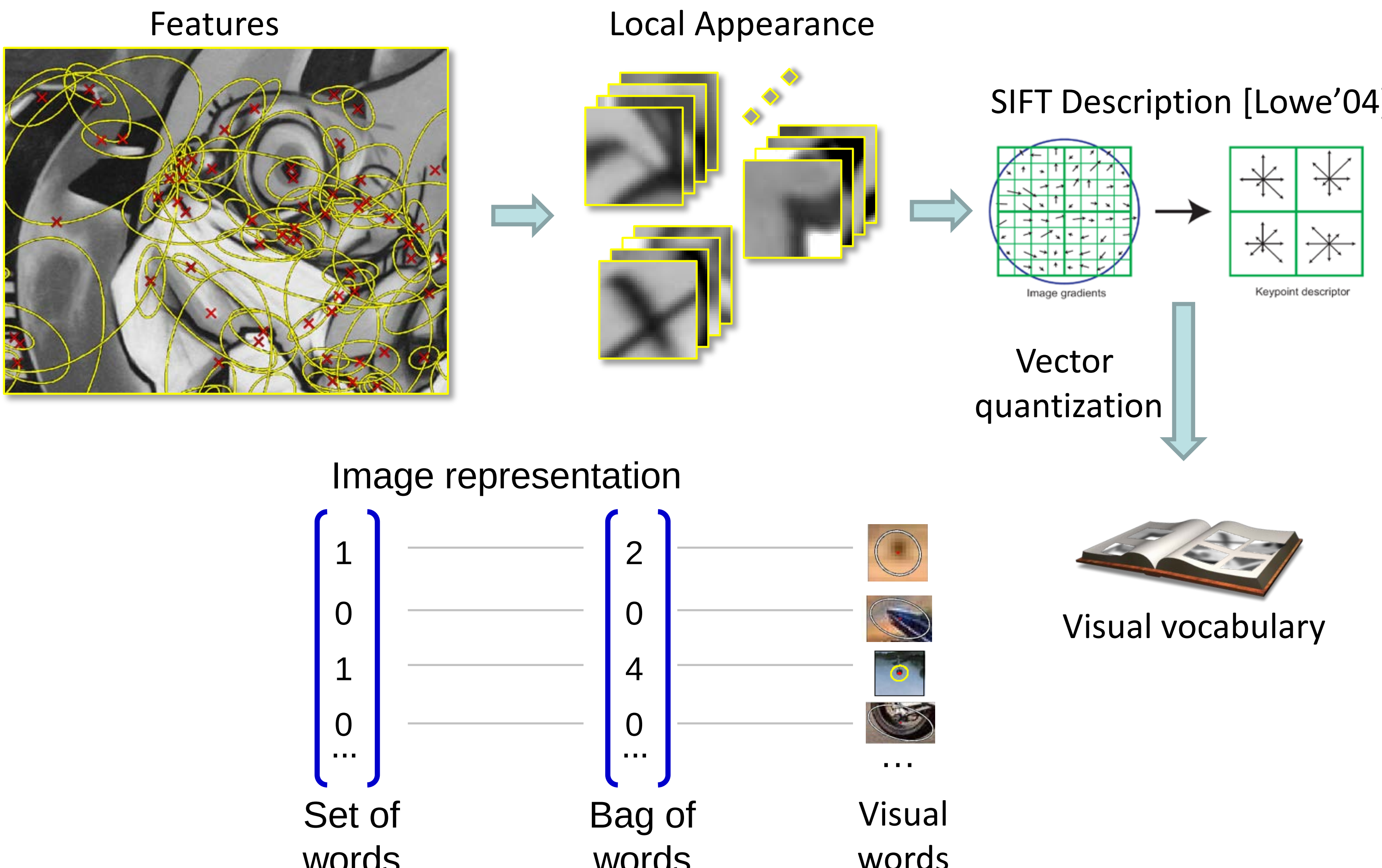
2. min-Hash is a powerful representation, but we show that Geometric min-Hash is significantly more powerful

3. Applications:
   - Large database clustering
   - Discovery of (even small) objects

Introduction to min-Hash for Images

min-Hash originates from the text retrieval community, originally used for detection of near duplicate documents

Image Representation: a Set of Words



min-Hash

min-Hash is a locality sensitive hashing (LSH) function m that selects an element (visual word) $m(\mathcal{I}_i)$ from each set $\mathcal{I}_i$ of visual words detected in image *i* so that

$$P\{m(\mathcal{I}_1) == m(\mathcal{I}_2)\} = \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1 \cup \mathcal{I}_2|}$$

This probability will be called "image similarity " and denoted

$$\mathrm{sim}(\mathcal{I}_1, \mathcal{I}_2) = \frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1 \cup \mathcal{I}_2|}$$

min-Hash

Vocabulary

A B C D E F

vocabulary of 1 M visual words

Set $\mathcal{I}_1$

A B C E F

~ 2000 words per image

Set $\mathcal{I}_2$

B C D F

Random orderings

3 6 2 5 4 <u>1</u>

<u>1</u> <u>2</u> 6 3 5 4

3 2 <u>1</u> 6 4 5

min-Hash

m_1 : F F

m_2 : A B

m_3: C C

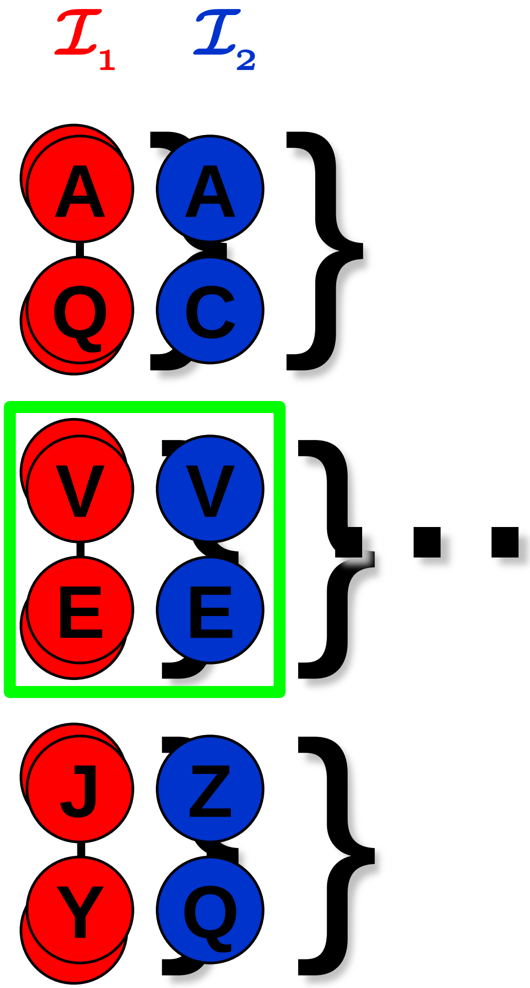
fixed size image representation ~100

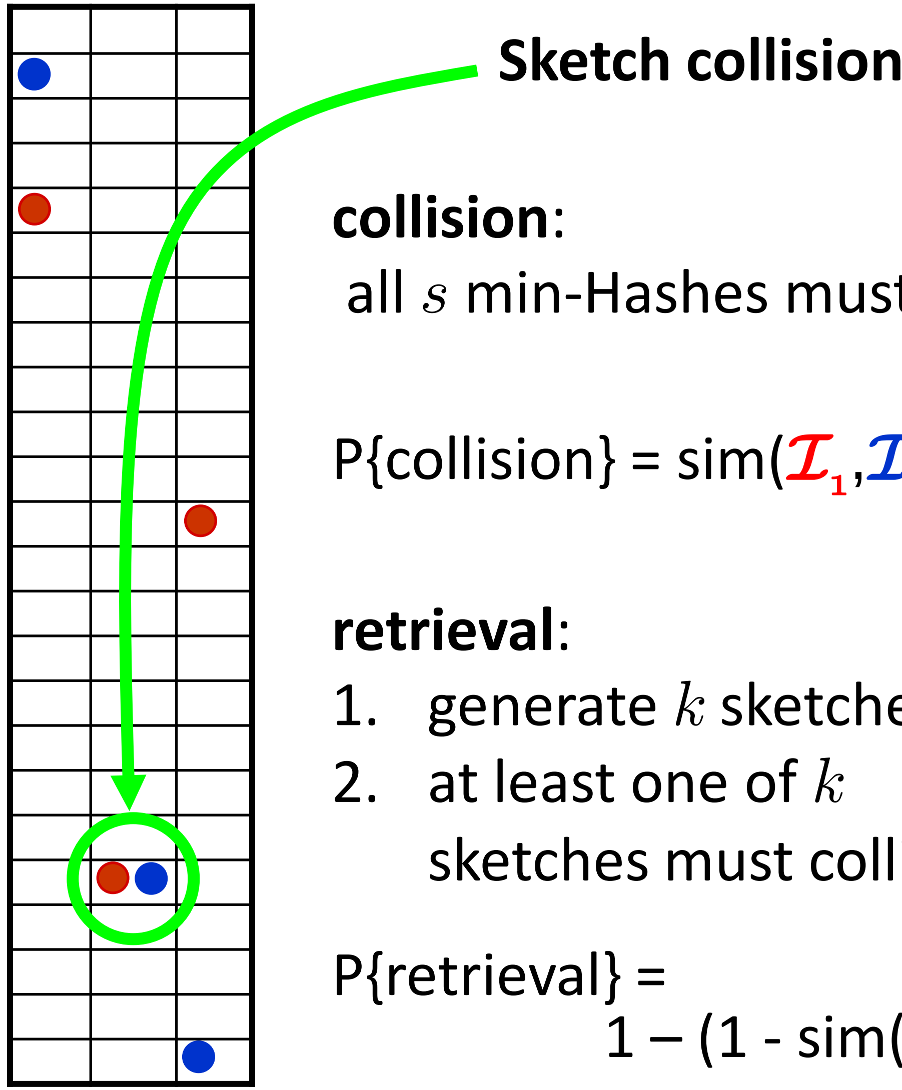
sim ($\mathcal{I}_1$, $\mathcal{I}_2$) = 1/2

Estimated similarity of $\mathcal{I}_1$ and $\mathcal{I}_2$ from 3 min-Hashes = 2/3

min-Hash Retrieval

a sketch = s-tuple of min-Hashes





k hash tables

collision:
all s min-Hashes must agree

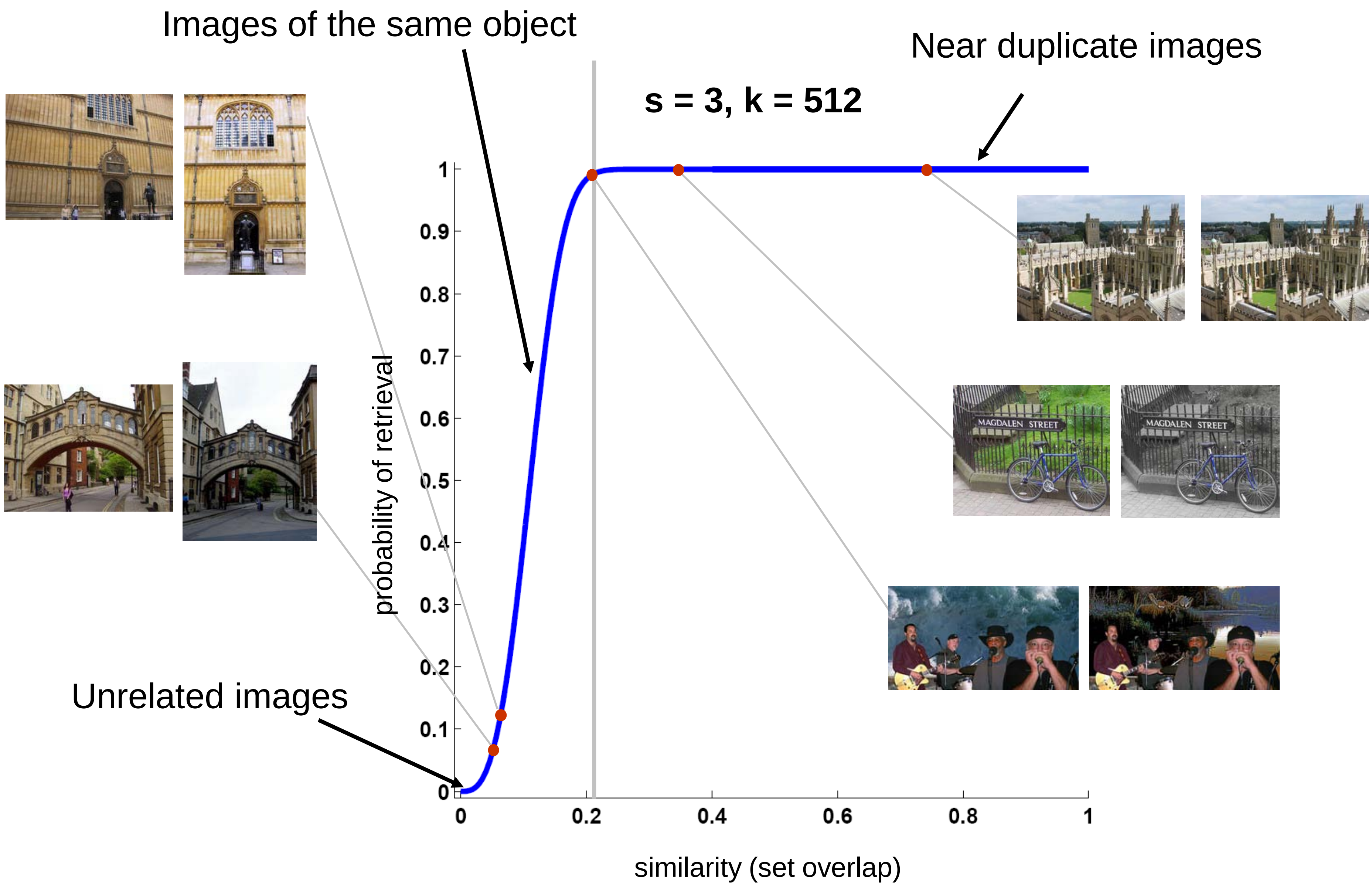
P{collision} = sim($\mathcal{I}_1$,$\mathcal{I}_2$)^s

retrieval:

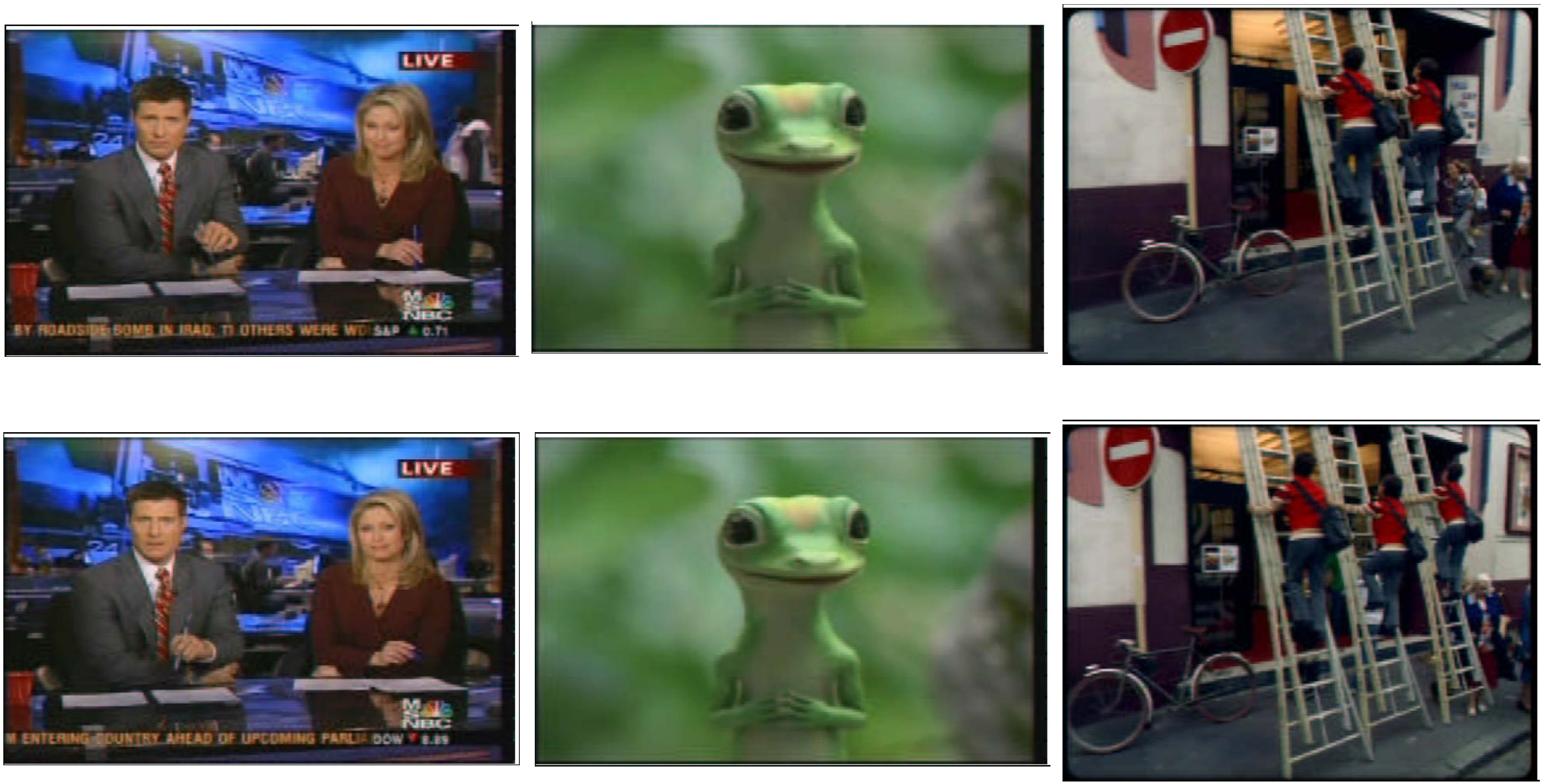
1. generate k sketches
2. at least one of k sketches must collide

P{retrieval} =
$1 - (1 - \mathrm{sim}(\mathcal{I}_1, \mathcal{I}_2)^s)^k$

Probability of Retrieving an Image Pair



Near Duplicate Images



Near Duplicate Images

Near Duplicate Images

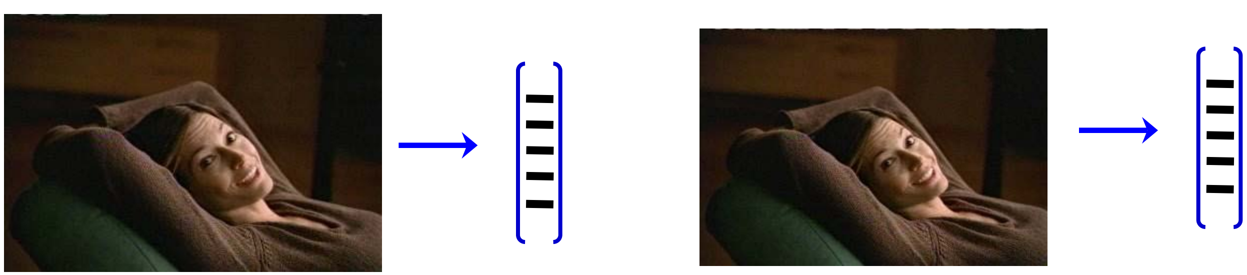
Scalable Near Duplicate Image Detection

- Images perceptually (almost) identical but not identical (noise, compression level, small motion, small occlusion)
- Similar images of the same object / scene
- Large databases
- Fast – linear in the number of duplicates
- Store small constant amount of data per image

Representation, Distance and Search

- Naïve method (space and time infeasible):
  - Representation: the whole image
  - Similarity: sophisticated visual similarity measure
  - Search: O(N^2) on N images

- Goal:
  - Compact representation – small constant number of bytes (that captures visual content addressed by near duplicate definition)
  - Similarity that allows for fast retrieval of similar images (that measures similarity according to near duplicate definition)
  - Search: linear in the number of near duplicates – needs to be indexing



Word Weighting for min-Hash

For hash function (set overlap similarity) $f_j(X_w) = x \quad x \sim \mathrm{Un}(1,0)$

all words X_w have the same chance to be a min-Hash

For hash function
$$\boxed{f_j(X_w) = \frac{-\log x}{d_w} \qquad x \sim \mathrm{Un}(1,0)}$$

the probability of X_w being a min-Hash is proportional to d_w

$\mathcal{A} \cup \mathcal{B}$: A C E J Q R V Y Z

d_A d_C d_E d_J d_Q d_R d_V d_Y d_Z

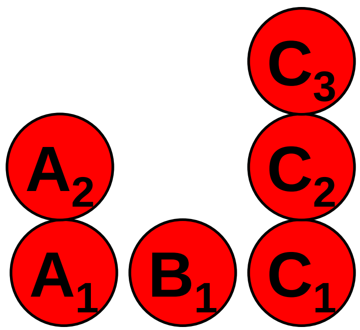
$$P(m(\mathcal{A}) = m(\mathcal{B})) = \frac{\sum_{X_w \in \mathcal{A} \cap \mathcal{B}} d_w}{\sum_{X_w \in \mathcal{A} \cup \mathcal{B}} d_w}$$

Histogram Intersection Using min-Hash

Idea: represent a histogram as a set, use min-Hash set machinery

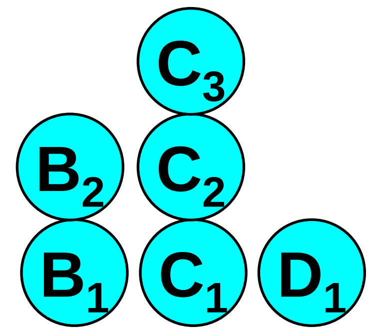
Visual words: A B C D

Bag of words ***A*** / set ***A'***



$t_A = (2,1,3,0)$

Bag of words ***B*** / set ***B'***



$t_B = (0,2,3,1)$

min-Hash vocabulary: A_1 A_2 B_1 B_2 C_1 C_2 C_3 D_1

A' U ***B'***: A_1 A_2 B_1 B_2 C_1 C_2 C_3 D_1

Set overlap of ***A'*** of ***B'*** is a histogram intersection of ***A*** and ***B***

Similarity Measures for min-Hash

	Set representation	Bag of words representation
Equal weights	$\mathrm{sim}_s(\mathcal{A}_1, \mathcal{A}_2) = \frac{\lvert \mathcal{A}_1 \cap \mathcal{A}_2 \rvert}{\lvert \mathcal{A}_1 \cup \mathcal{A}_2 \rvert}$	$\mathrm{sim}_{h_0}(\mathcal{A}_1, \mathcal{A}_2) = \frac{\sum_w \min(t_1^w, t_2^w)}{\sum_w \max(t_1^w, t_2^w)}$
Weighted	$\mathrm{sim}_w(\mathcal{A}_1, \mathcal{A}_2) = \frac{\sum_{X_w \in \mathcal{A}_1 \cap \mathcal{A}_2} d_w}{\sum_{X_w \in \mathcal{A}_1 \cup \mathcal{A}_2} d_w}$	$\mathrm{sim}_h(\mathcal{A}_1, \mathcal{A}_2) = \frac{\sum_w d_w \min(t_1^w, t_2^w)}{\sum_w d_w \max(t_1^w, t_2^w)}$

Min-Hash on TrecVid

- DoG features
- vocabulary of 64,635 visual words
- 192 min-Hashes, 3 min-Hashes per a sketch, 64 sketches
- similarity threshold 35%

Set overlap similarity measure

Query

Retrieved images with decreasing similarity

Query Examples

Query image:

Results

Set overlap, weighted set overlap, weighted histogram intersection

Geometric min-Hash (GmH)

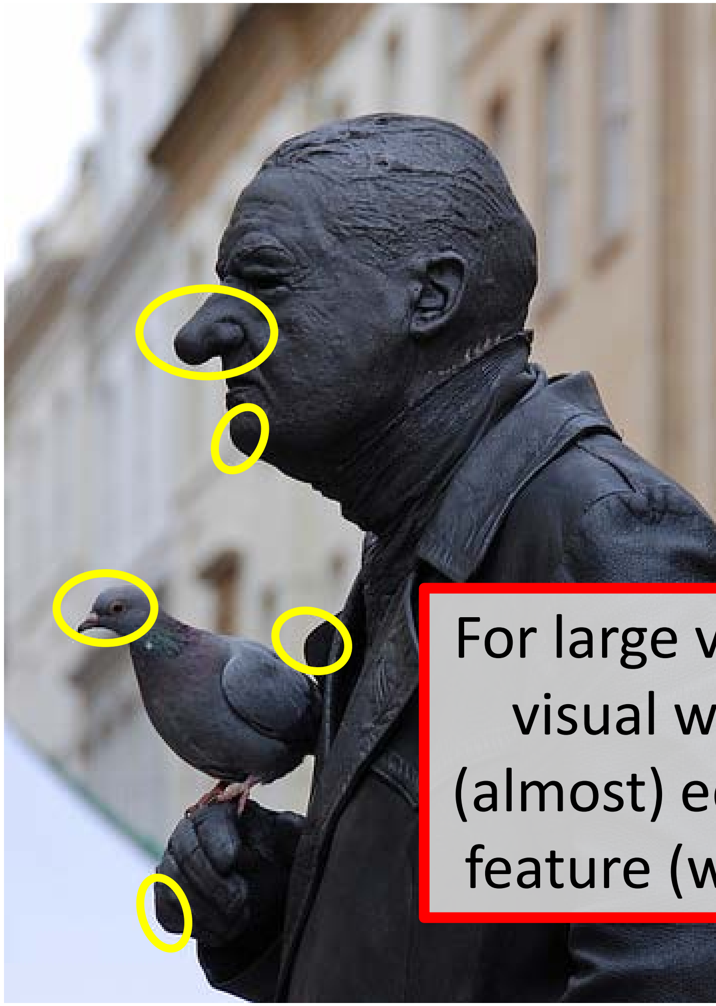
Can geometry help us in finding sketches of min-Hashes with much higher repeatability than random s-tuples?

The Idea

- For a sketch collision all s min-Hashes in the sketch must agree
- In the construction of a sketch, we can assume that the first min-Hash is matching
- If the assumption is violated, no harm is done, the sketch would not collide

We show how to exploit the assumption of matching $m_1(\mathcal{I})$ to design the rest of the sketch (**not independent min-Hashes** anymore). The new procedure – Geometric min-Hash - is superior to the standard min-Hash.

Vocabulary Size and Set Representation



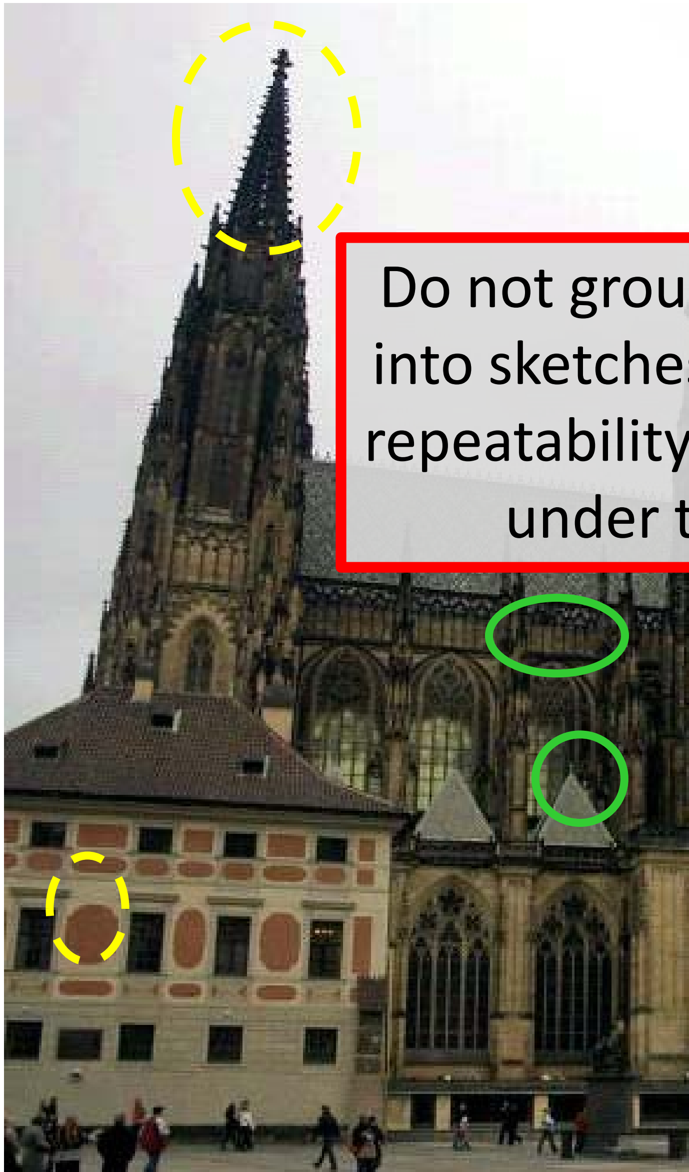
Small vocabulary (1k visual words)

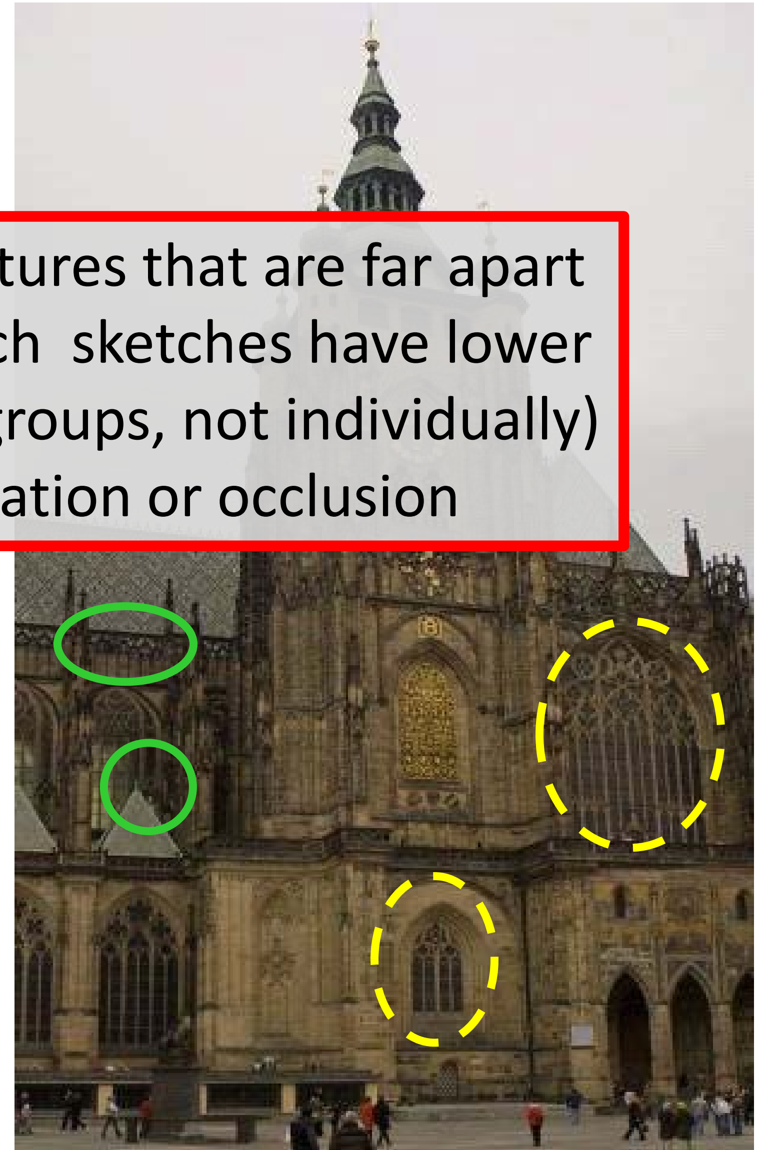
Large vocabulary (1M visual words)

For large vocabularies selecting a visual word from an image is (almost) equivalent to selecting a feature (with location and scale)

On a 100k dataset of images, 95% of features have a unique visual word in an image

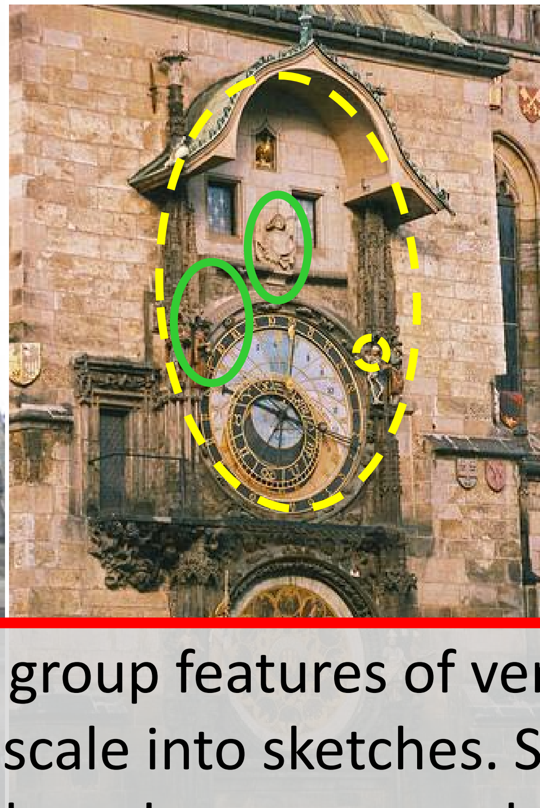
Repeatability of Feature Sets: Translation and Occlusion

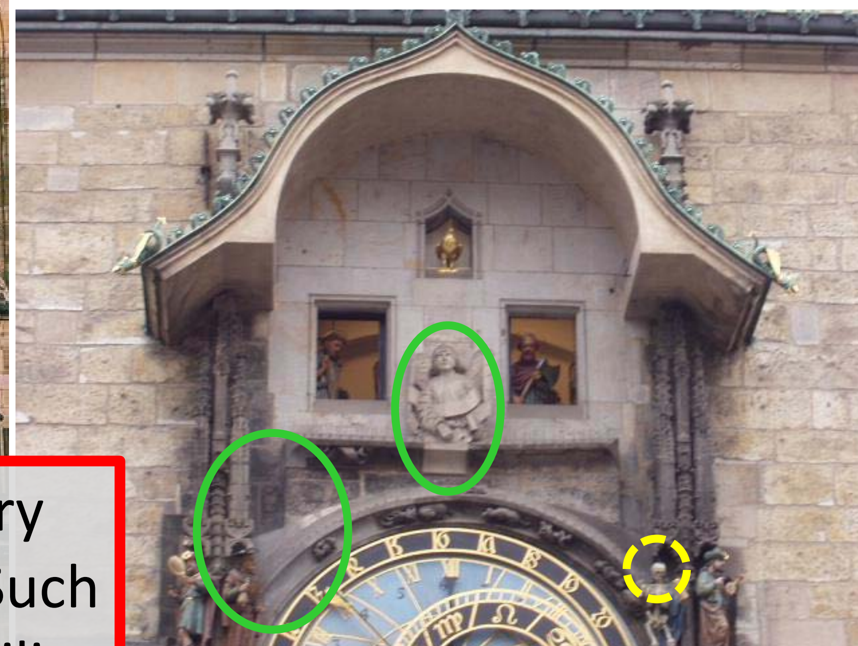




Do not group features that are far apart into sketches. Such sketches have lower repeatability (as groups, not individually) under translation or occlusion

Repeatability of Feature Sets: Scale Change



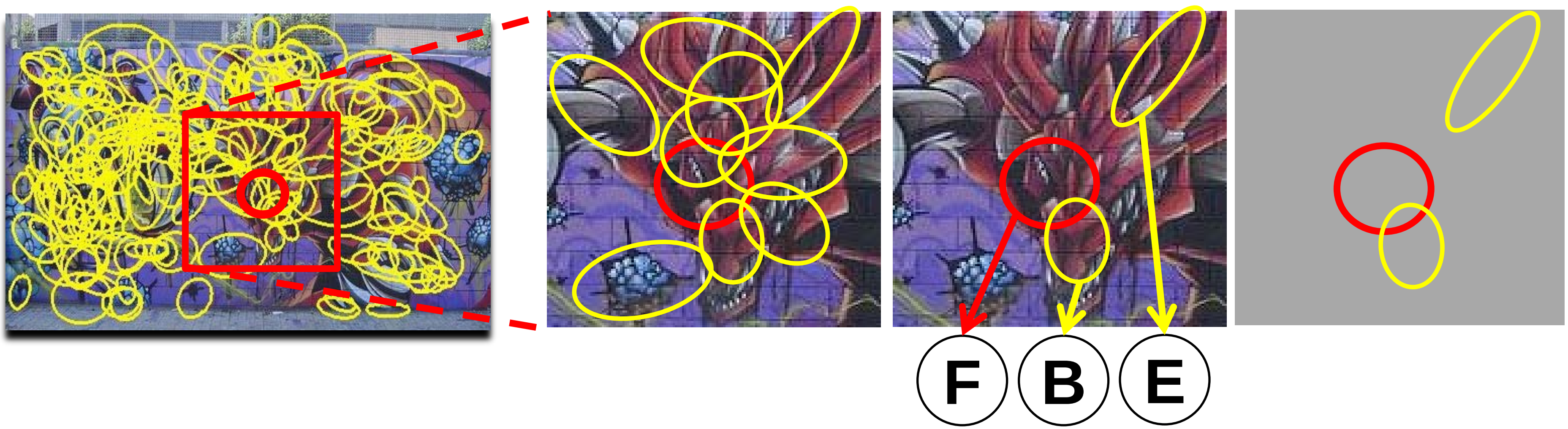


Do not group features of very different scale into sketches. Such sketches have lower repeatability under scale change.

Geometric min-Hash algorithm

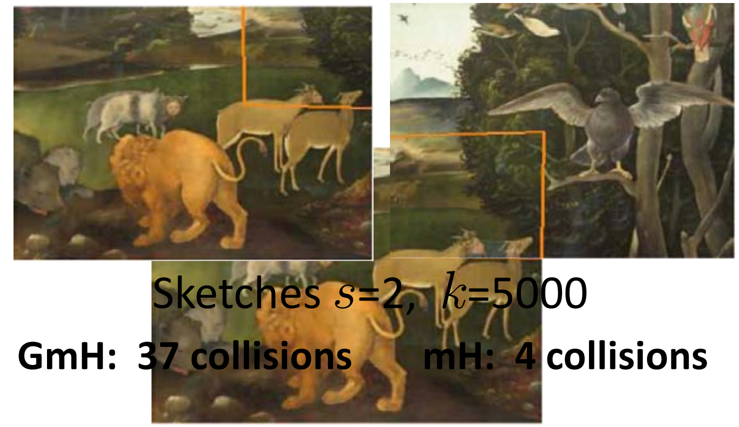
1. Keep features with unique visual word in the image
2. Obtain the “central feature” by min-Hash
3. Select scale and spatial neighbourhood of the central feature
4. Select secondary min-Hash(es) from the neighbourhood
5. Relative pose of the sketch features is a geometric invariant (as in geometric hashing)

Sketch of GmH: s-tuple of visual words + geometric invariant



Geometric vs. Standard min-Hash

- Higher true positives
  - View point change
  - Severe occlusion
  - Scale change
  - Object on a different background



Sketches s=2, k=5000

GmH: 37 collisions **mH: 4 collisions**

- Lower false positives
  - Additional geometric invariant (part of the hash key or verification)
  - Lower probability random sketch collisions (next slide)

Sketches s=2, k=5000

GmH: 0 collisions **mH: 4 collisions**

- Faster spatial verification
  - Sketch collision defines geometric transformation

Overlap of Random Sets

False positive = sketch collision of two random images

The probability of two random sets $\mathcal{I}_1$ and $\mathcal{I}_2$ having a common min-Hash (*i.e.* the average overlap of two random sets)

$$\frac{\min(|\mathcal{I}_1|, |\mathcal{I}_2|)}{2w} \leq \mathrm{E}\left(\frac{|\mathcal{I}_1 \cap \mathcal{I}_2|}{|\mathcal{I}_1 \cup \mathcal{I}_2|}\right) \leq \frac{\min(|\mathcal{I}_1|, |\mathcal{I}_2|)}{w}$$

where w is the size of the vocabulary

The smaller the sets, the smaller probability of random collision

- Min-Hash: features in the whole image
- Geometric min-Hash: only small subset of the image

Experiments

Experiment 1: Clustering on Oxford 100k DB

Randomized clustering

Cluster hypotheses by hashing, completion by retrieval

All Soul's

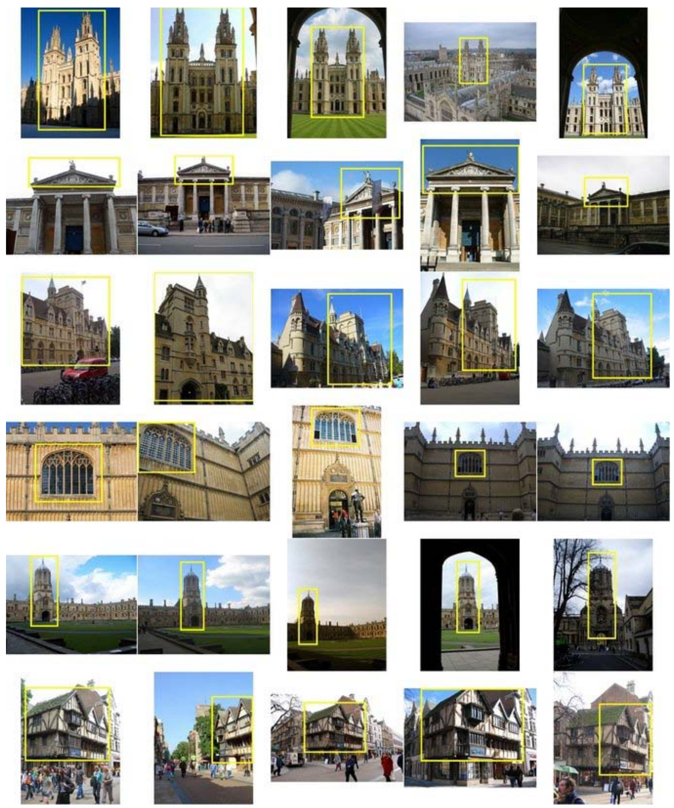
Ashmolean

Balliol

Bodleian

Christ Church

Cornmarket



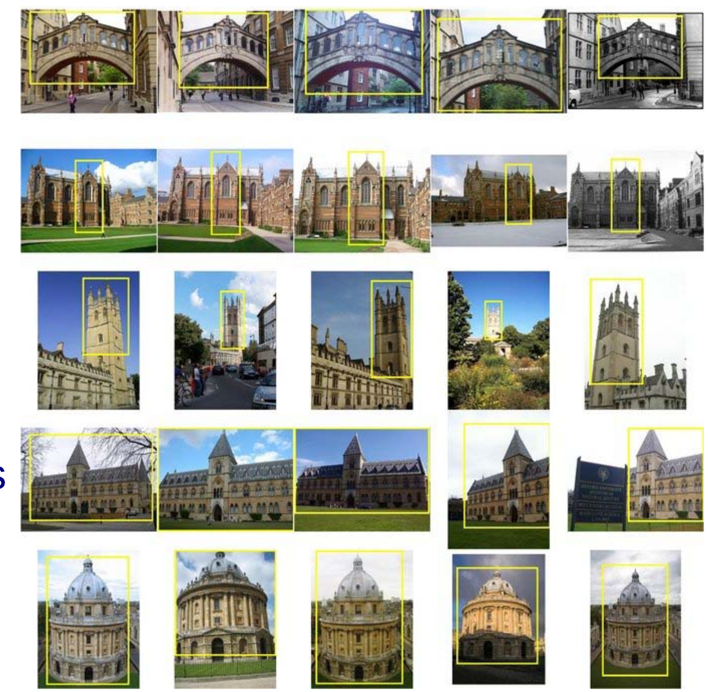
Hertford

Keble

Magdalen

Pitt Rivers

Radcliffe Camera



100 000 Images downloaded from FLICKR

Includes 11 Oxford Landmarks with manually labeled ground truth

Clustering Results on Oxford 100k

	Geometric min-Hash		[Chum TR 2008]	
	Component Recall	fp	Component Recall	fp
all souls	**98.72**	0	97.44	0
ashmolean	**76.00**	0	68.00	0
balliol	**91.67**	0	33.33	0
bodleian	**100**	0	95.83	1
christ church	**97.44**	1	89.74	0
cornmarket	**77.78**	1	66.67	0
hertford	**100**	0	96.30	1
keble	**100**	0	85.71	0
magdalen	**38.89**	0	5.56	0
pitt rivers	**100**	0	100	0
radcliffe	**99.55**	0	98.64	0

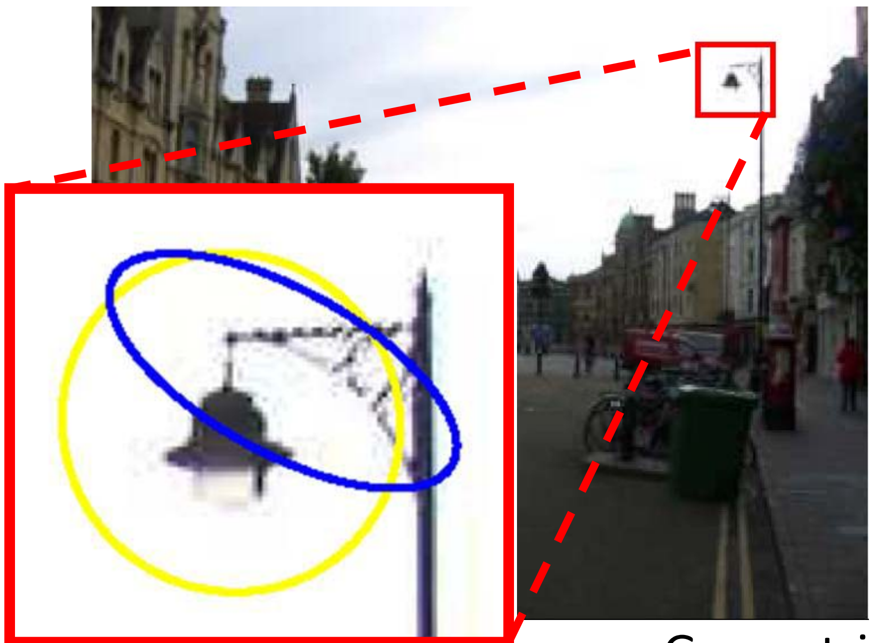
16 min* — 33 min*

s = 2, k = 64 < s = 3, k = 256

~550 bytes per image — ~1600 bytes per image

* The time does not include feature detection, SIFT computation, vector quantization

Experiment 2: Object Discovery



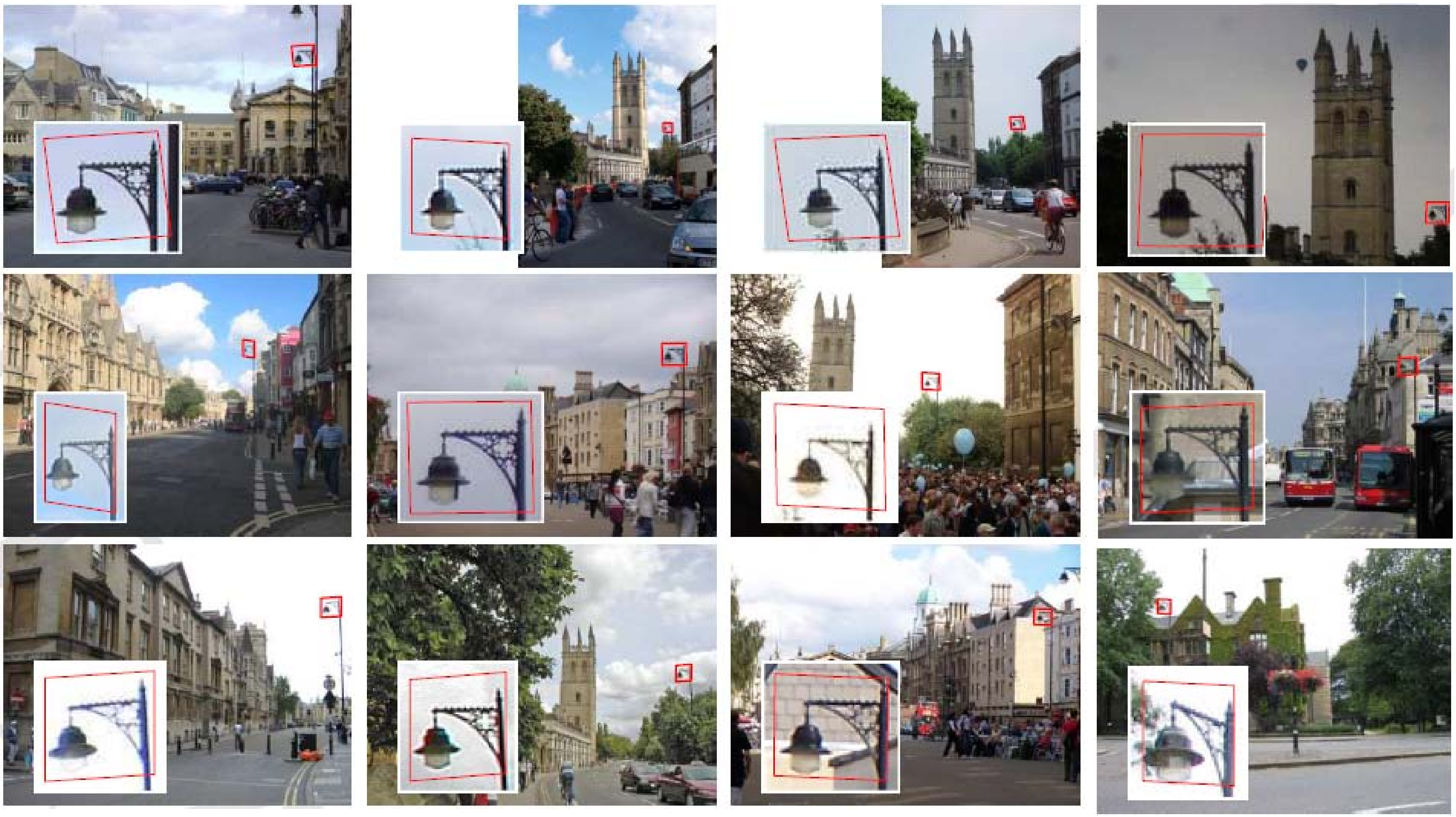
Geometric min-Hash
sketch collision
$s = 2, k = 256$

Verification by co-segmentation critical for small objects

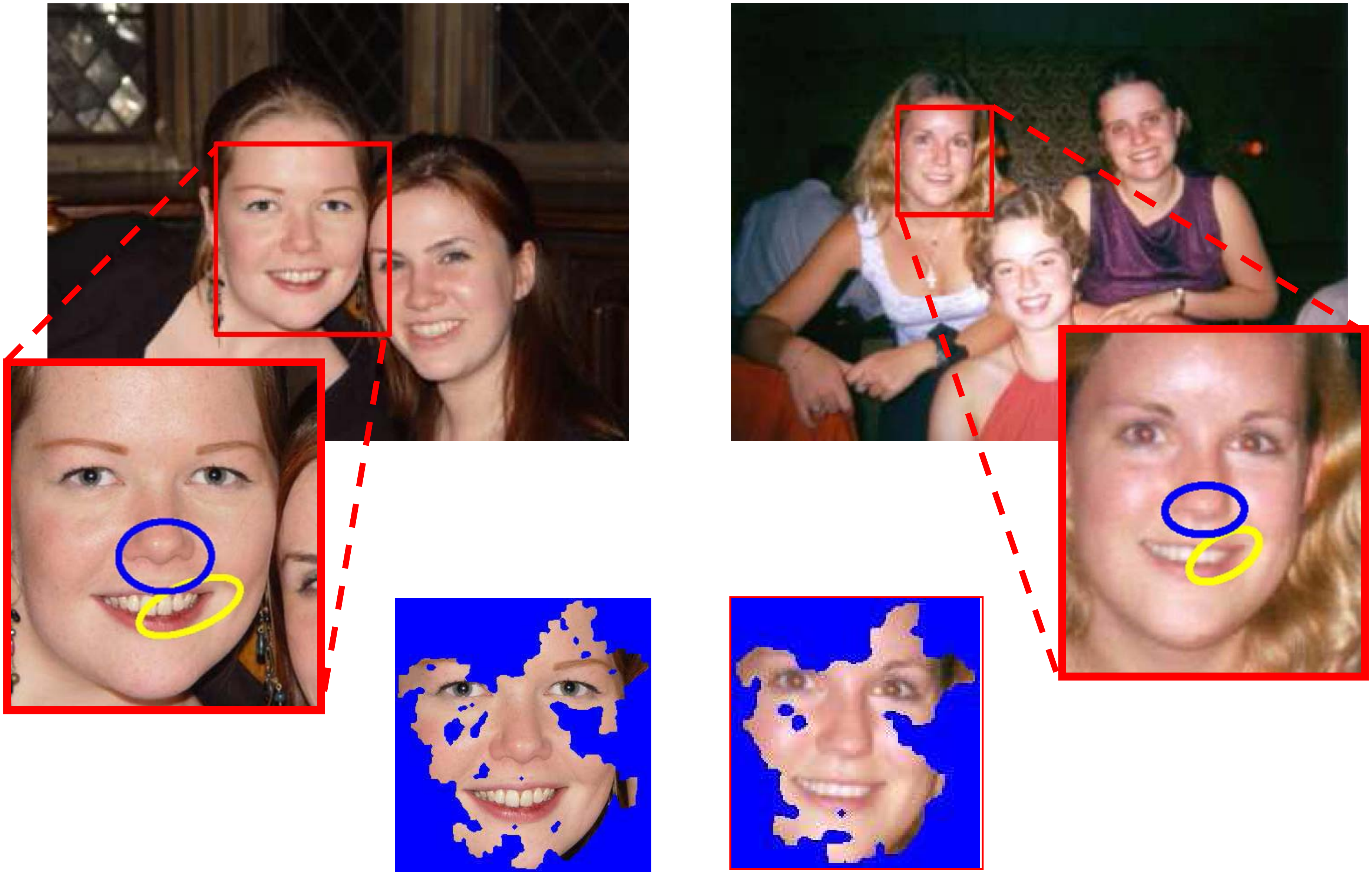
[Cech, Matas, Perdoch CVPR 08], code available on WWW
[Ferrari, Tuytelaars,Van Gool, ECCV 2004]

Small Object Discovery

Other instances of the discovered object by (sub)image retrieval

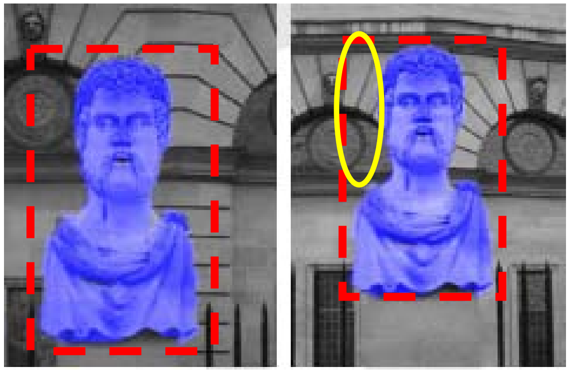


Faces are Small Objects

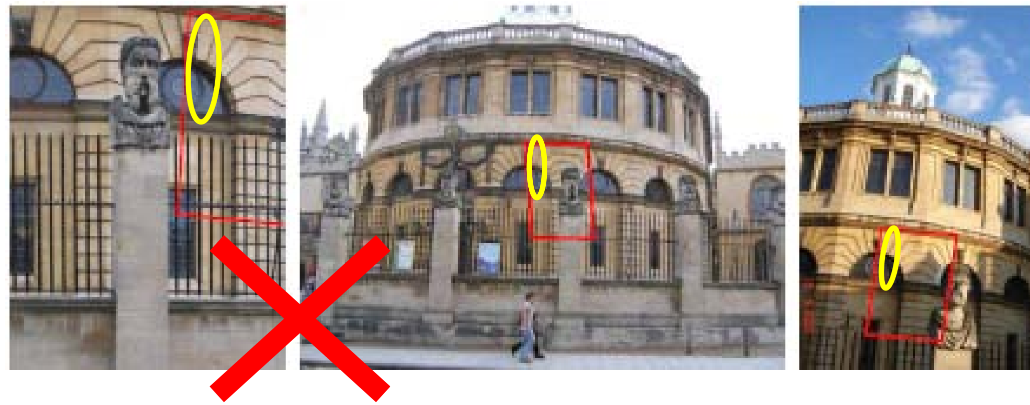


Discovery of the Face Category: the Largest Cluster in Oxford 100k

The Importance of the Co-segmentation in Object Discovery



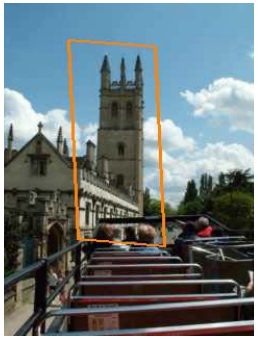
Seed image pair

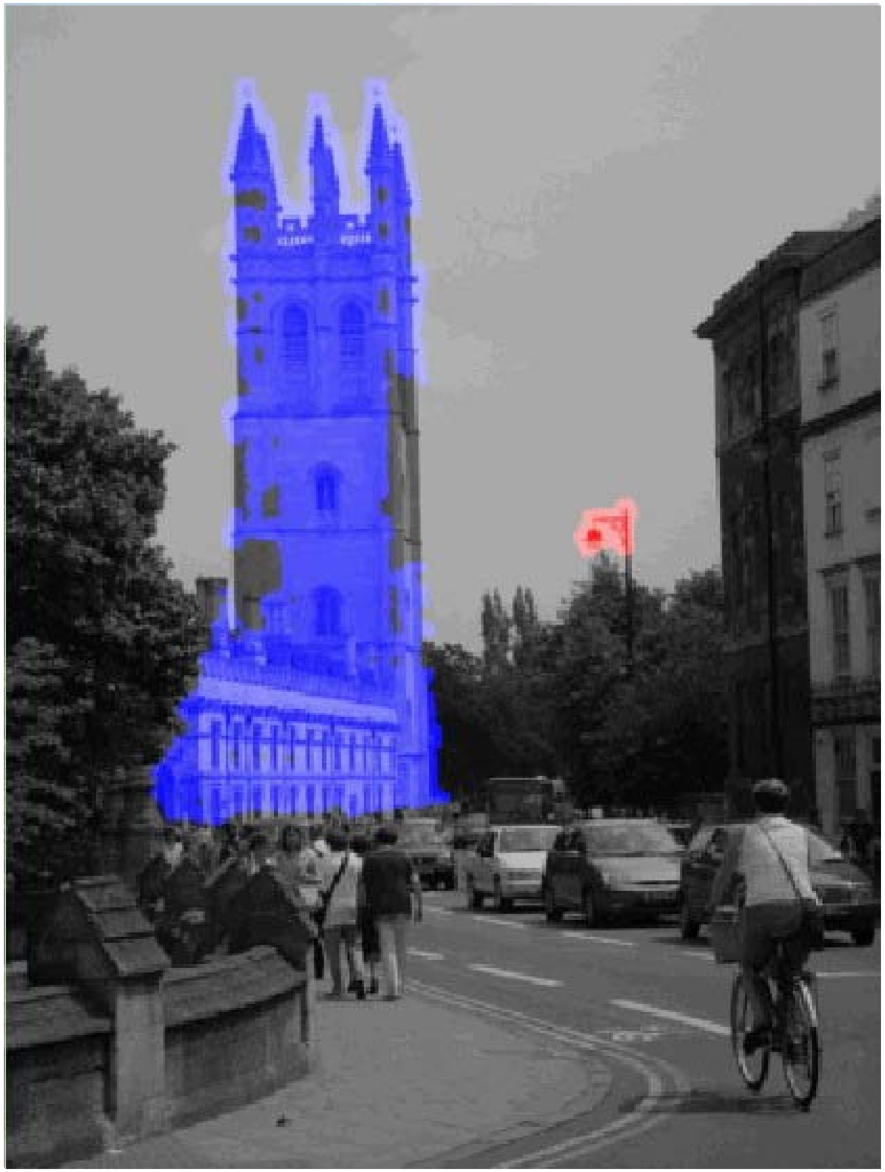


Using the whole bounding box

Result by query expansion using features inside the segmentation

Visual Content Hyperlinks





Conclusions

- Novel representation for hashing was introduced
  - Significantly improves the recall (reduces false negatives)
  - Reduces the number of false positives
  - Reduces memory footprint of image representation
  - Connects min-Hash and geometric Hashing
  - The first efficient combination of appearance and geometry in large scale indexing
- Applications
  - Clustering of spatially related images
  - Discovery of small objects

Thank you!

(cut outs)

Example of discovered object