Image Retrieval for Image-Based Localization Revisited

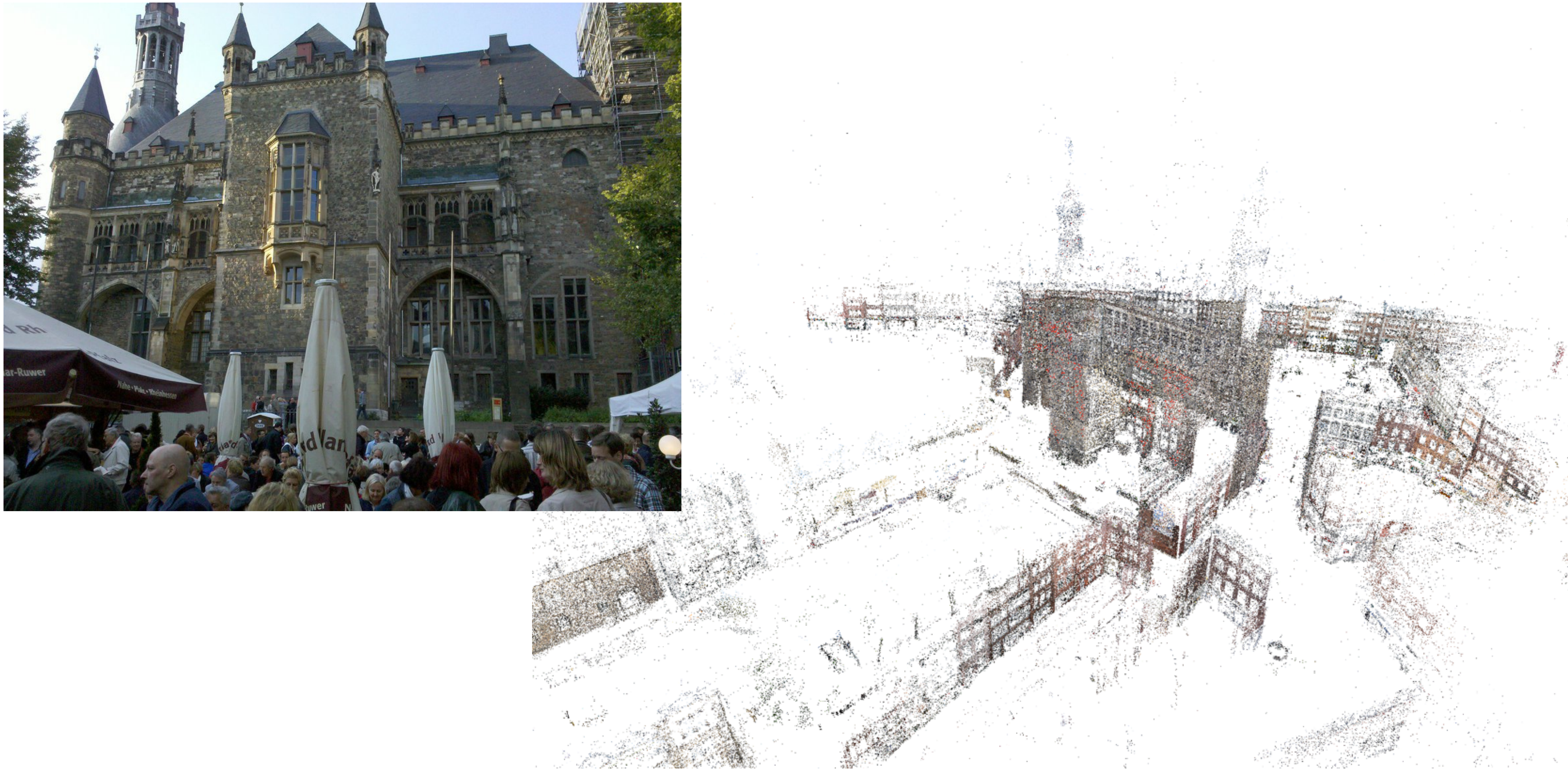
Torsten Sattler[1] Tobias Weyand[2]
Bastian Leibe[2] Leif Kobbelt[1]

[1]Computer Graphics Group, RWTH Aachen University
[2]Computer Vision Group, RWTH Aachen University

Image-Based Localization

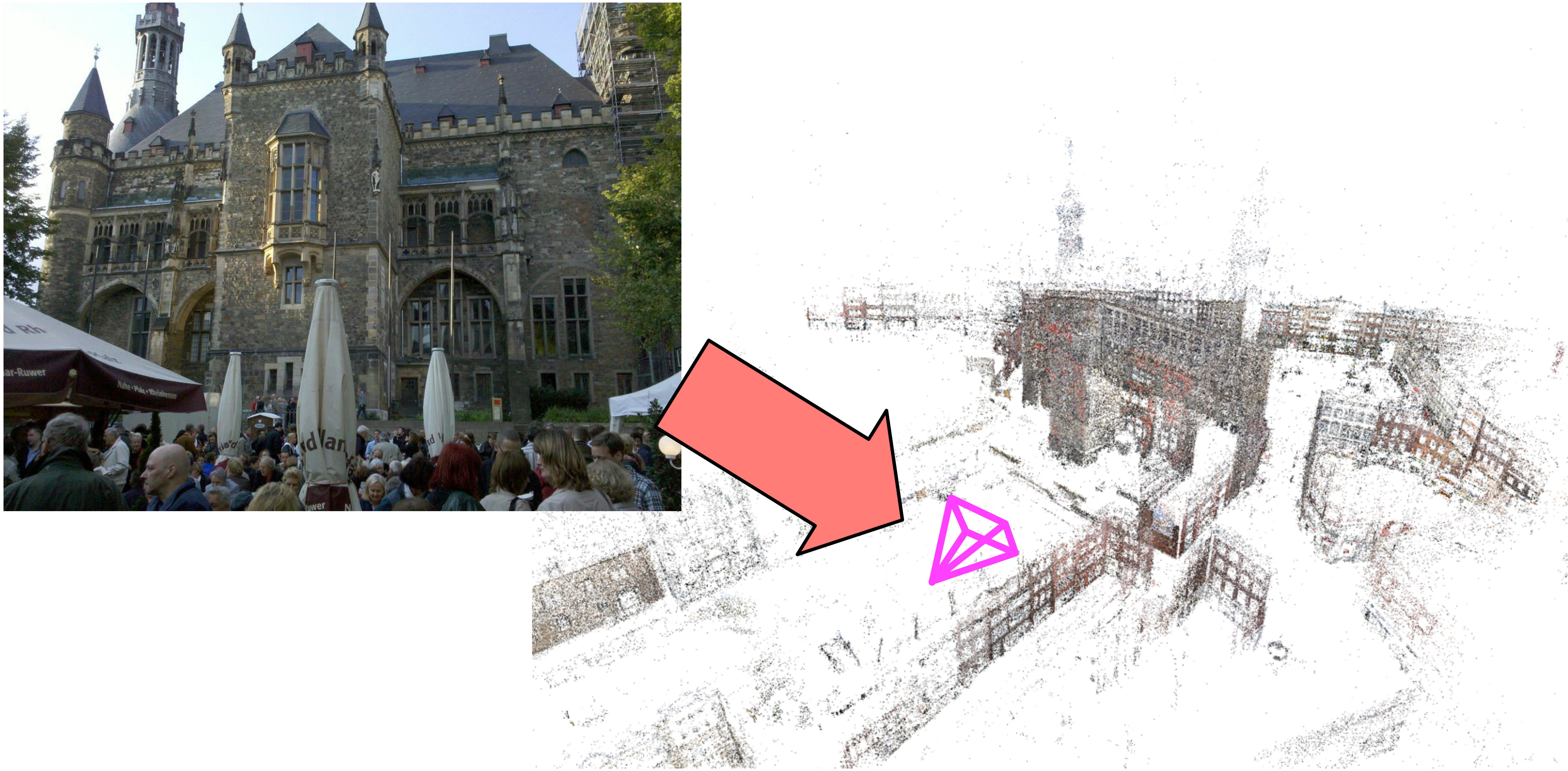
Determine **position & orientation** of query image

Image-Based Localization



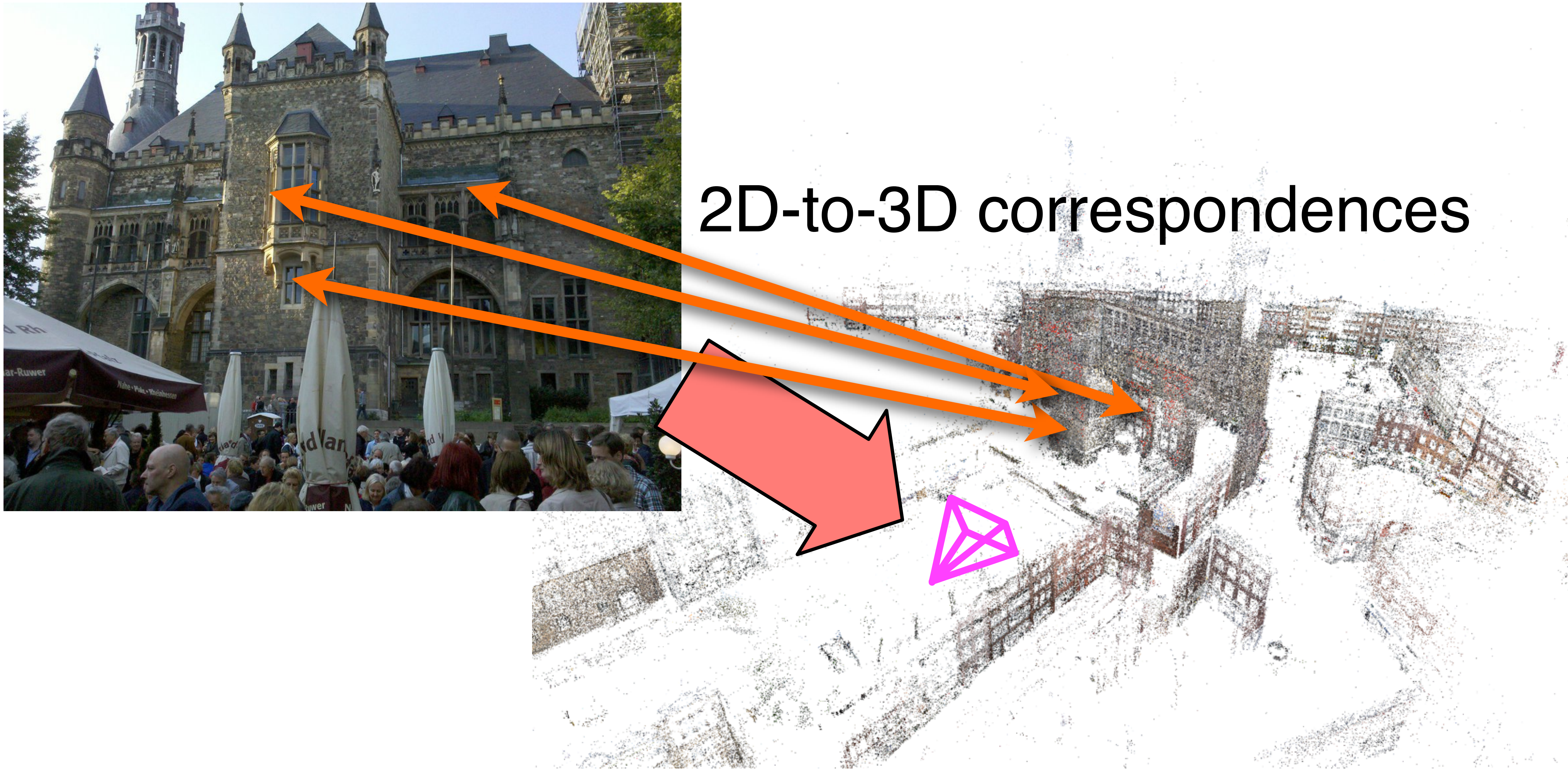
Determine **position & orientation** of query image

Image-Based Localization



Determine **position & orientation** of query image

Image-Based Localization



2D-to-3D correspondences

Determine **position & orientation** of query image

Image-Based Localization

- Structure-from-Motion point cloud
  - associate image descriptors with 3D points
  - ➡ **descriptor matching** problem

Image-Based Localization

- Structure-from-Motion point cloud
  - associate image descriptors with 3D points

➡ **descriptor matching** problem

	Scalability	Performance
Image retrieval	✓	✗

Image-Based Localization

- Structure-from-Motion point cloud
  - associate image descriptors with 3D points
  - ➡ **descriptor matching** problem

	Scalability	**Performance**
Image retrieval	✓	✗
Direct matching	✗	✓

Image-Based Localization

- Structure-from-Motion point cloud
  - associate image descriptors with 3D points
  - ➡ **descriptor matching** problem

	Scalability	Performance
Image retrieval	✓	✗
Direct matching	✗	✓

Overview

- Image Retrieval & Direct Matching
- Image Retrieval Revisited
- Efficient Correspondence Selection

Image Retrieval for Localization

Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09

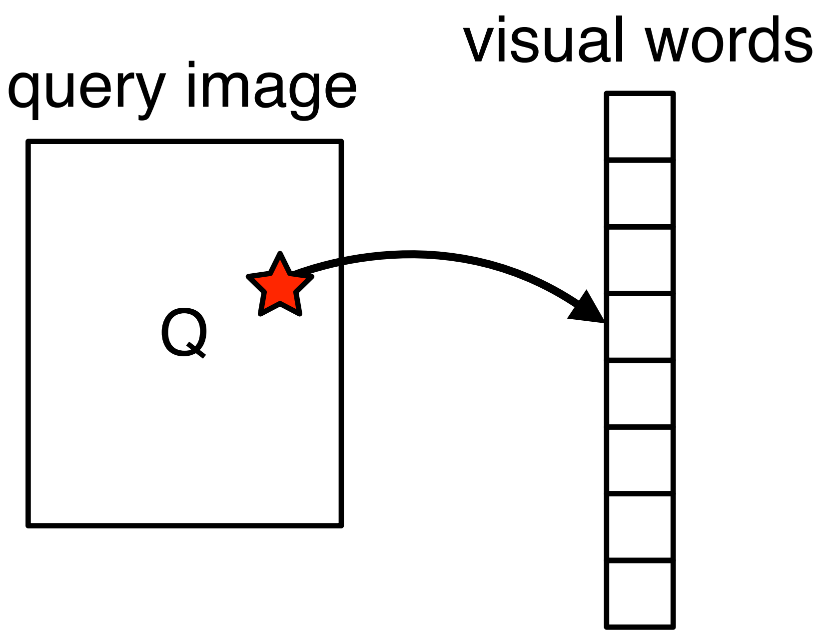
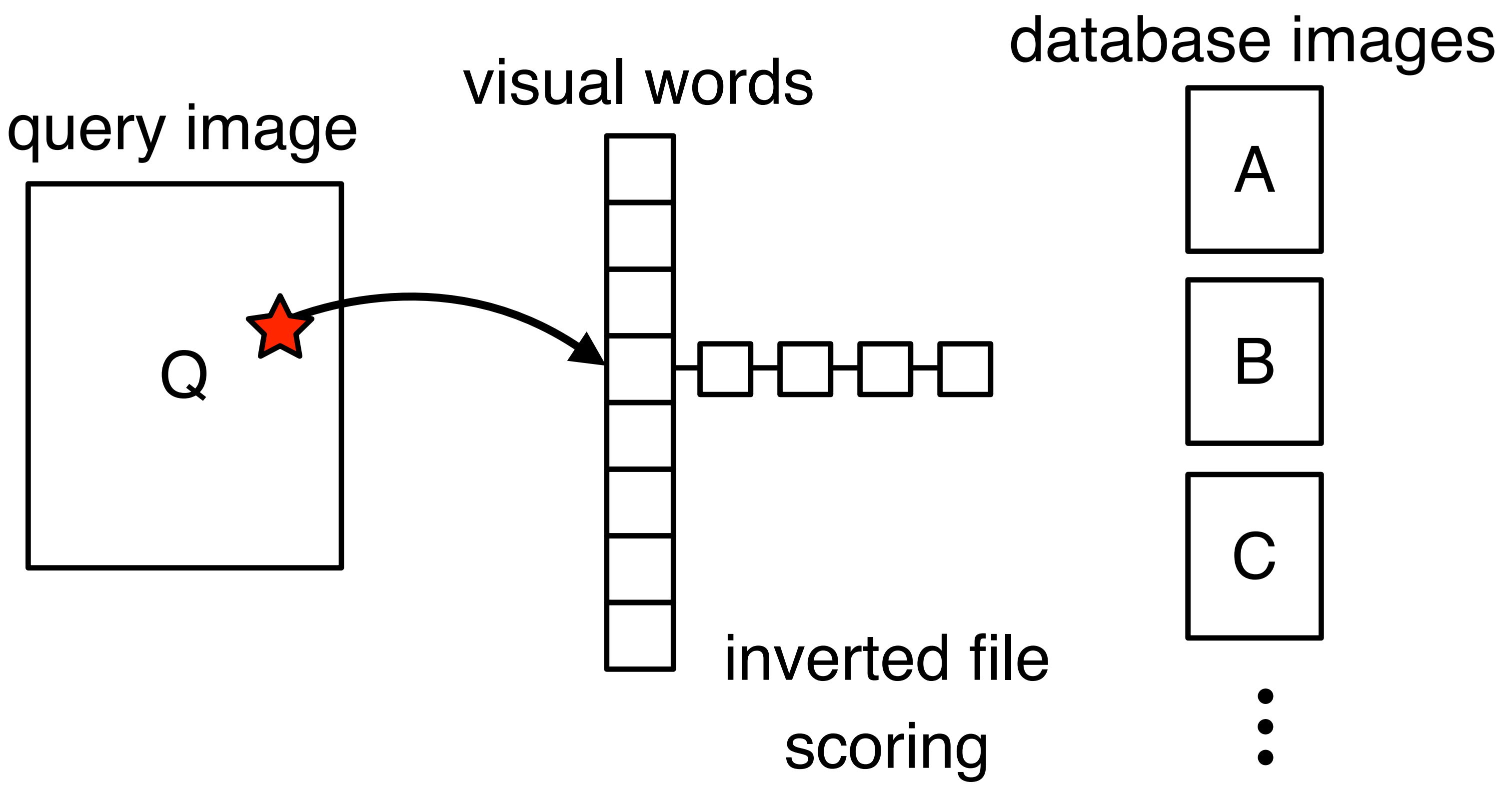


Image Retrieval for Localization

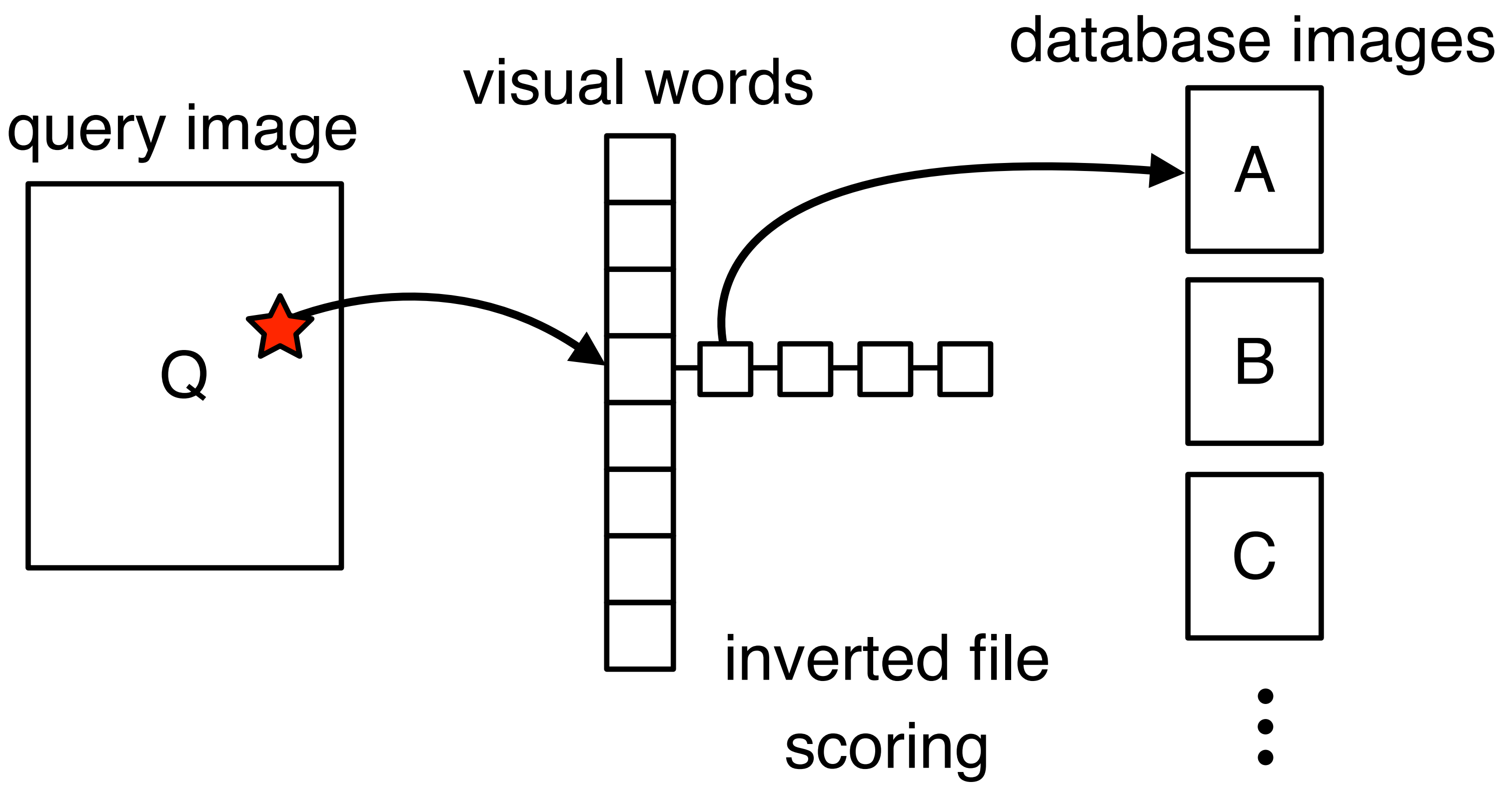
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

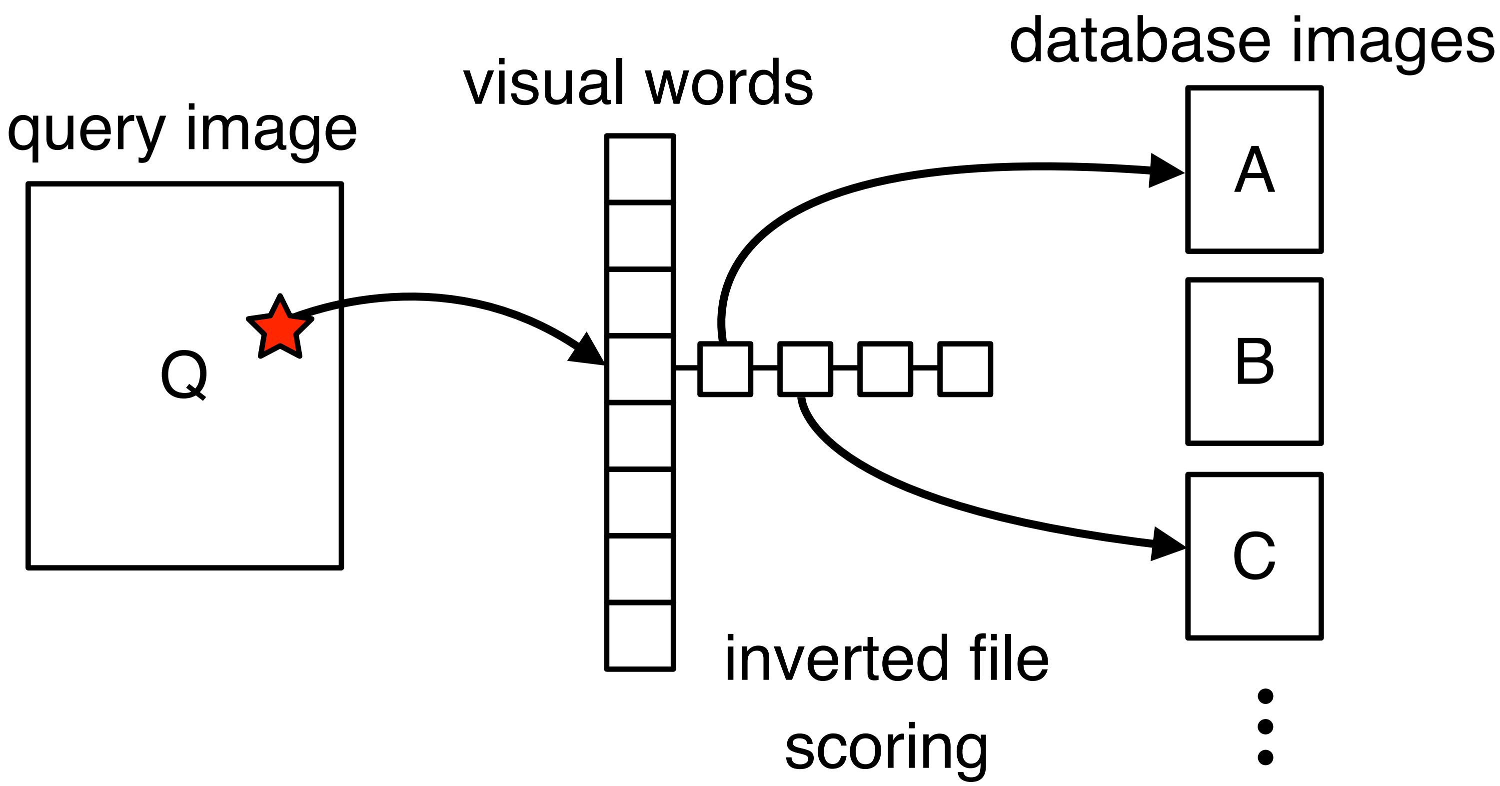
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

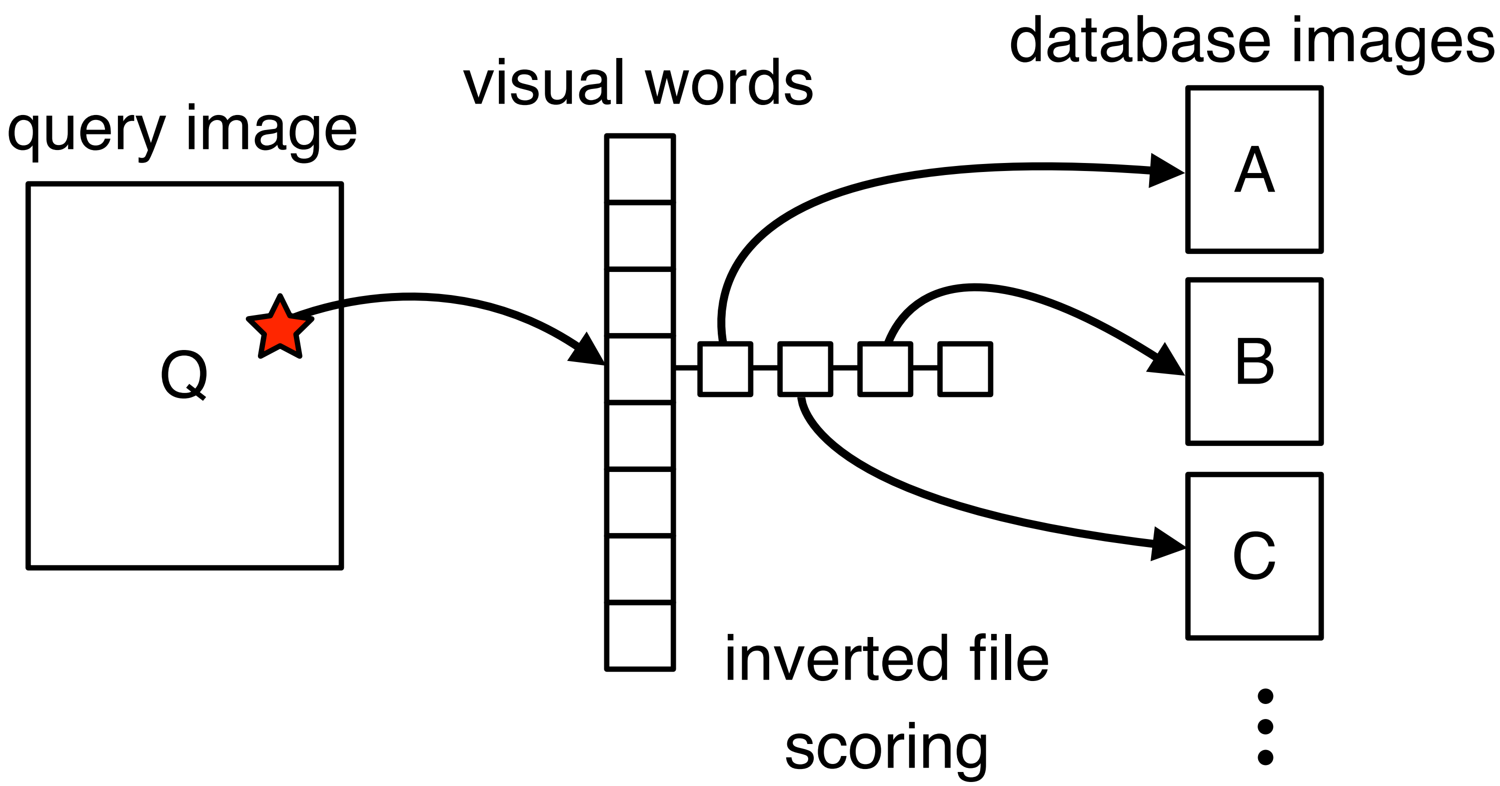
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

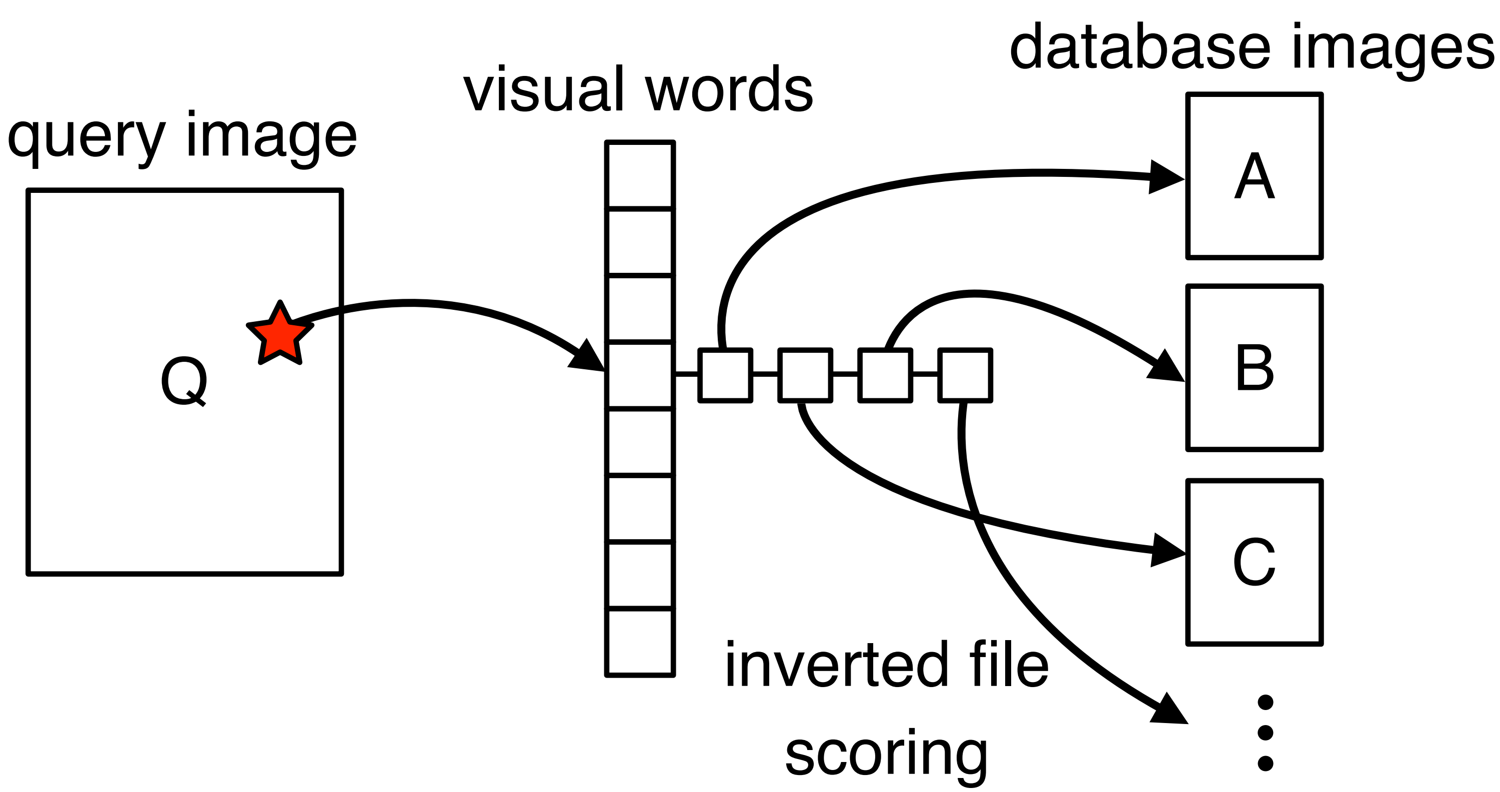
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR’09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

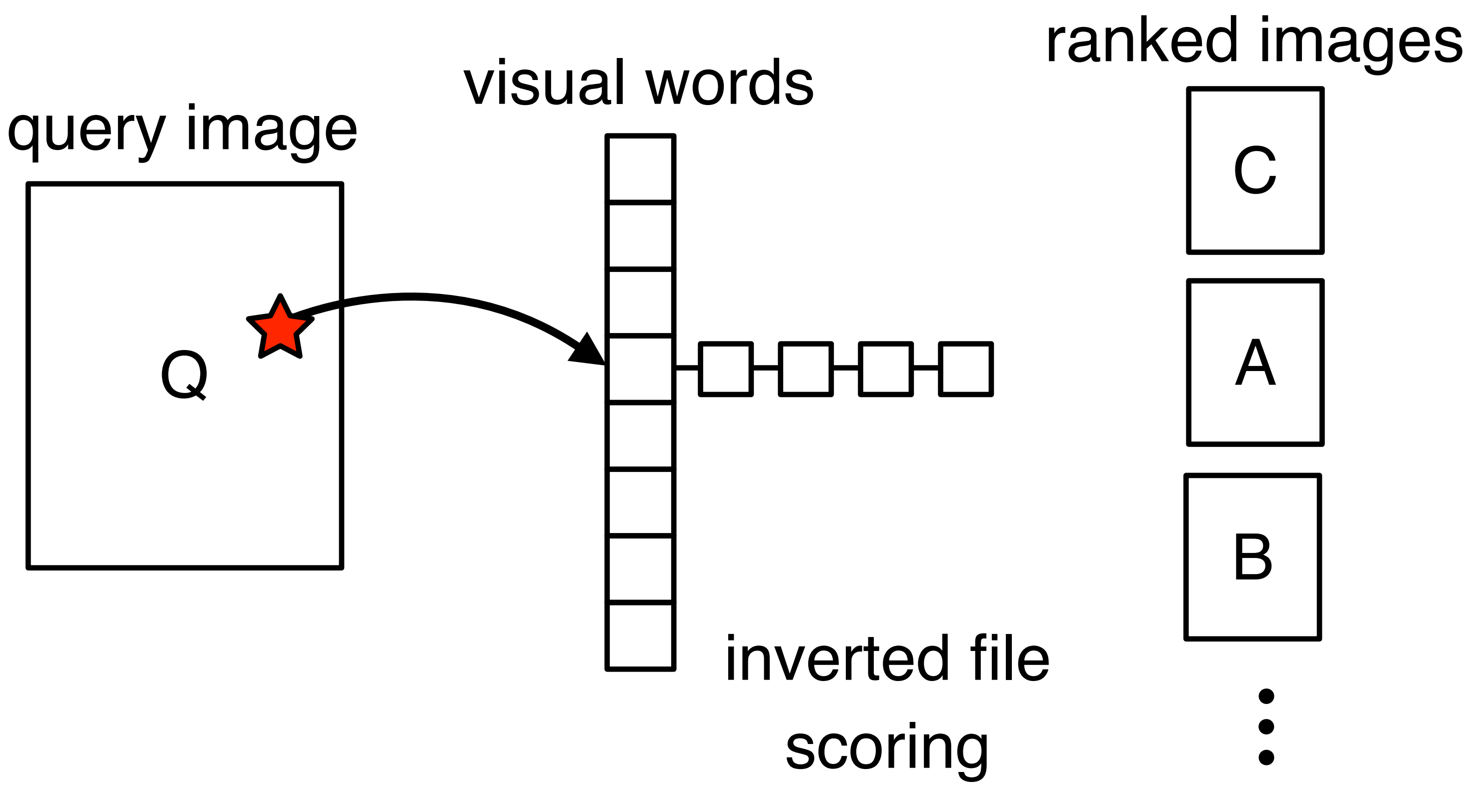
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

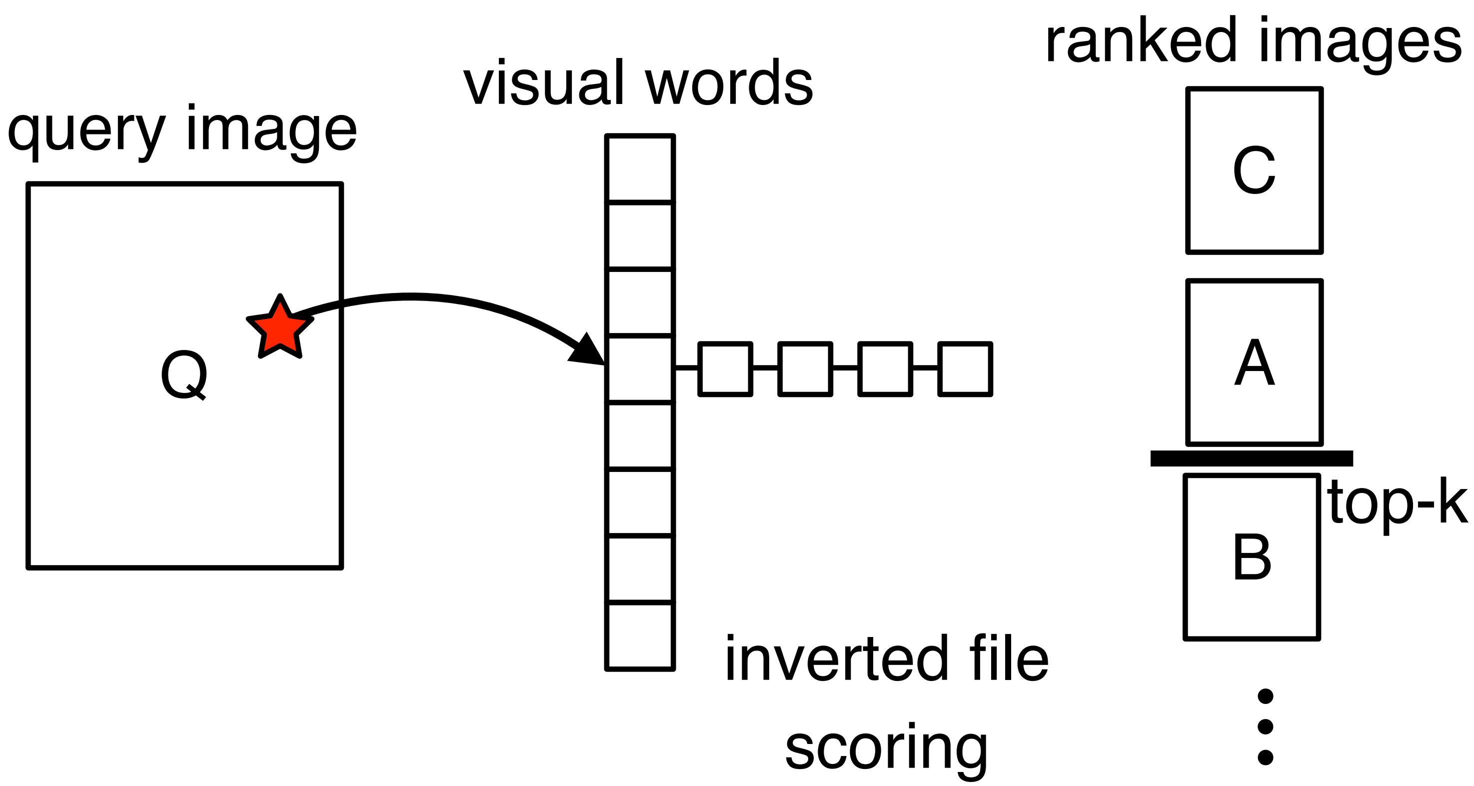
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

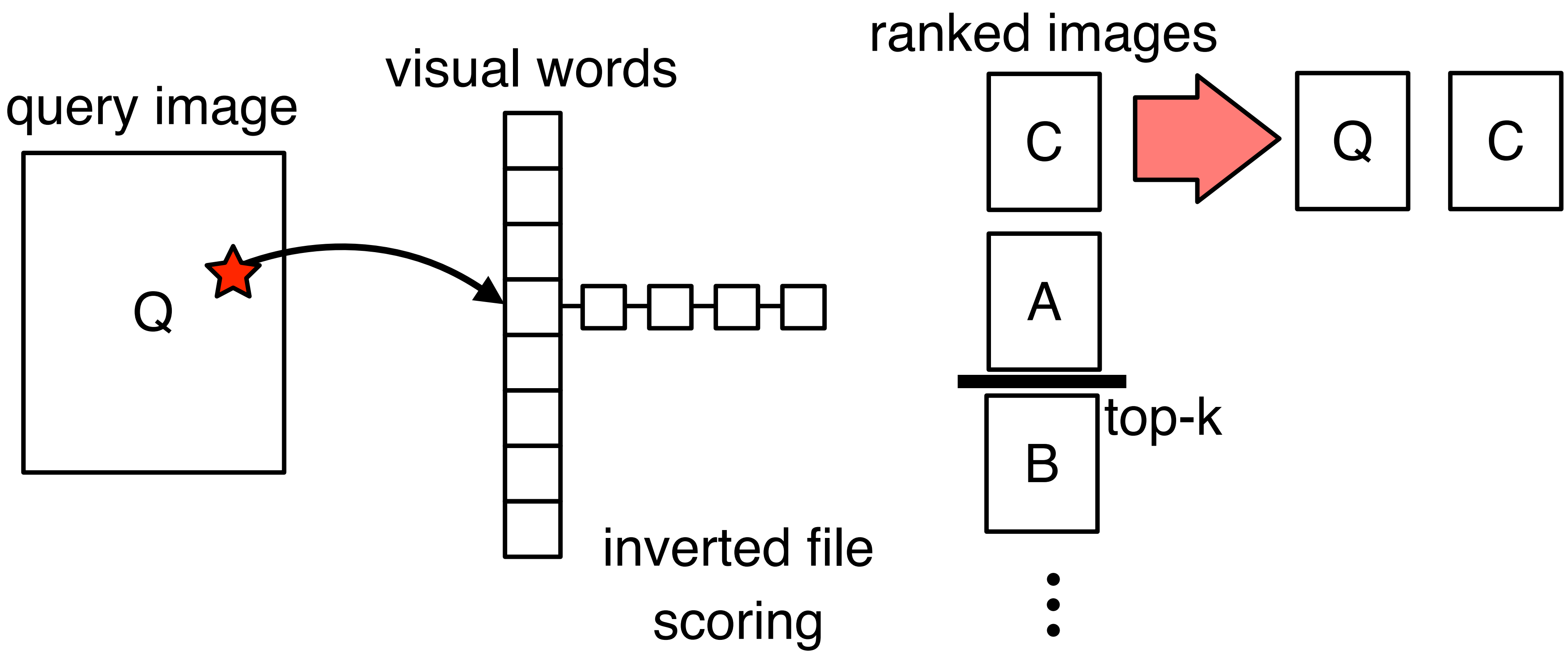
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

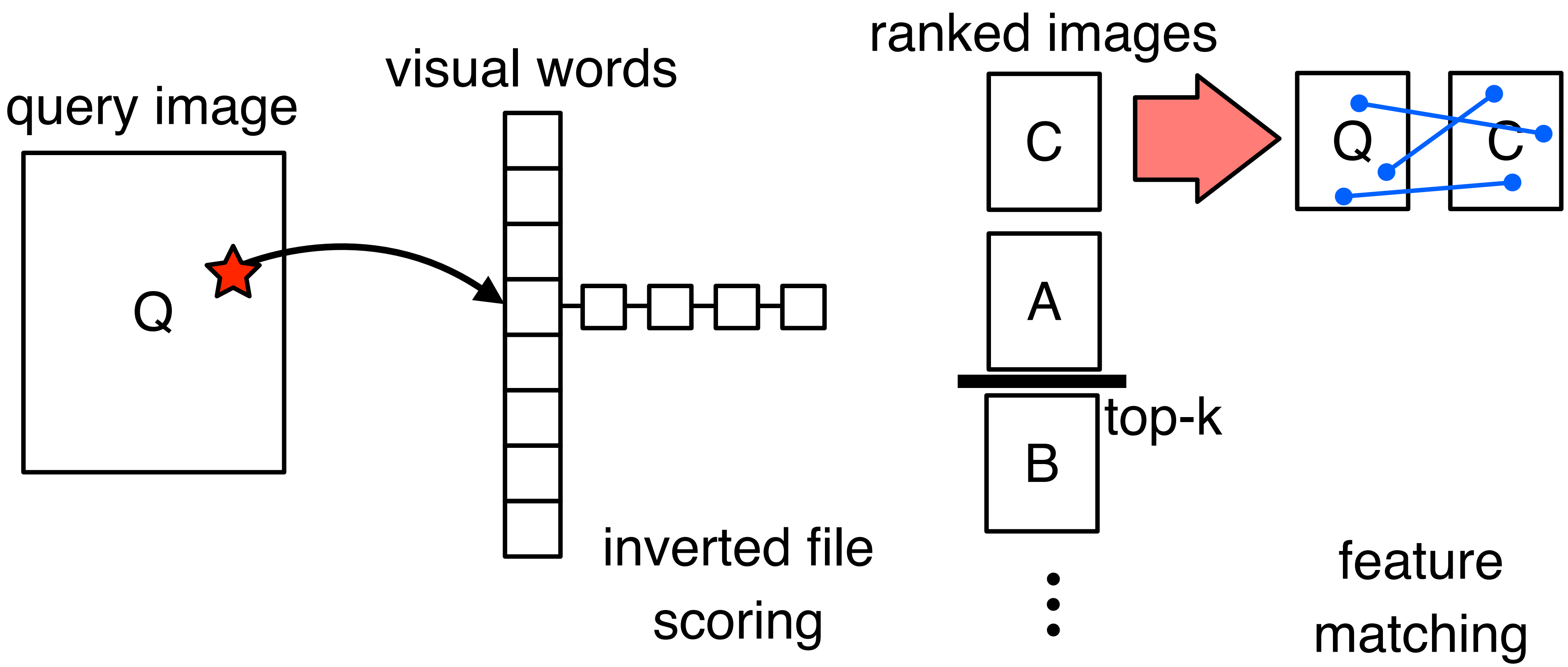
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

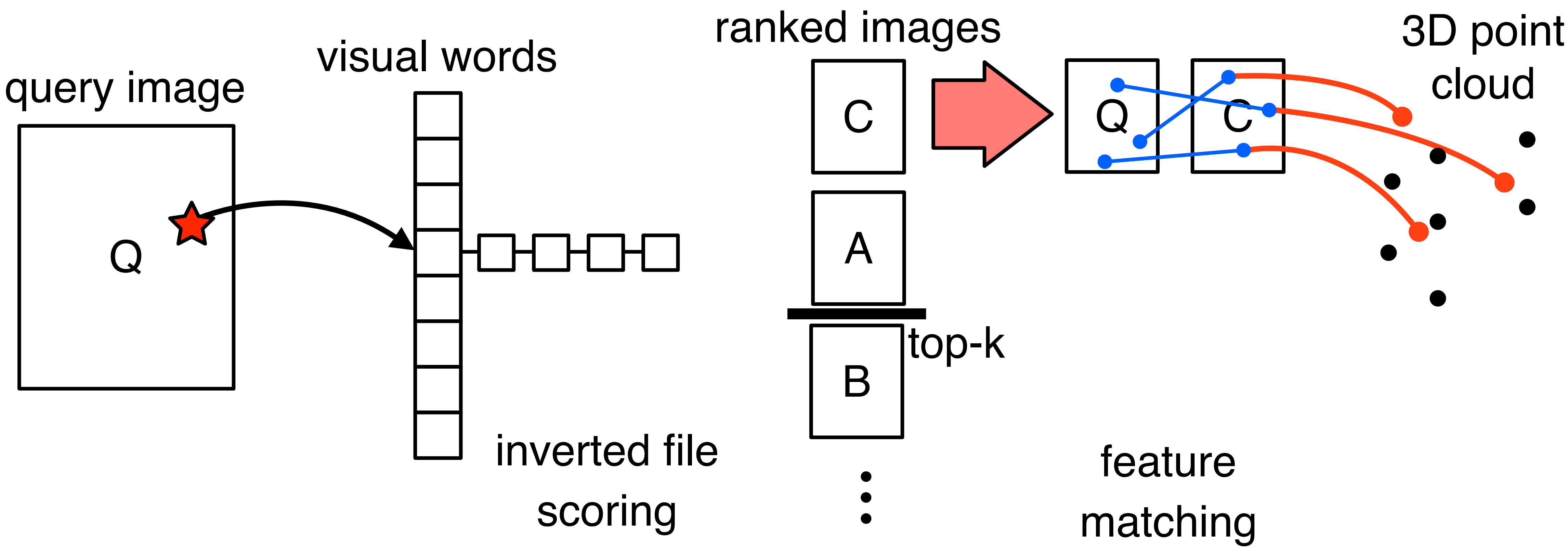
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

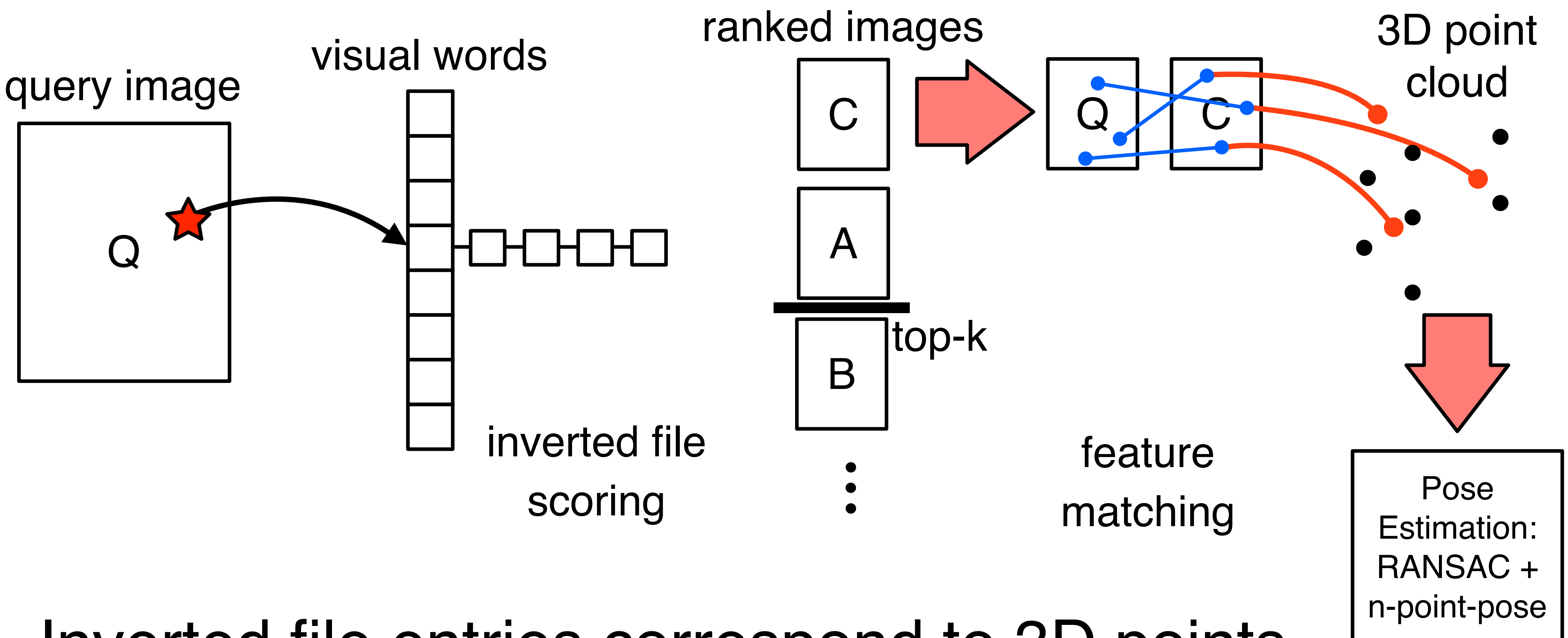
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

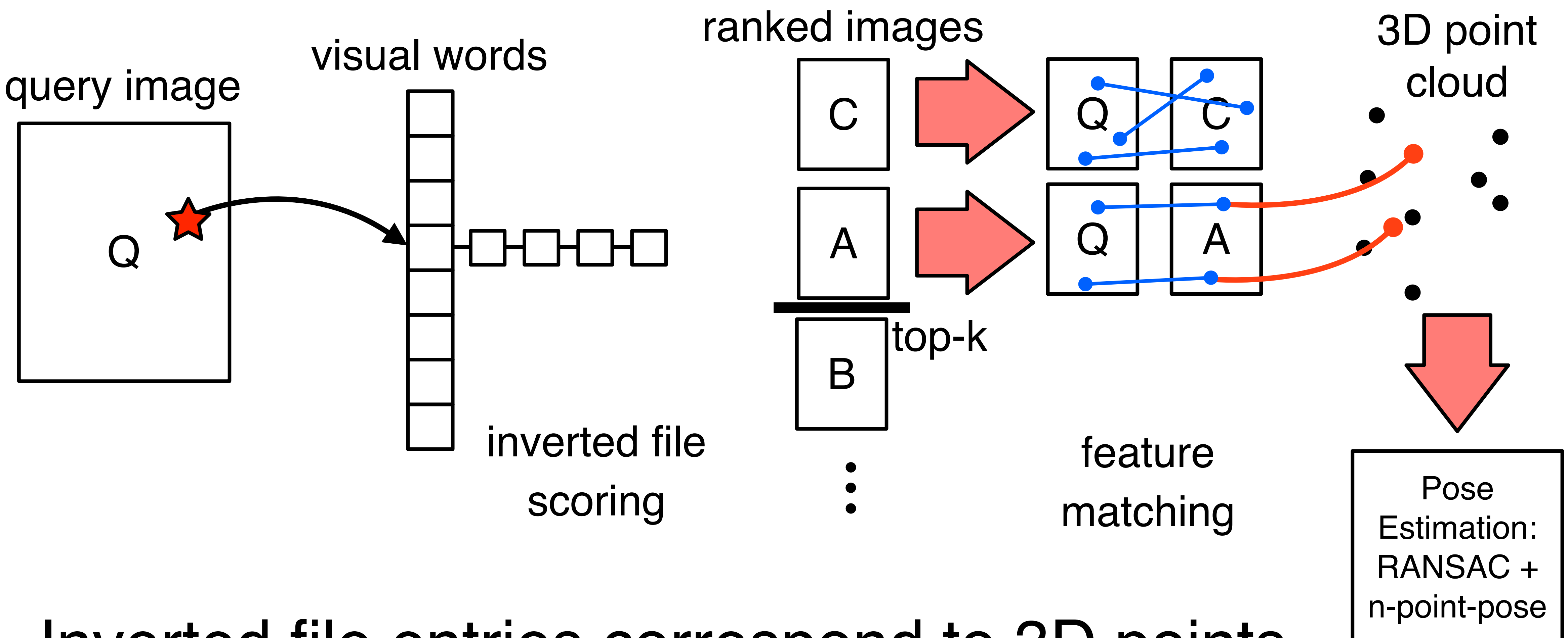
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

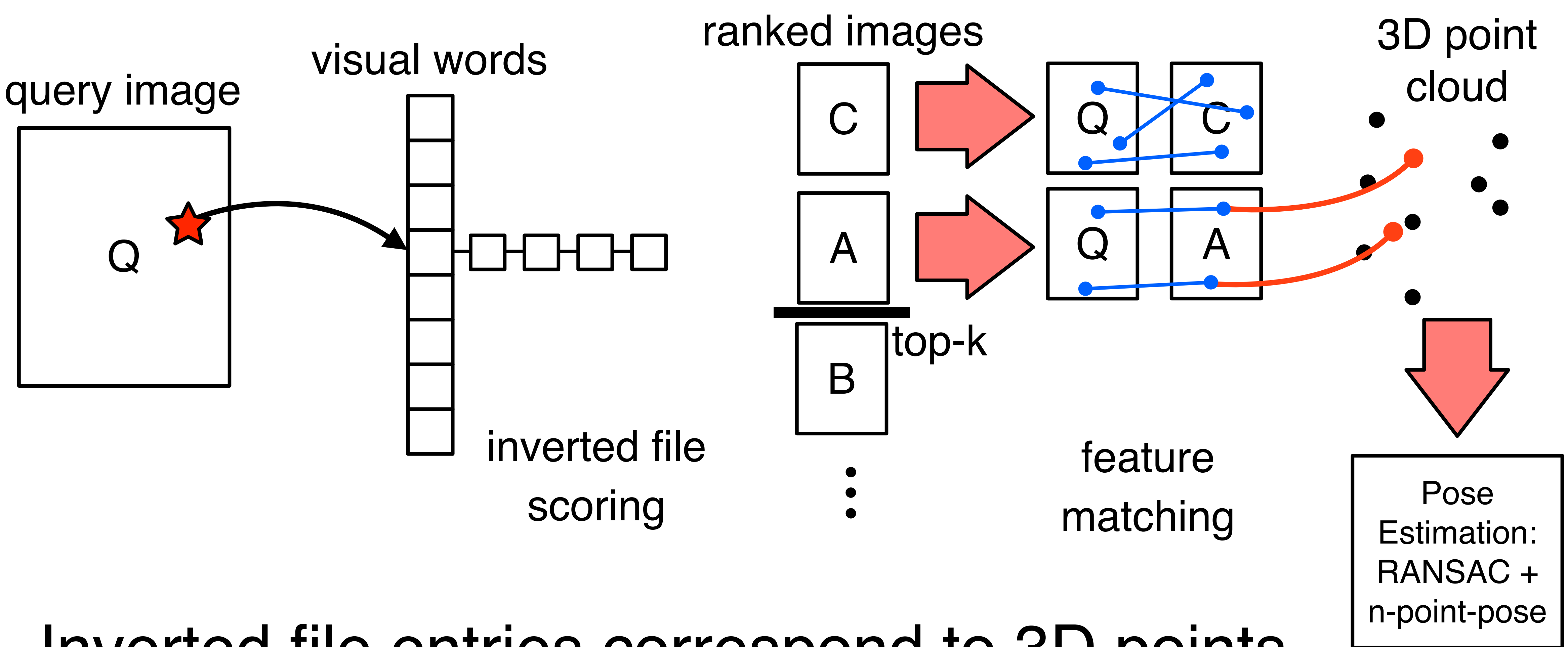
Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09



Inverted file entries correspond to 3D points

Image Retrieval for Localization

Irschara, Zach, Frahm, Bischof. *From Structure-from-Motion Point Clouds to Fast Location Recognition*. CVPR'09

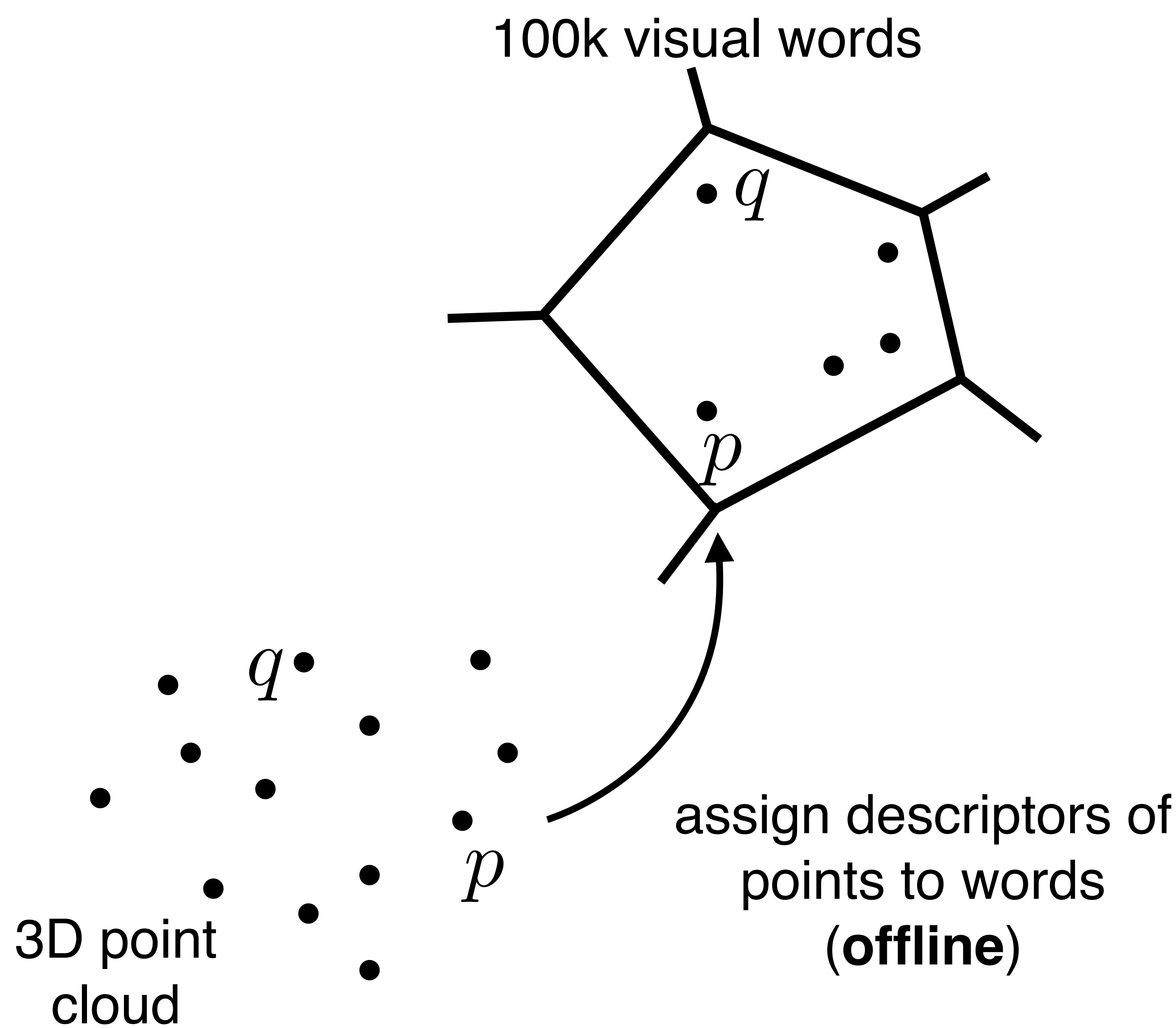


Inverted file entries correspond to 3D points

Choose pose with most inliers as final pose

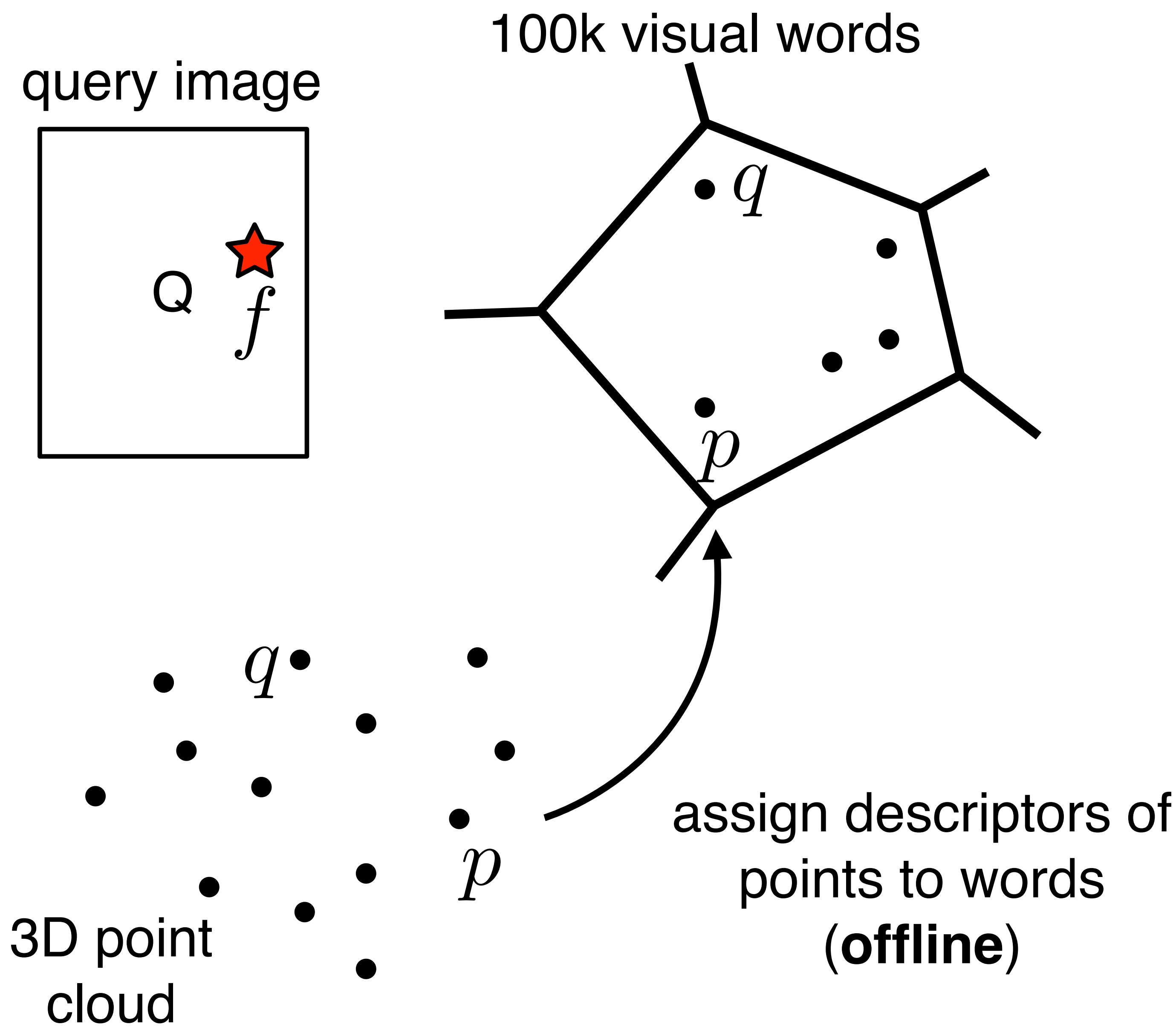
Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11



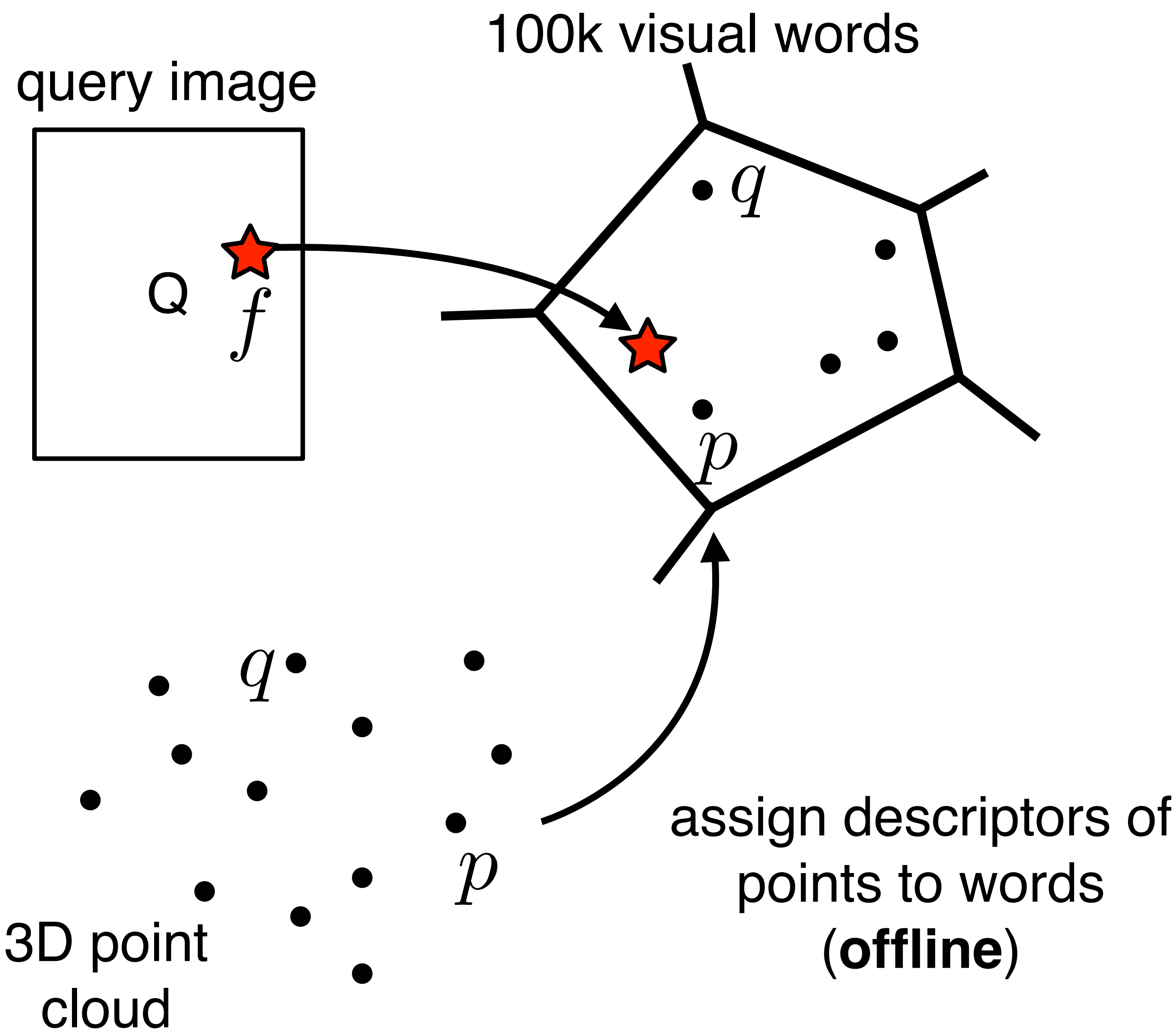
Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11



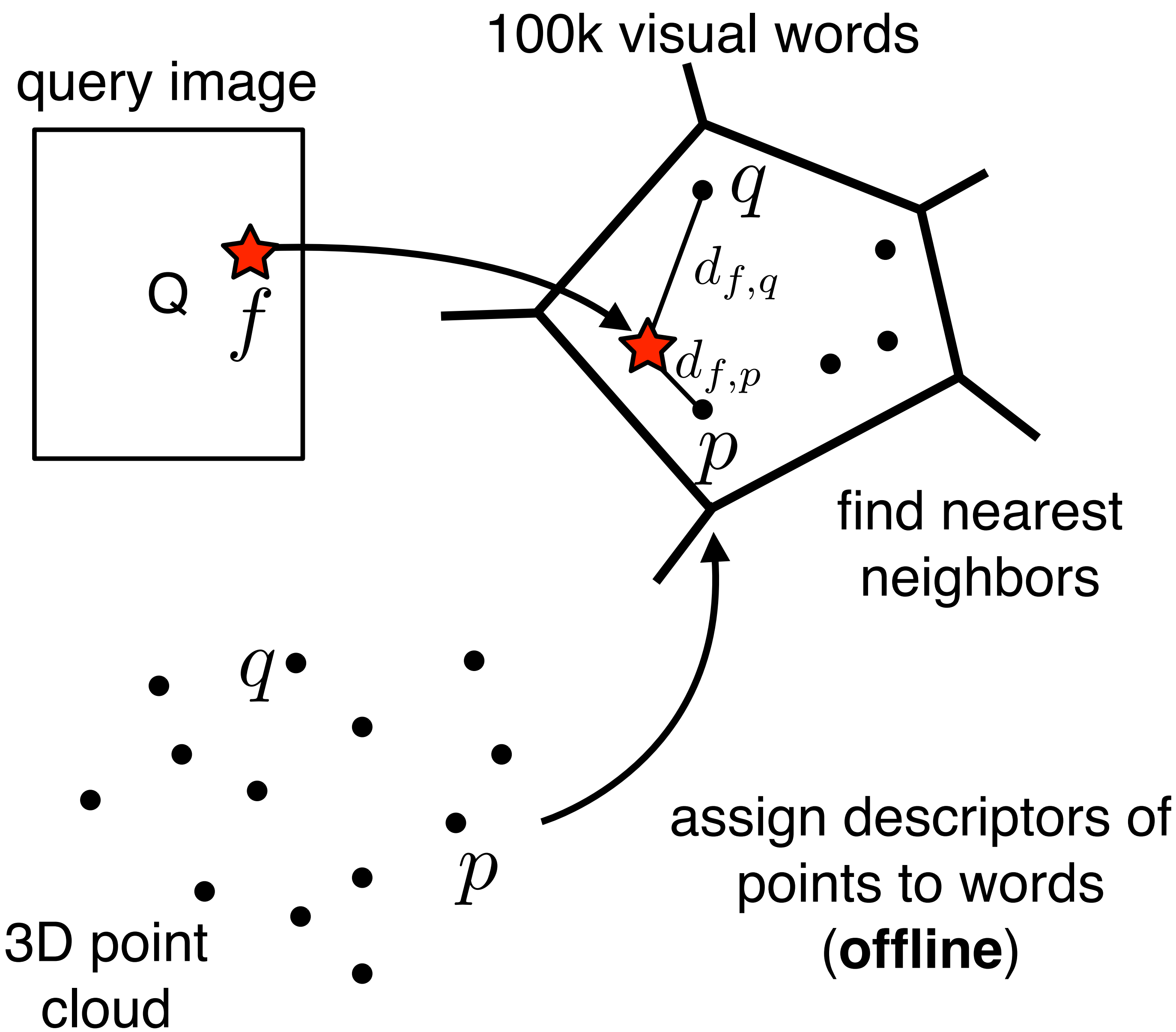
Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11



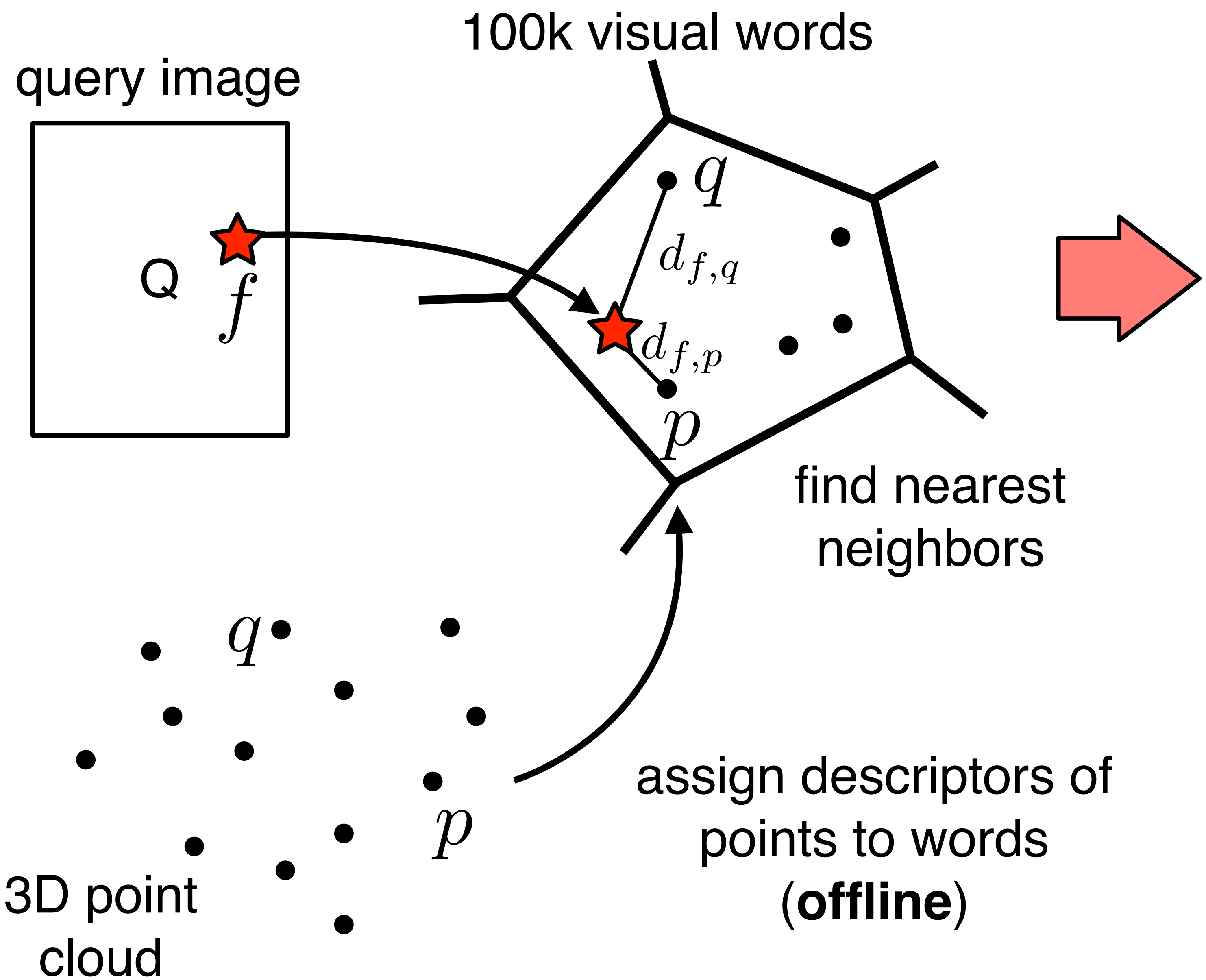
Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11



Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11

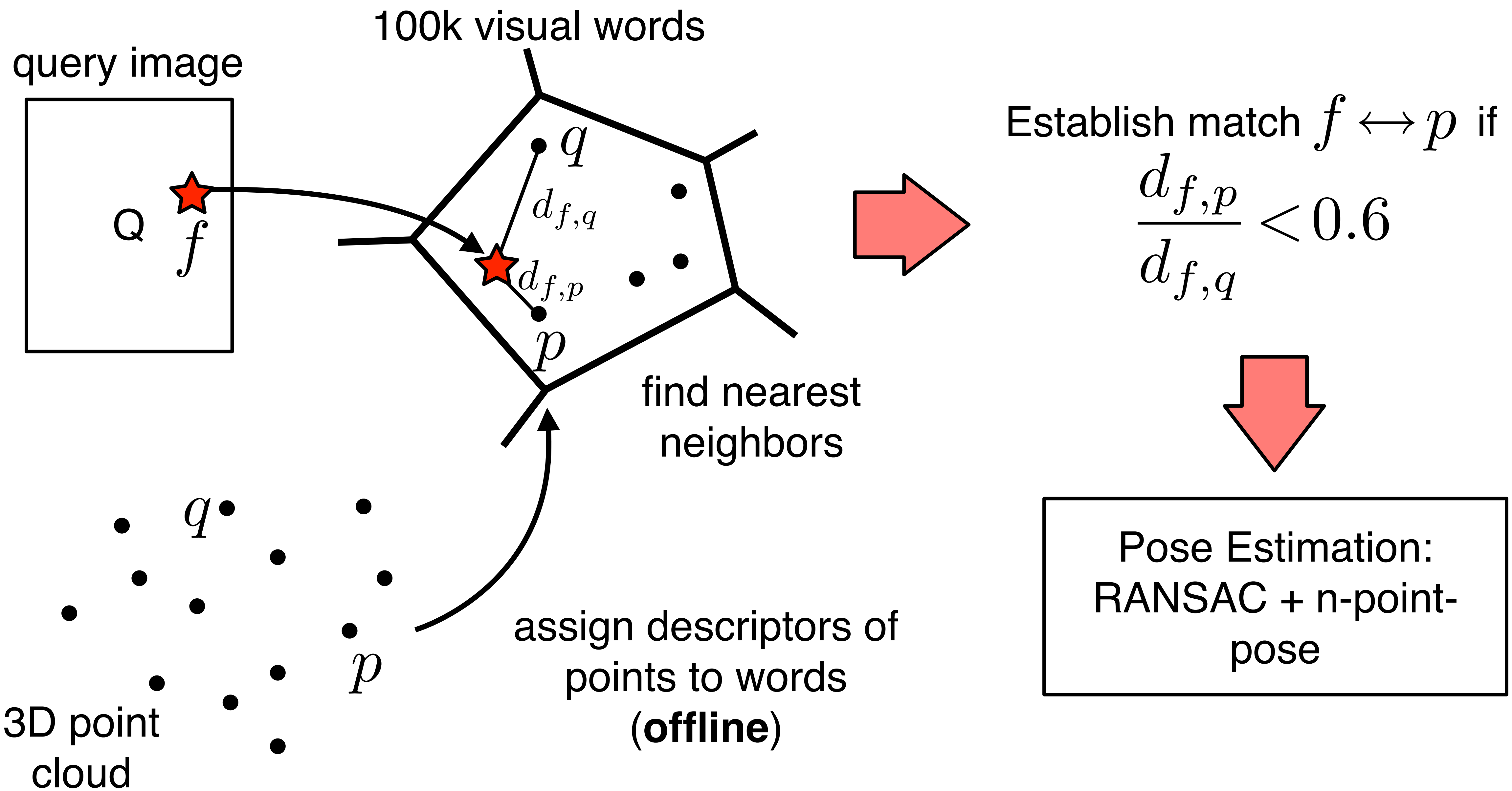


Establish match $f \leftrightarrow p$ if

$$\frac{d_{f,p}}{d_{f,q}} < 0.6$$

Direct Matching

Sattler, Leibe, Kobbelt. *Fast Image-Based Localization using Direct 2D-to-3D Matching.* ICCV'11



The Performance Gap

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art

The Performance Gap

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art

The Performance Gap

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art

The Performance Gap

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art

Performance gap caused by **failure to rank any relevant image high enough**

Image Retrieval Revisited

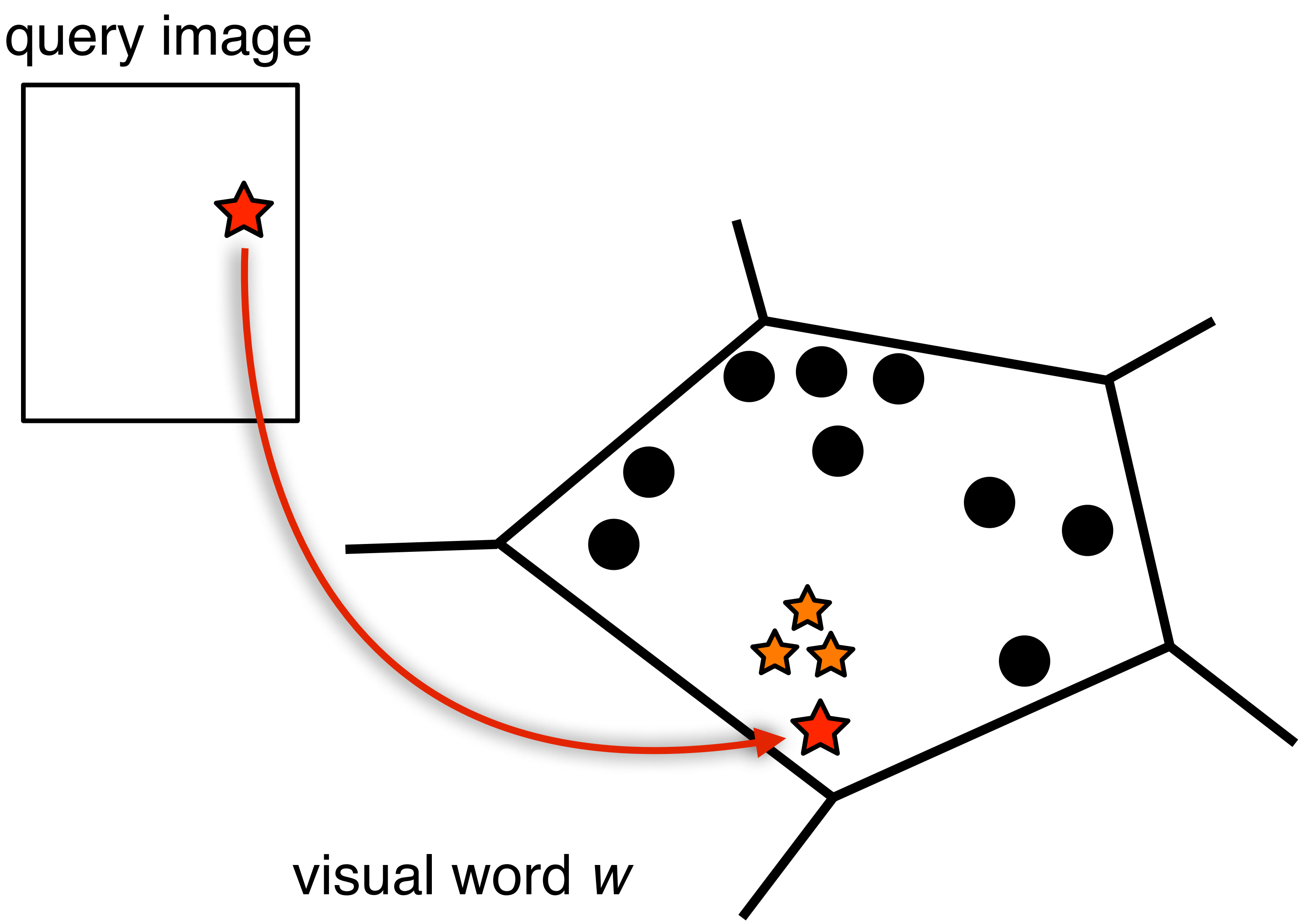


Image Retrieval Revisited

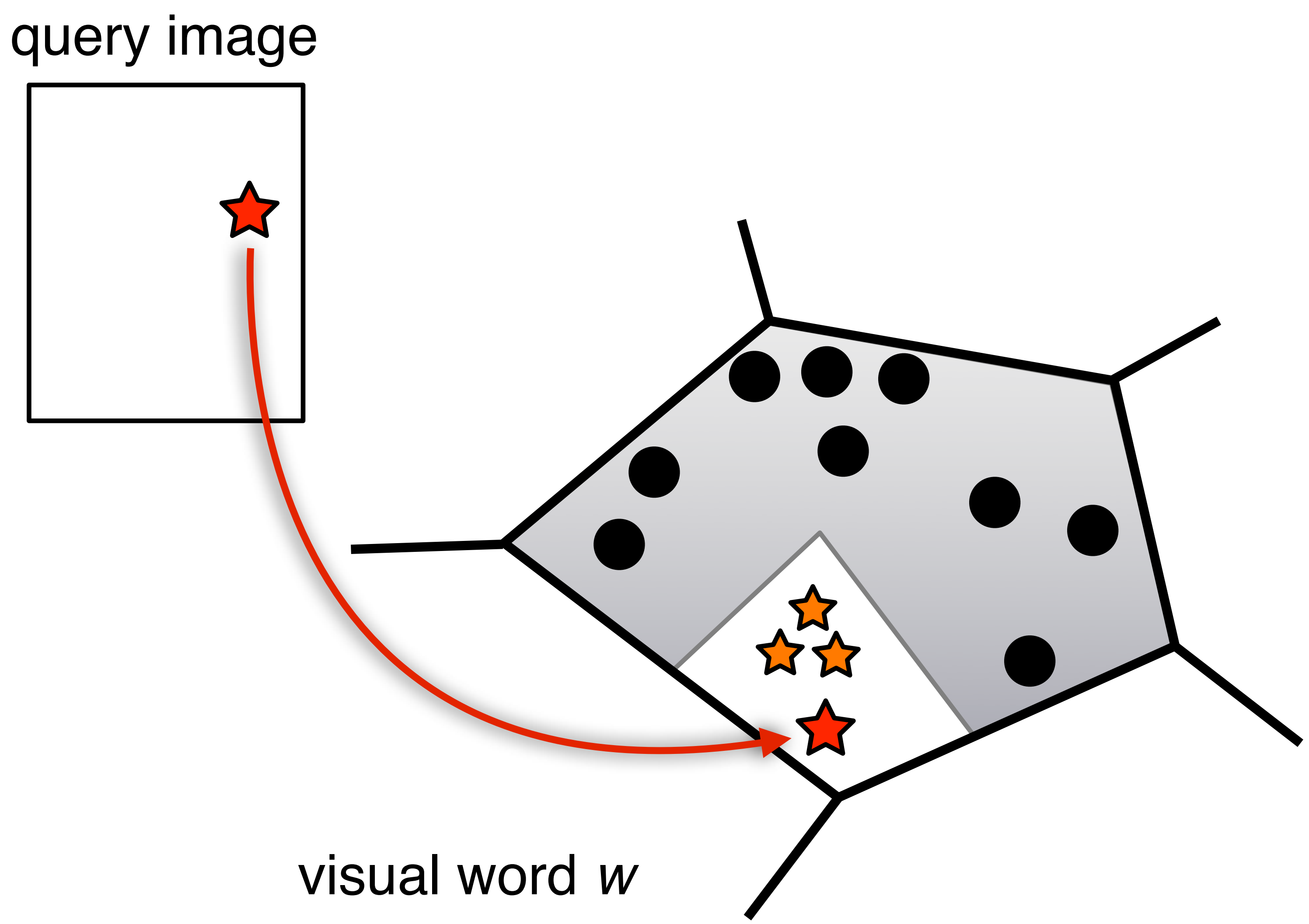


Image Retrieval Revisited

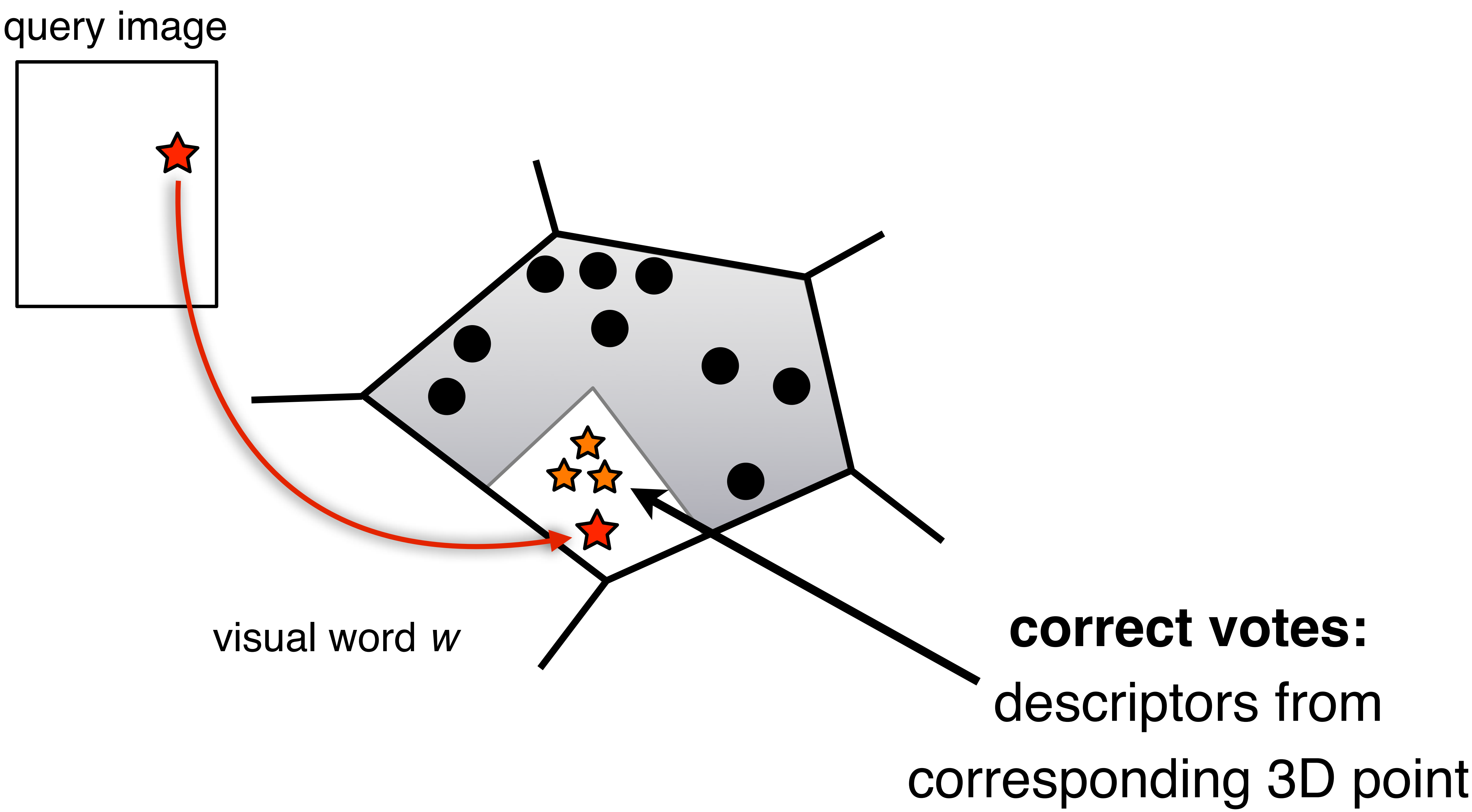


Image Retrieval Revisited

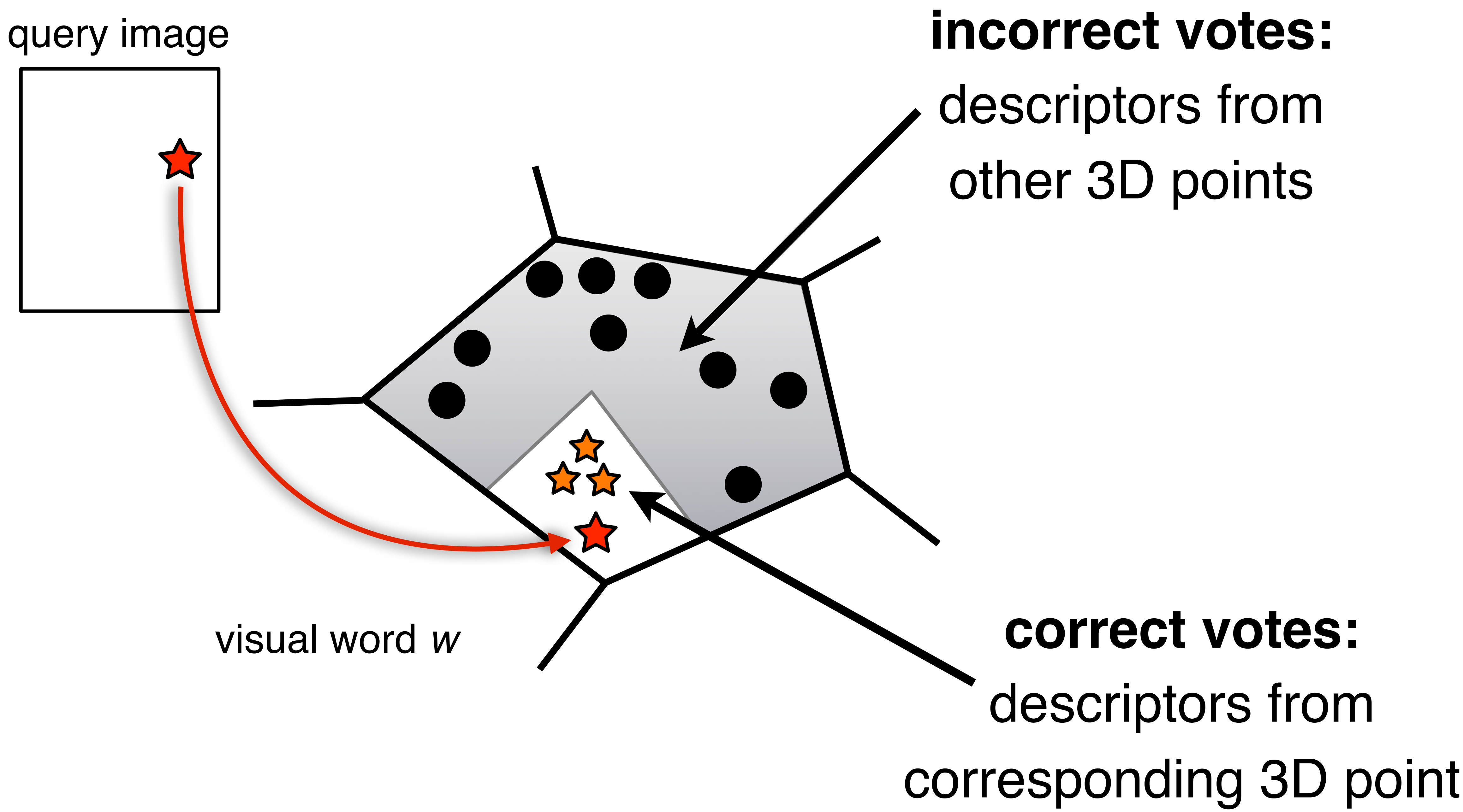
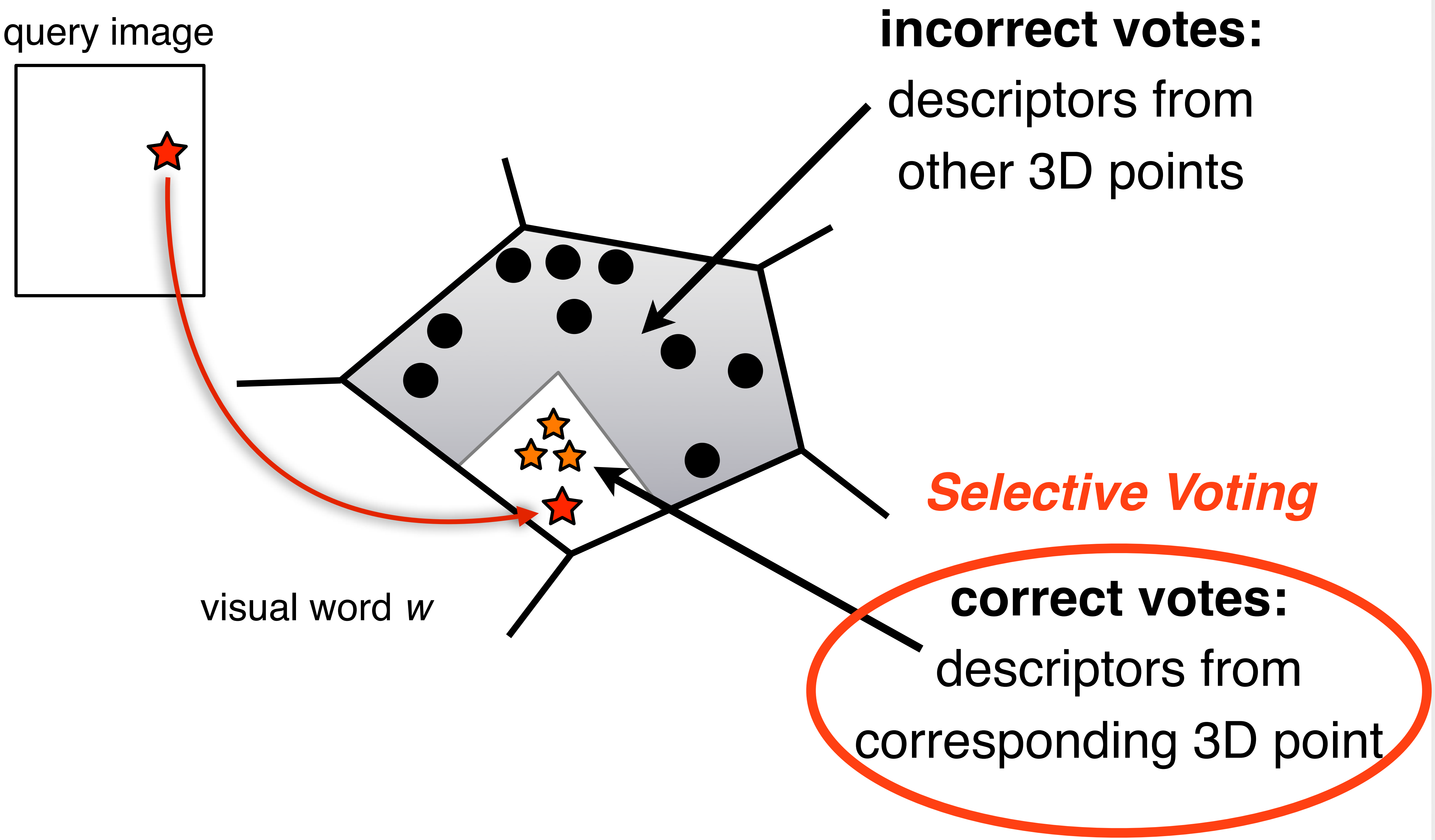
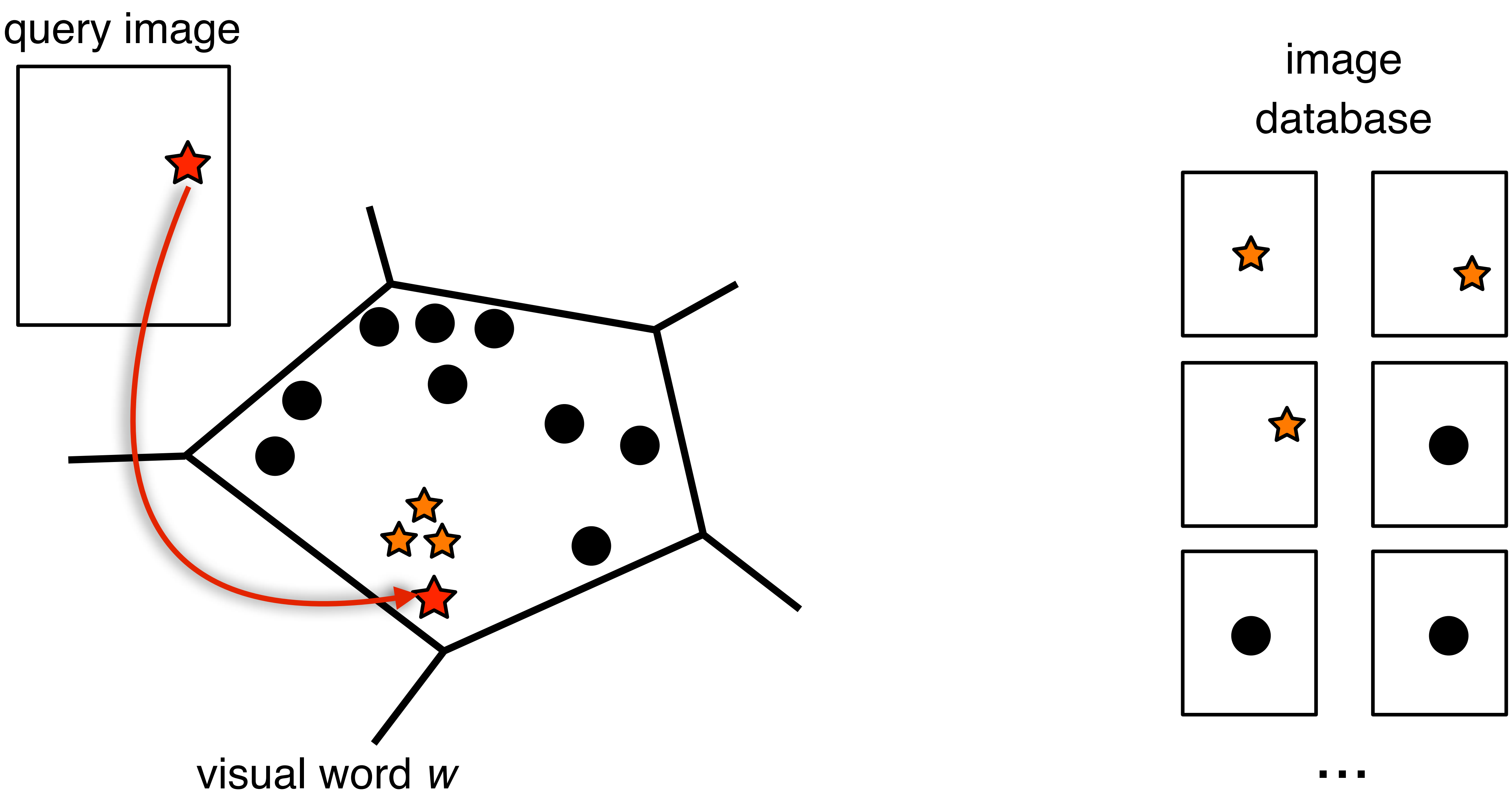


Image Retrieval Revisited



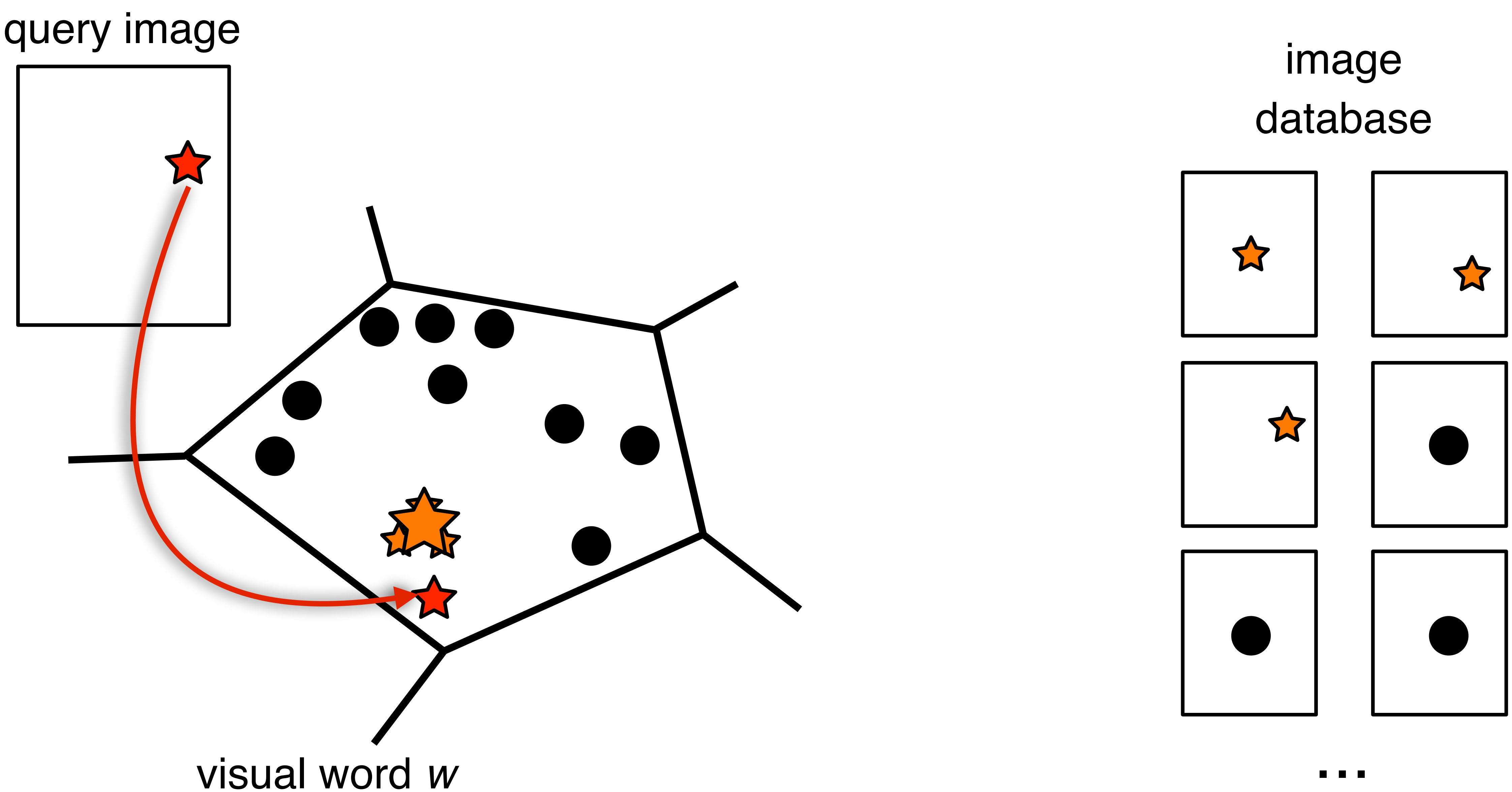
Correspondence Voting

Idea: Find corresponding 3D point



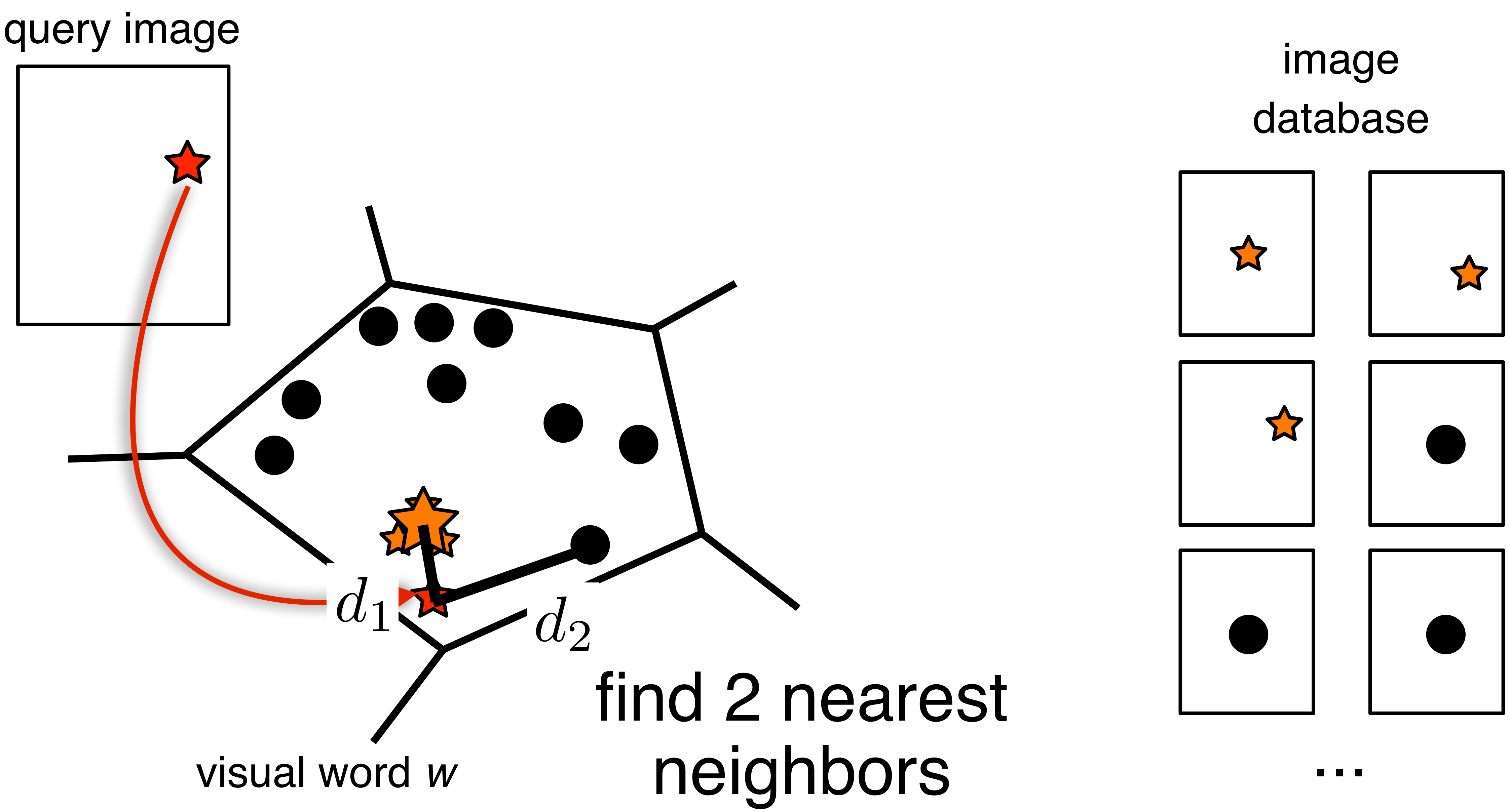
Correspondence Voting

Idea: Find corresponding 3D point



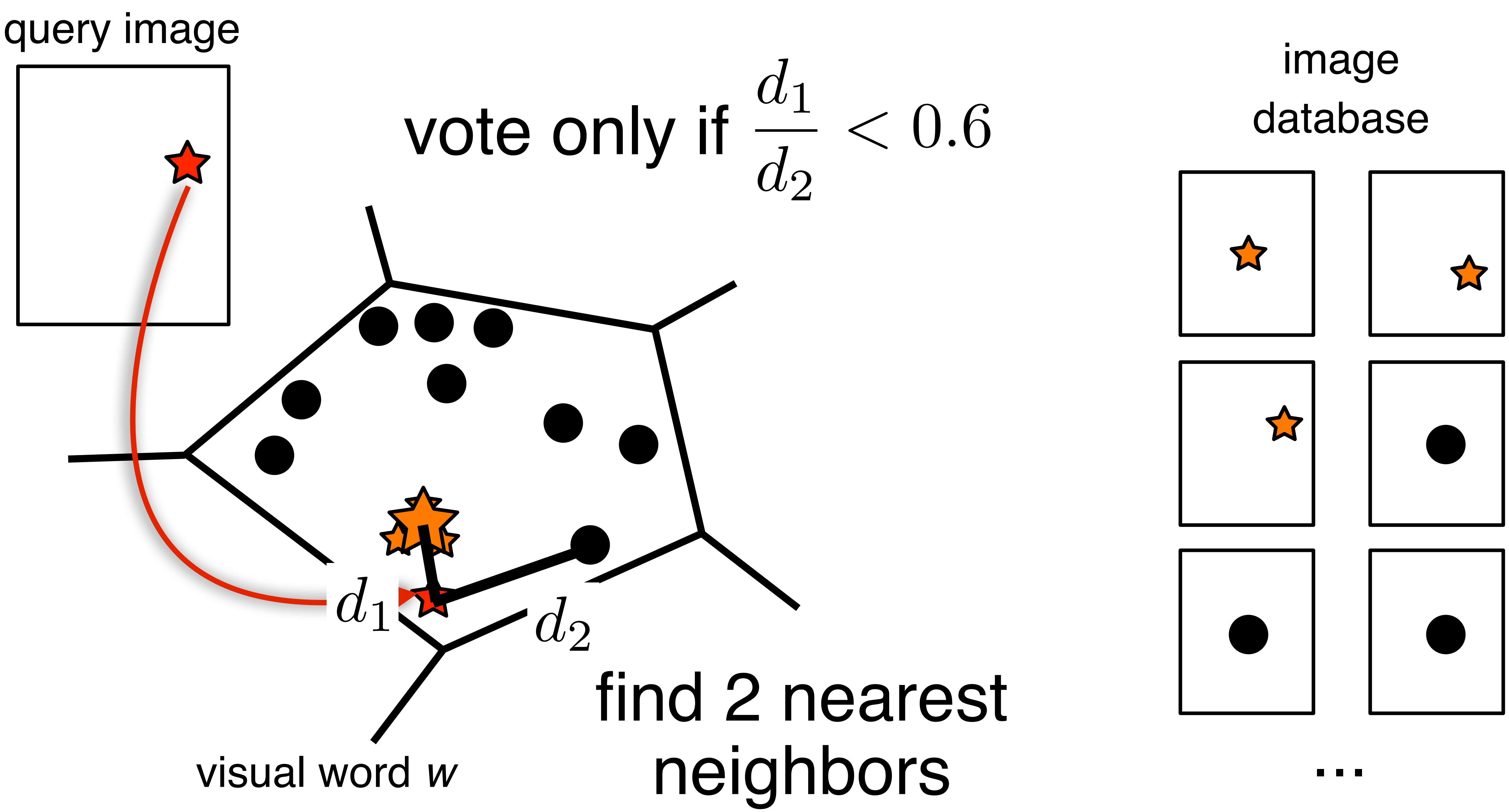
Correspondence Voting

Idea: Find corresponding 3D point



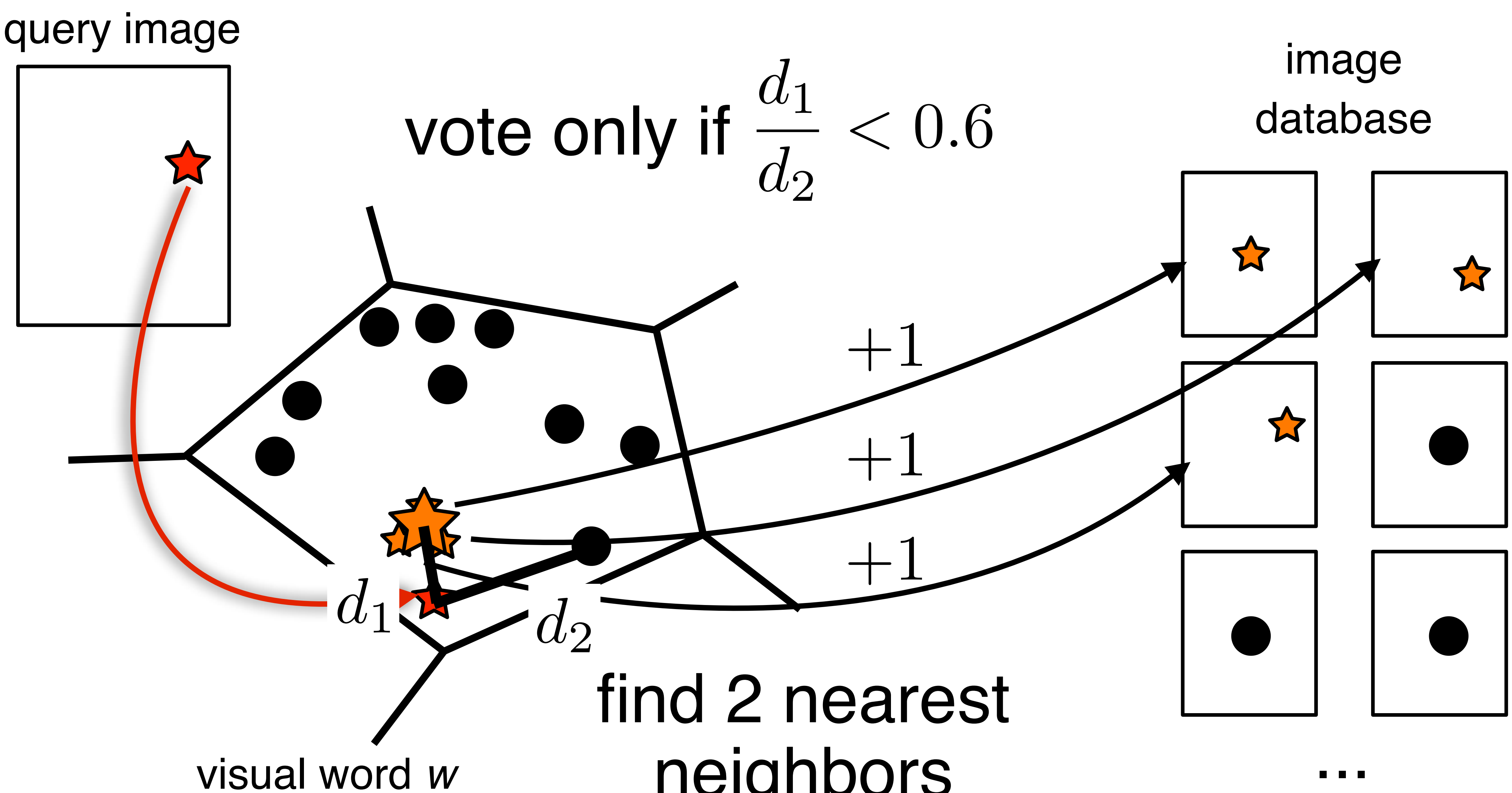
Correspondence Voting

Idea: Find corresponding 3D point



Correspondence Voting

Idea: Find corresponding 3D point

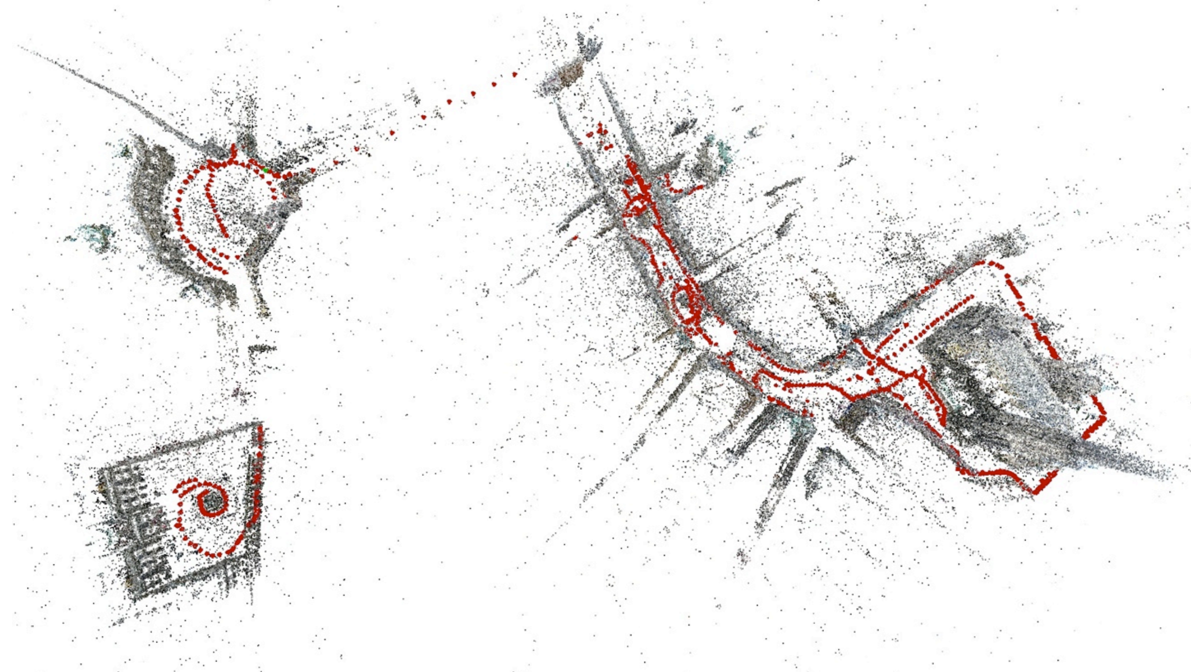


Experimental Evaluation

Aachen

dataset available at
http://www.graphics.rwth-aachen.de/localization

Vienna



dataset kindly provided by
A. Irschara [Irschara,CVPR'09]
used in [Irschara,CVPR'09], [Li,ECCV'10], [Sattler,ICCV'11]

Dataset	# 3D points	# db images	# query images	mean # features per query
Aachen	1.54M	3047	369	9707.29
Vienna	1.12M	1324	266	8648.66

Registration Performance

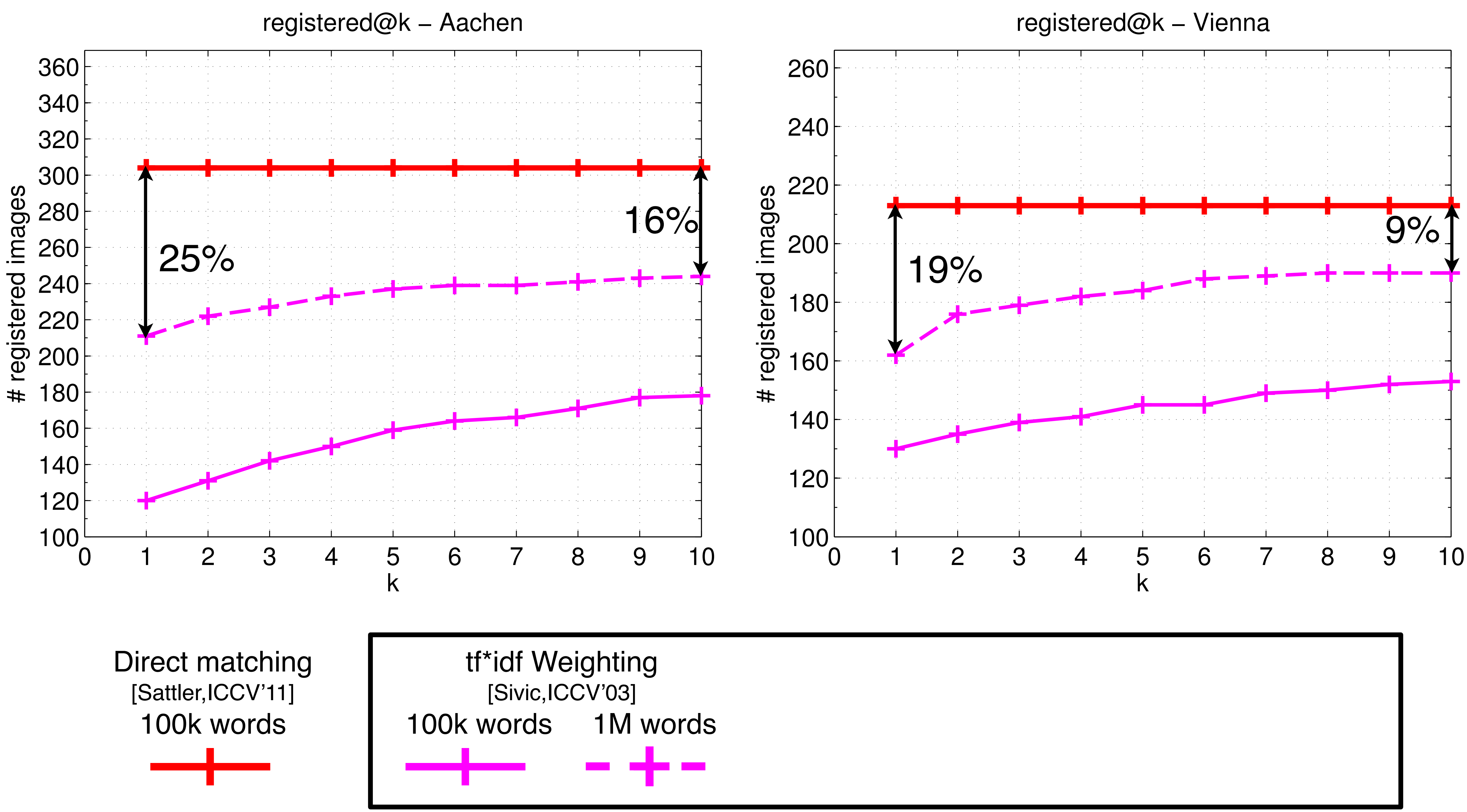
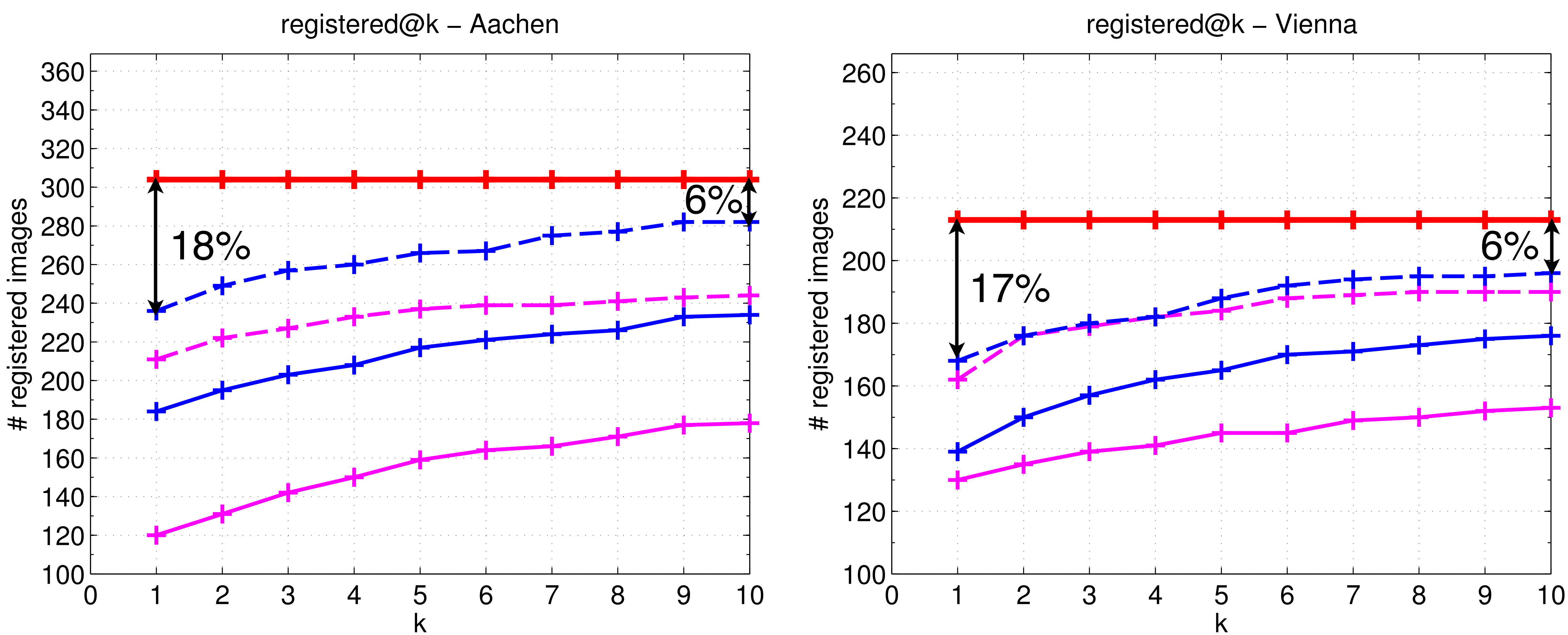


image retrieval-based

Registration Performance



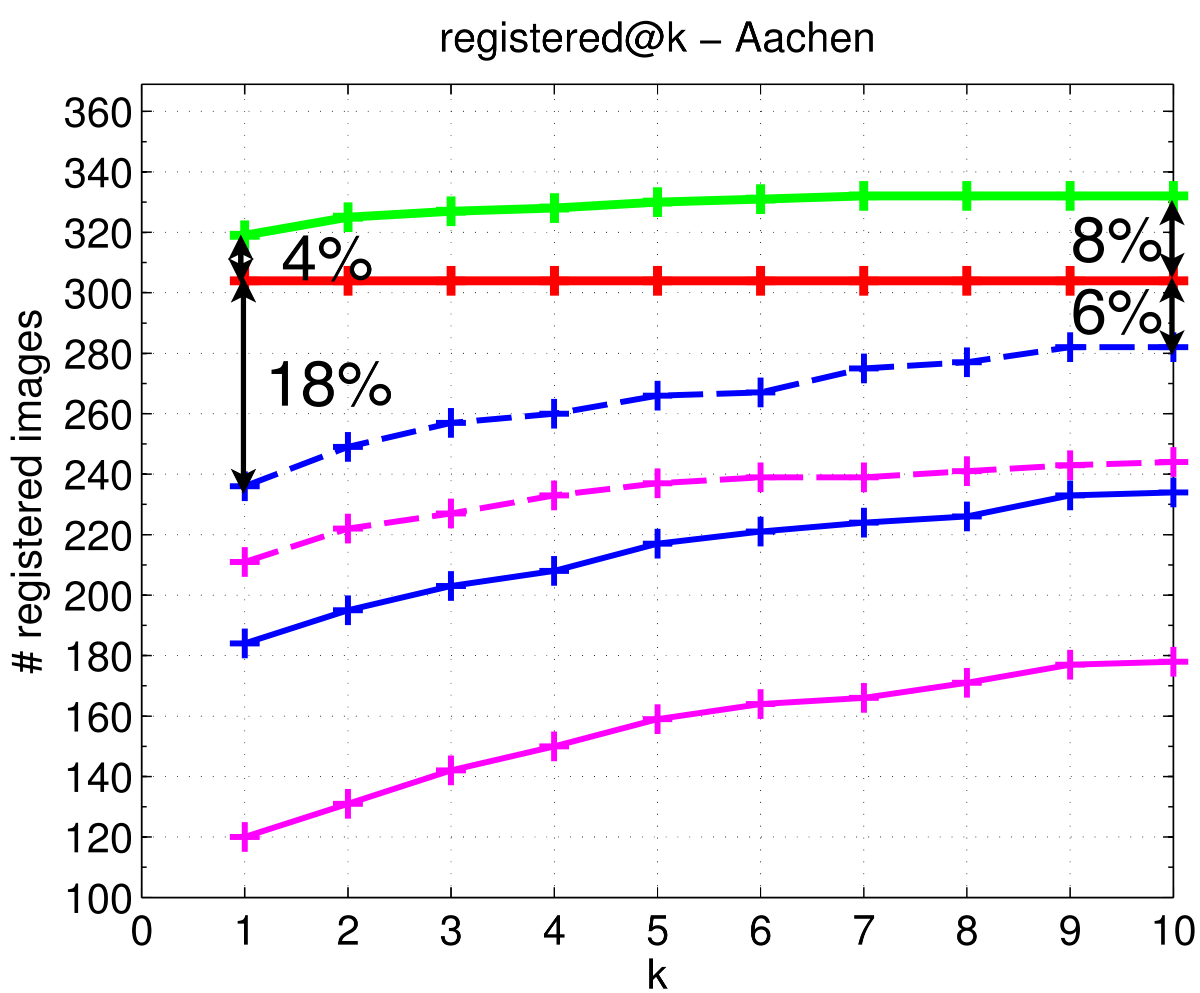
Direct matching
[Sattler,ICCV'11]
100k words

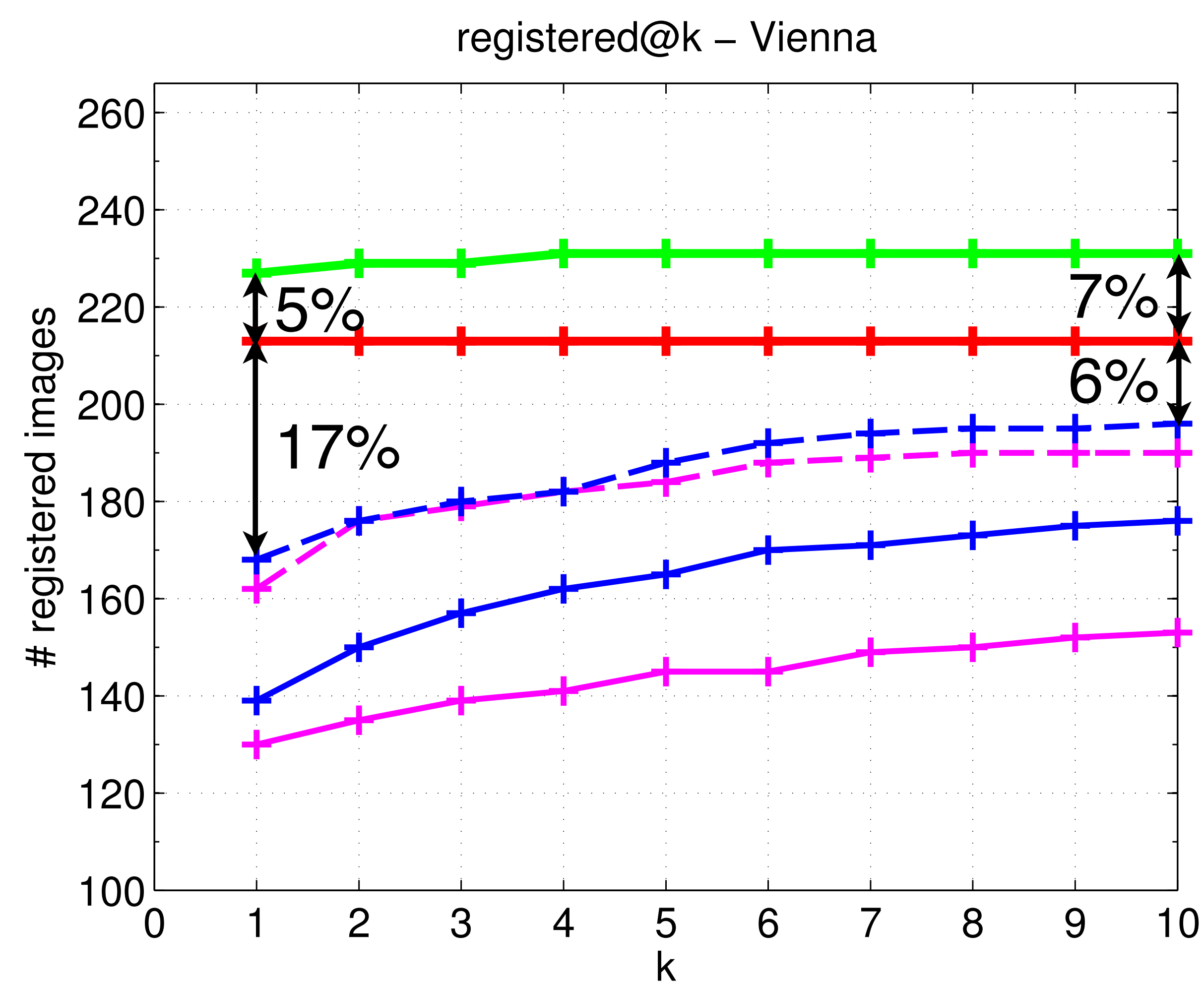
tf*idf Weighting
[Sivic,ICCV'03]
100k words 1M words

Probabilistic Scoring
[Irschara,CVPR'09]
100k words 1M words

image retrieval-based

Registration Performance





Direct matching
[Sattler,ICCV'11]
100k words

tf*idf Weighting
[Sivic,ICCV'03]
100k words 1M words

Probabilistic Scoring
[Irschara,CVPR'09]
100k words 1M words

Correspondence Voting
100k words

image retrieval-based

Comparison

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art
Correspondence Voting	SIFT descriptor (128 bytes)	descriptor distance computation	8% more images

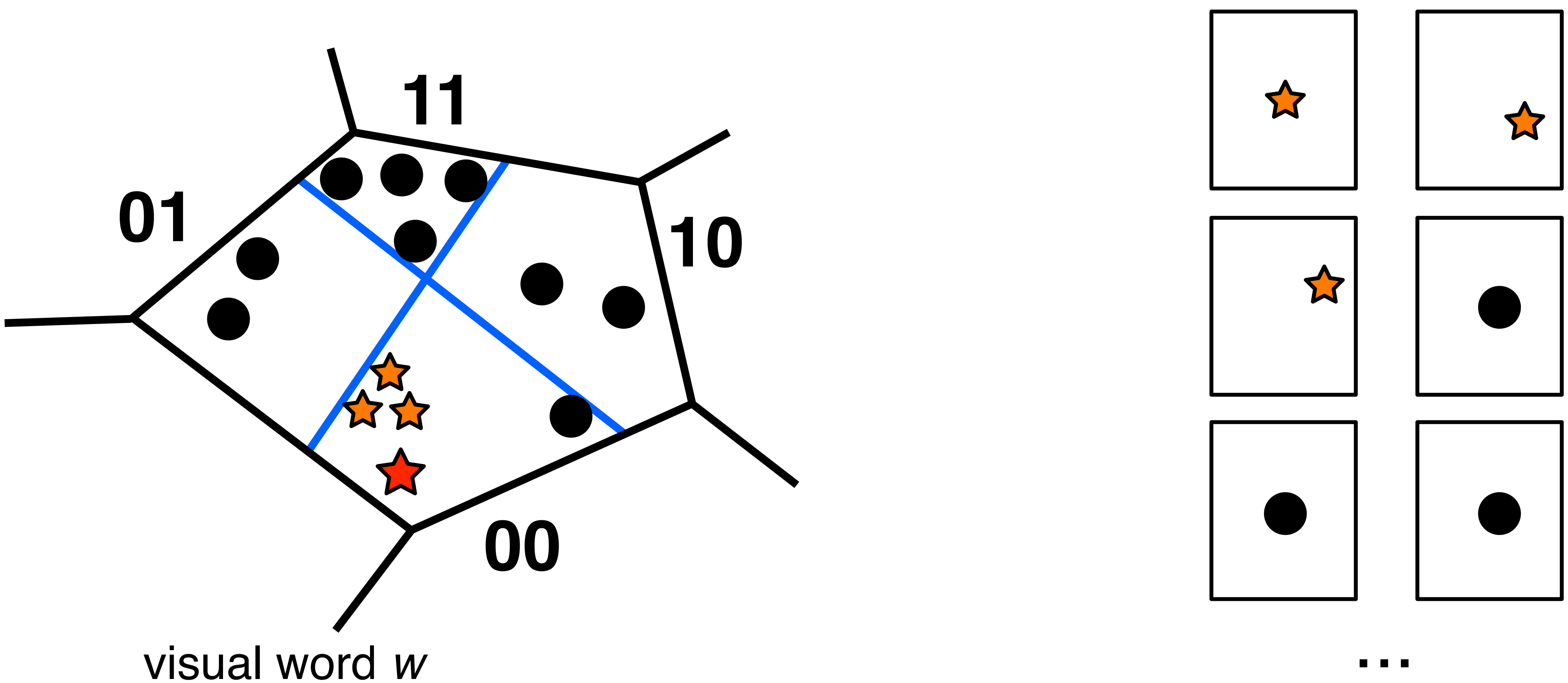
Comparison

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art
Correspondence Voting	SIFT descriptor (128 bytes)	descriptor distance computation	8% more images

Hamming Voting

Jégou, Douze, Schmid. ***Hamming Embedding*** *and Weak Geometric consistency for large-scale image search*. ECCV'08

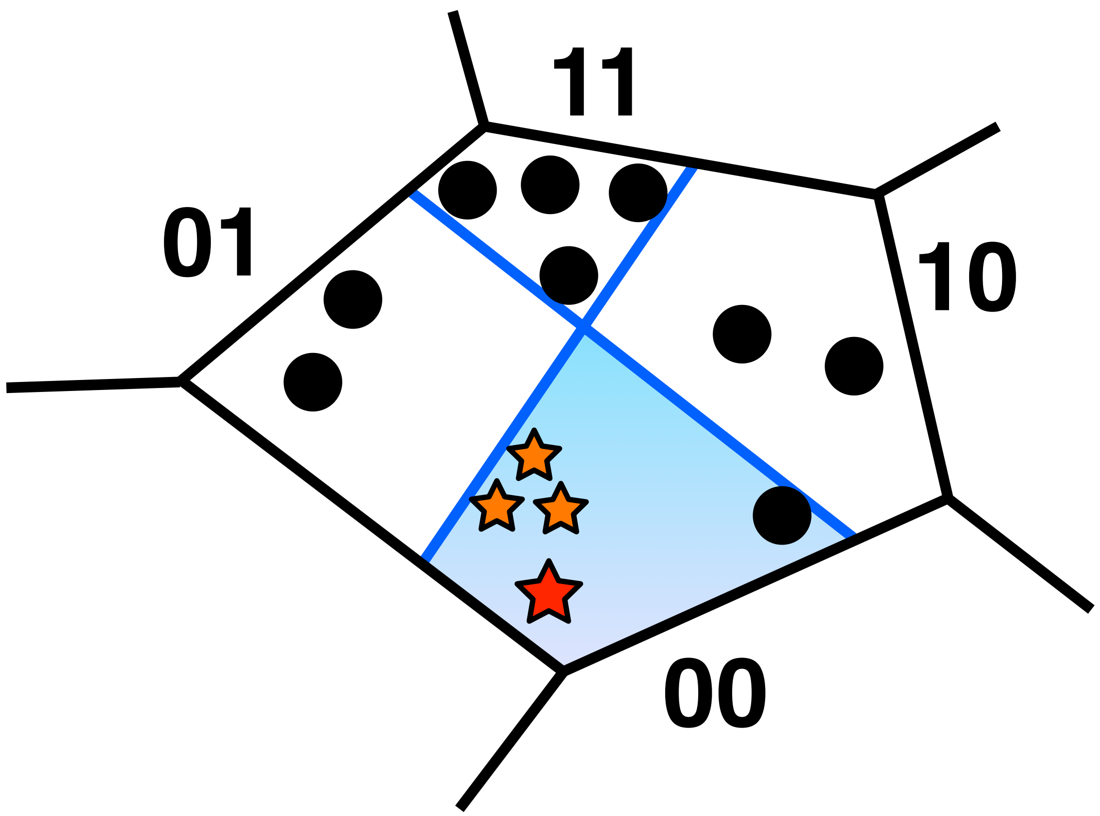
- Random projection: $\mathbb{R}^{128} \rightarrow \mathbb{R}^d$
- Thresholding per visual word: $\mathbb{R}^d \rightarrow \{0,1\}^d$



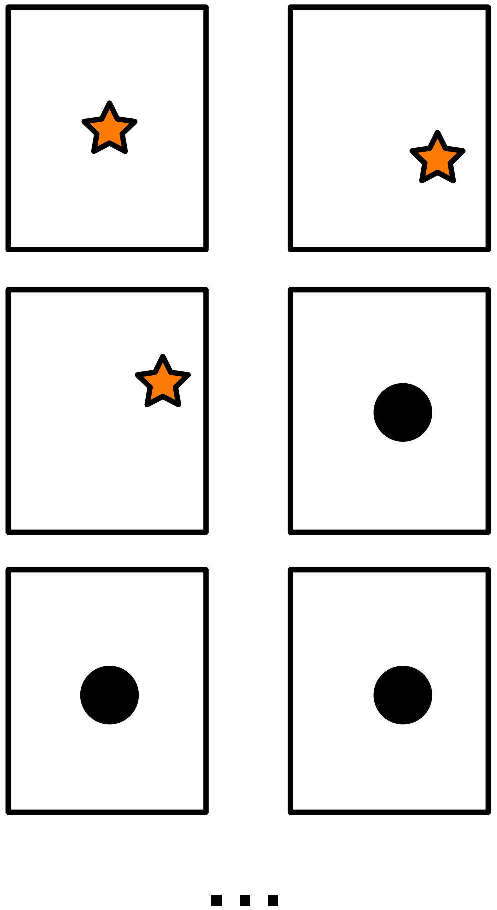
Hamming Voting

Jégou, Douze, Schmid. ***Hamming Embedding*** *and Weak Geometric consistency for large-scale image search*. ECCV'08

- Random projection: $\mathbb{R}^{128} \rightarrow \mathbb{R}^d$
- Thresholding per visual word: $\mathbb{R}^d \rightarrow \{0,1\}^d$



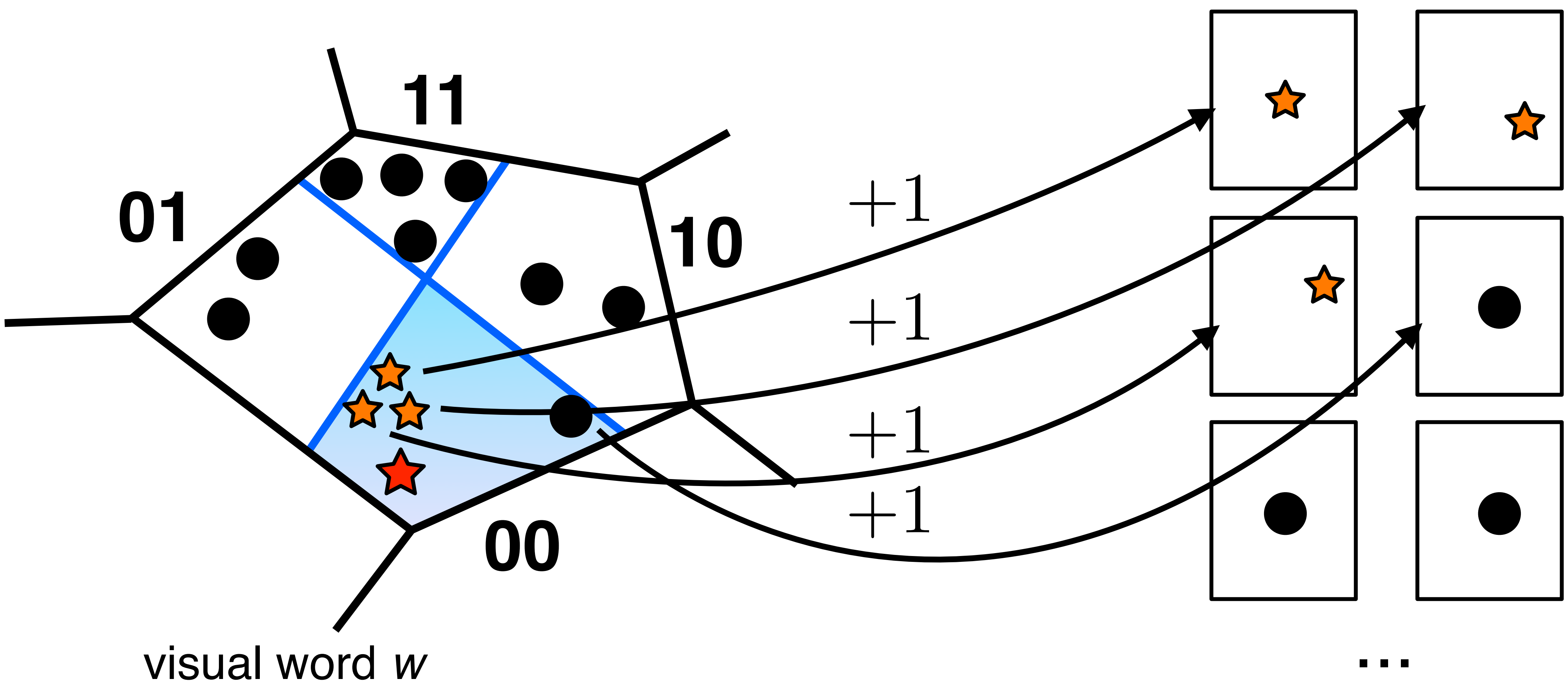
visual word *w*



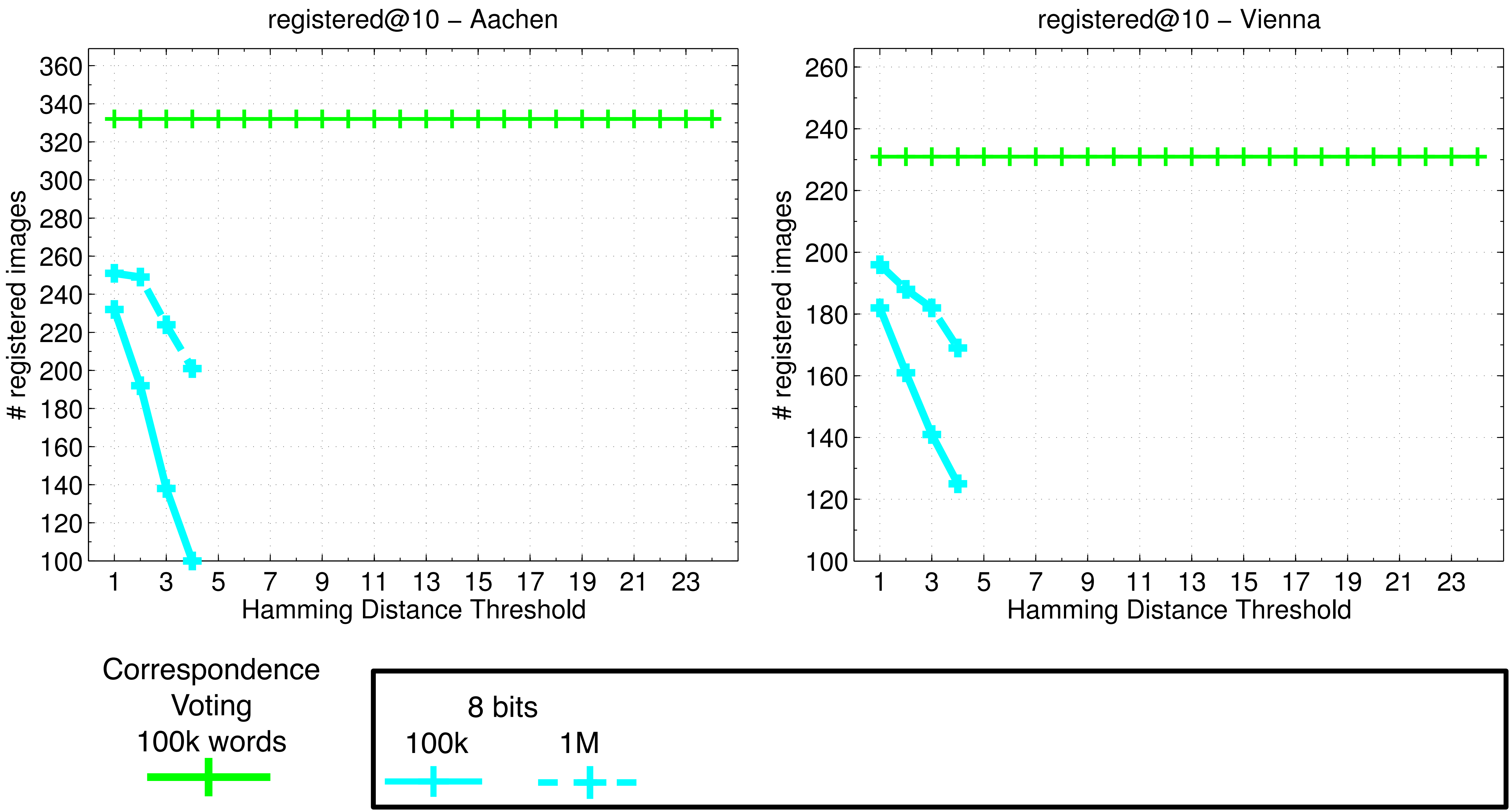
Hamming Voting

Jégou, Douze, Schmid. ***Hamming Embedding*** *and Weak Geometric consistency for large-scale image search*. ECCV'08

- Random projection: $\mathbb{R}^{128} \rightarrow \mathbb{R}^d$
- Thresholding per visual word: $\mathbb{R}^d \rightarrow \{0,1\}^d$

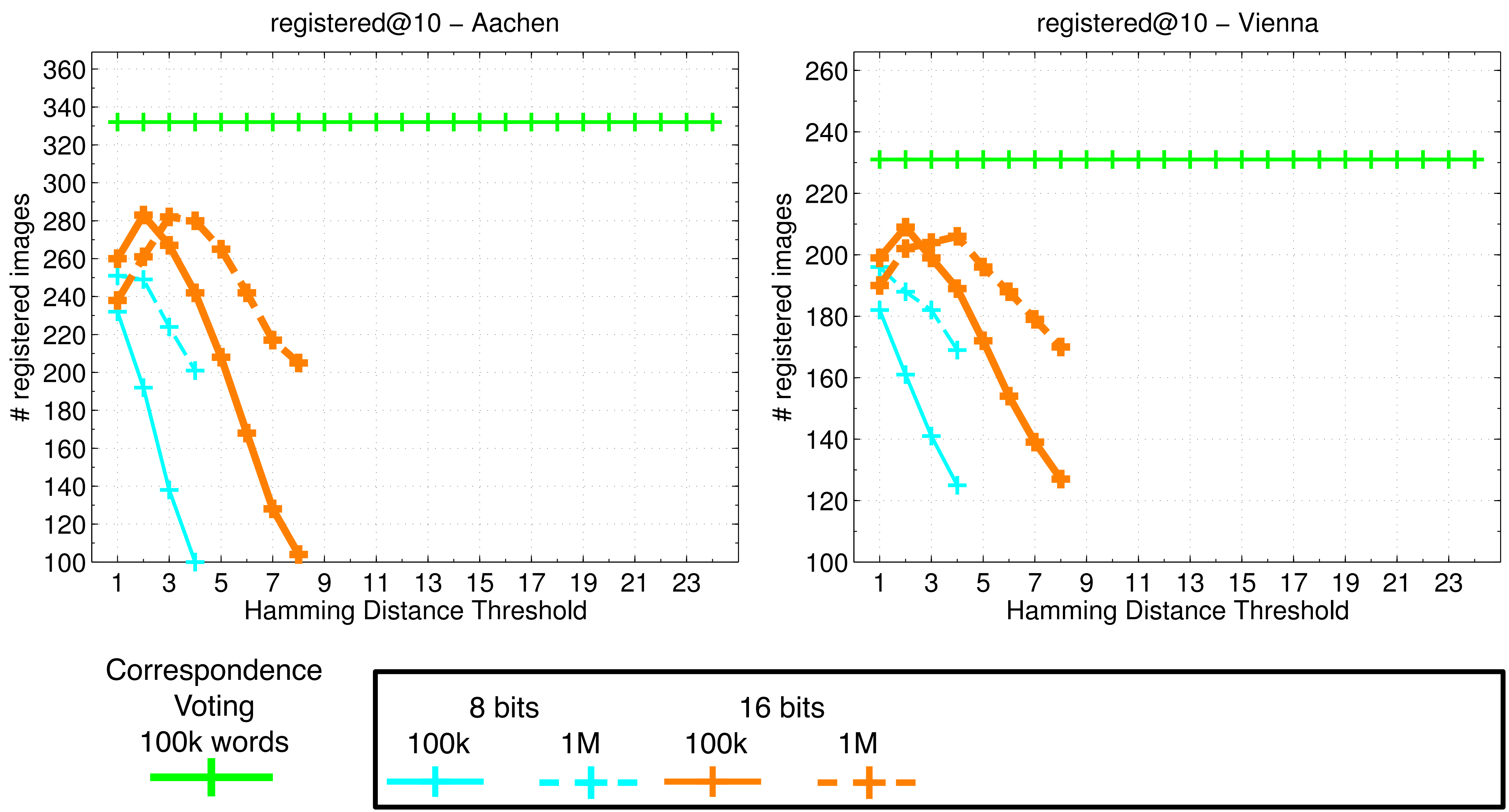


Hamming Voting



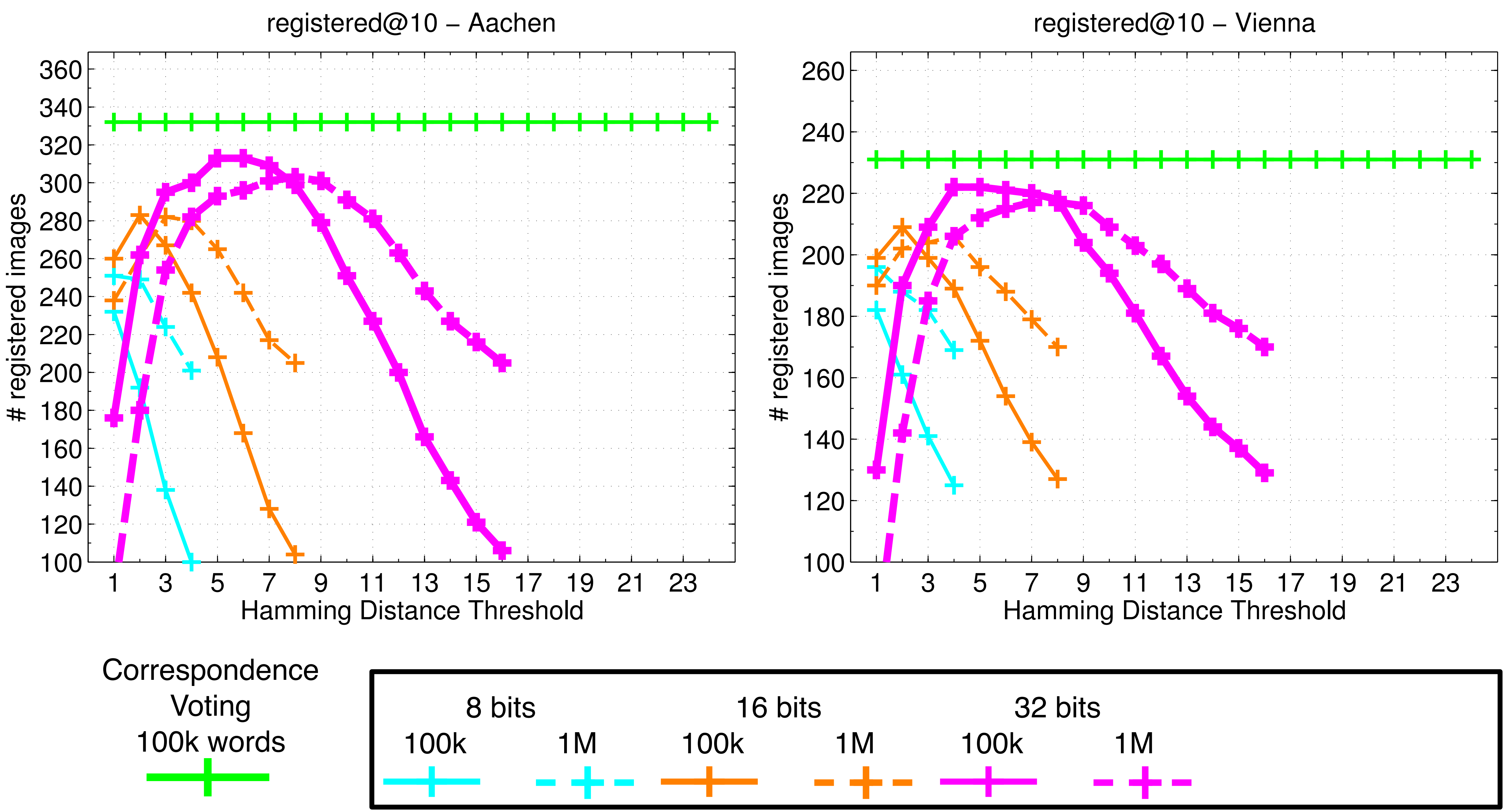
Hamming Voting

Hamming Voting



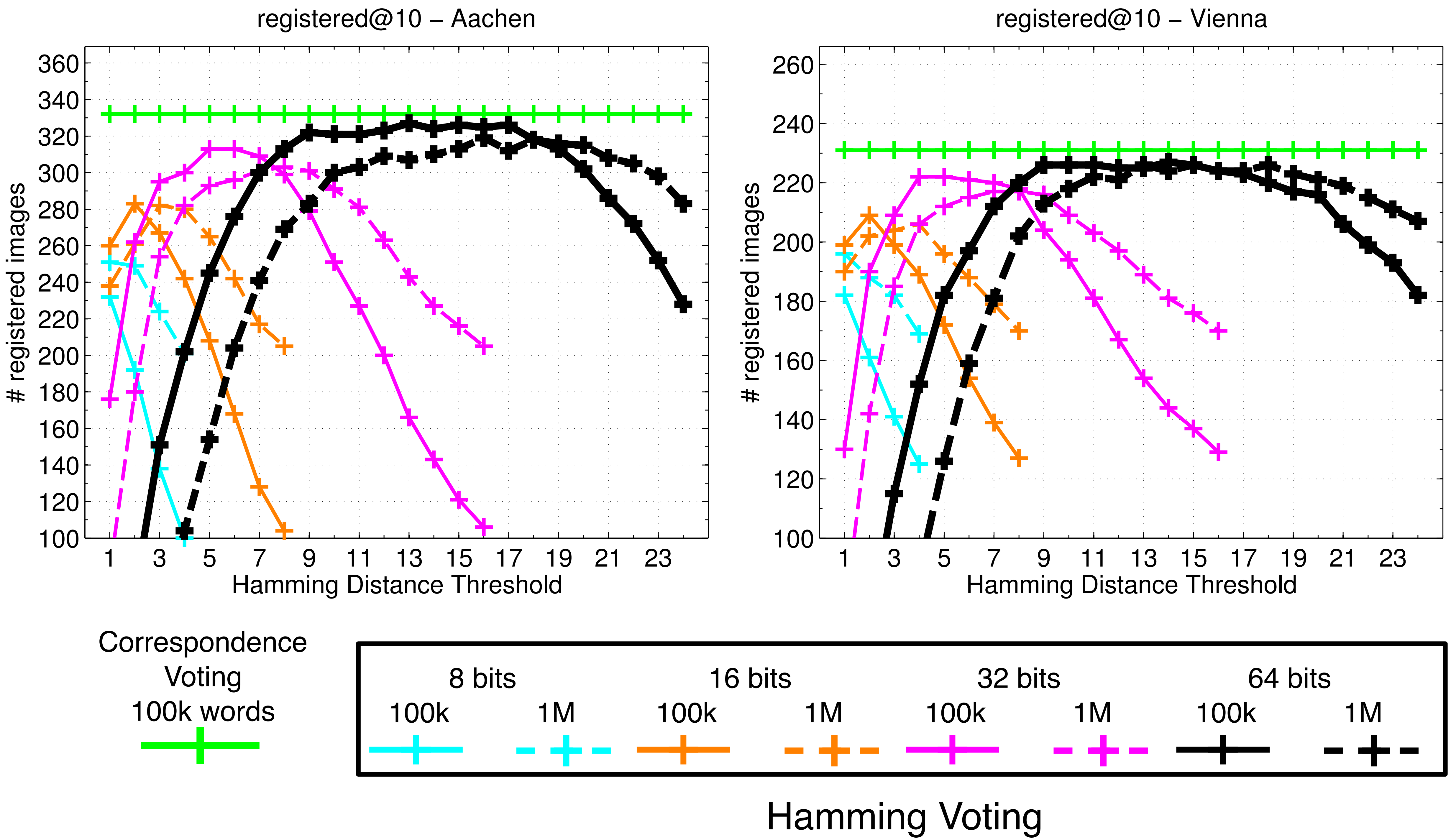
Hamming Voting

Hamming Voting



Hamming Voting

Hamming Voting



Comparison

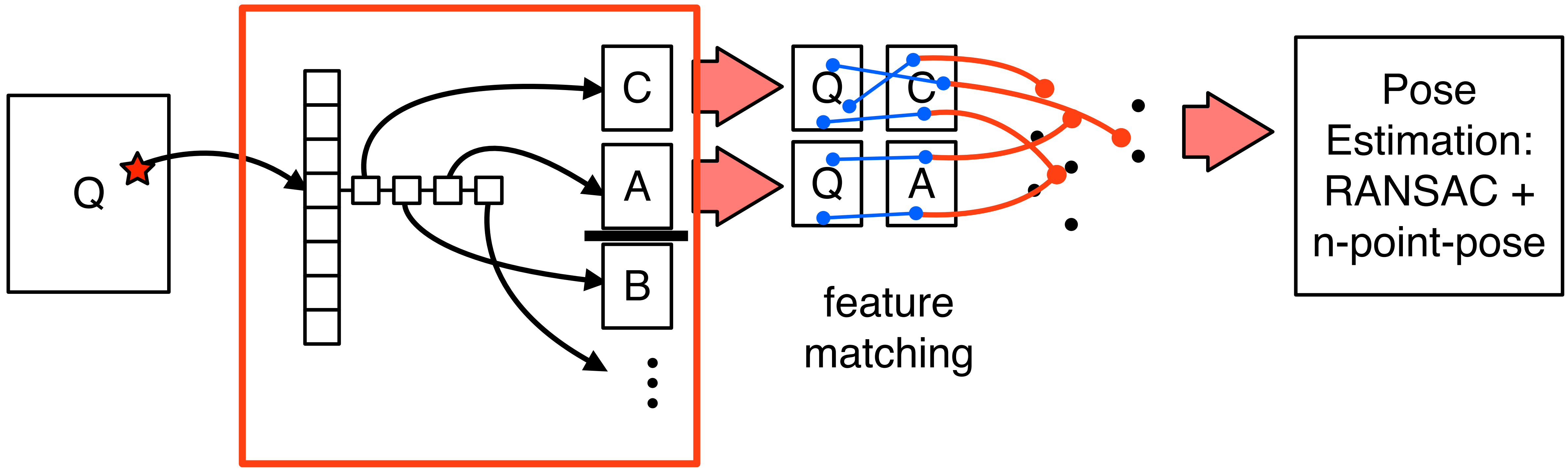
	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art
Correspondence Voting	SIFT descriptor (128 bytes)	descriptor distance computation	8% more images
Hamming Voting (64 bits)	binary descriptor (8 bytes)	Hamming distance computation (10^6 computations ≈ 2ms) + vote	6% more images

Comparison

	Scalability		Registration Performance
	Inverted file entry size	Run time cost / entry	
Image retrieval	image id (4 bytes)	vote for image	6-18% less images
Direct matching	SIFT descriptor (128 bytes)	descriptor distance computation	state-of-the-art
Correspondence Voting	SIFT descriptor (128 bytes)	descriptor distance computation	8% more images
Hamming Voting (64 bits)	binary descriptor (8 bytes)	Hamming distance computation (10^6 computations ≈ 2ms) + vote	6% more images

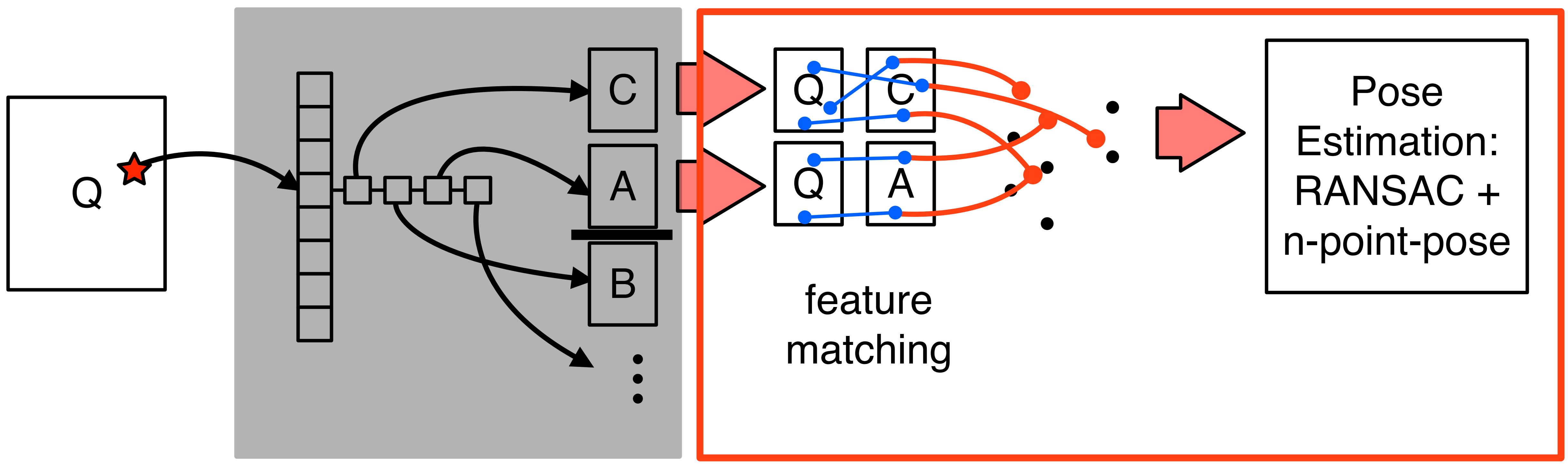
Additional cost for Hamming Voting: **+ ~23ms per query image** (projection, thresholding)

Correspondence Selection



- Run time cost: Voting + **Regular SIFT matching**
  - Build kd-tree for query features
  - Match database features against kd-tree
  - Introduces additional computations

Correspondence Selection



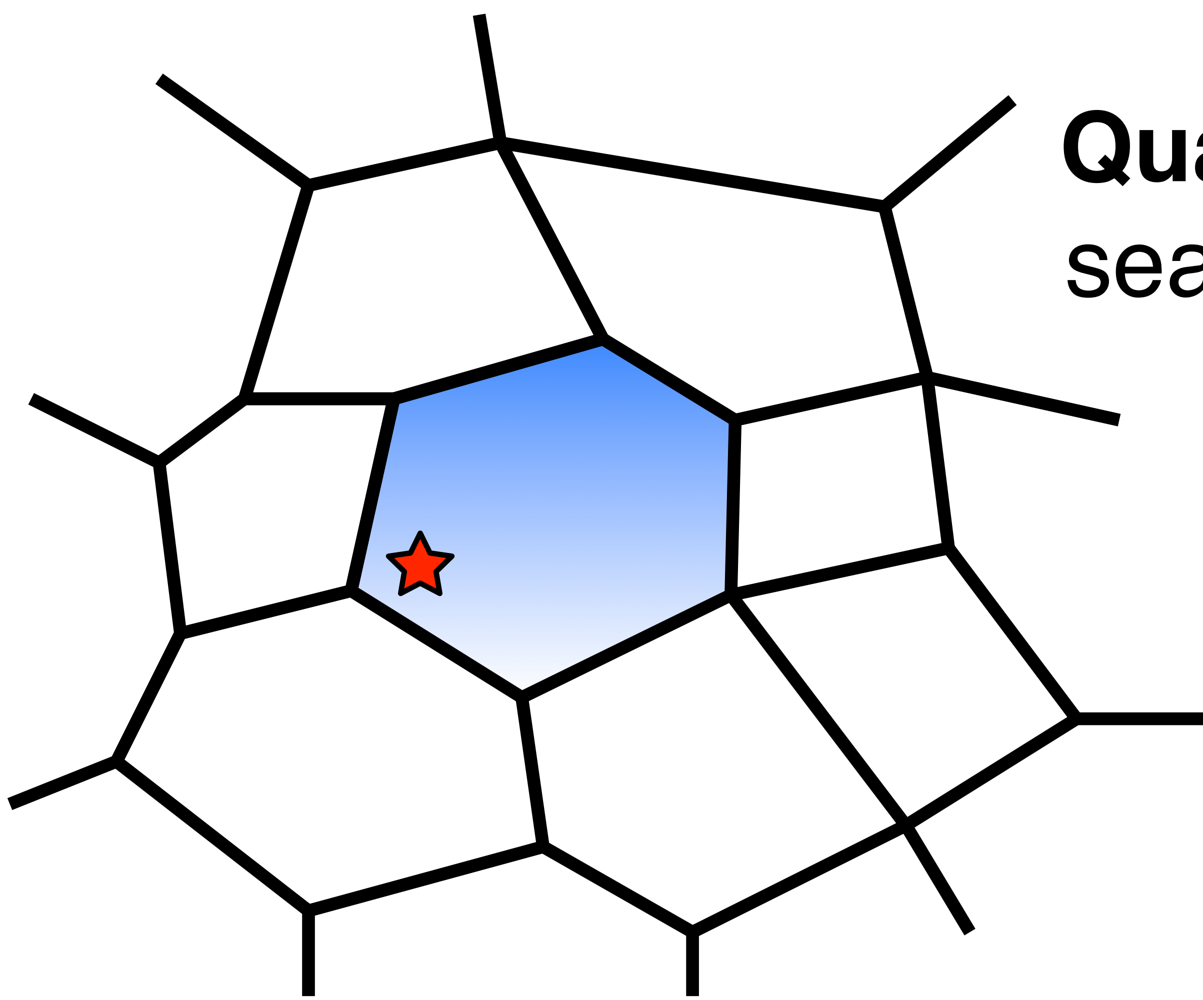
- Run time cost: Voting + **Regular SIFT matching**
  - Build kd-tree for query features
  - Match database features against kd-tree
  - Introduces additional computations

Correspondence Selection

- Idea: Re-use matches from voting stage
- Problem: Not enough correspondences

Correspondence Selection

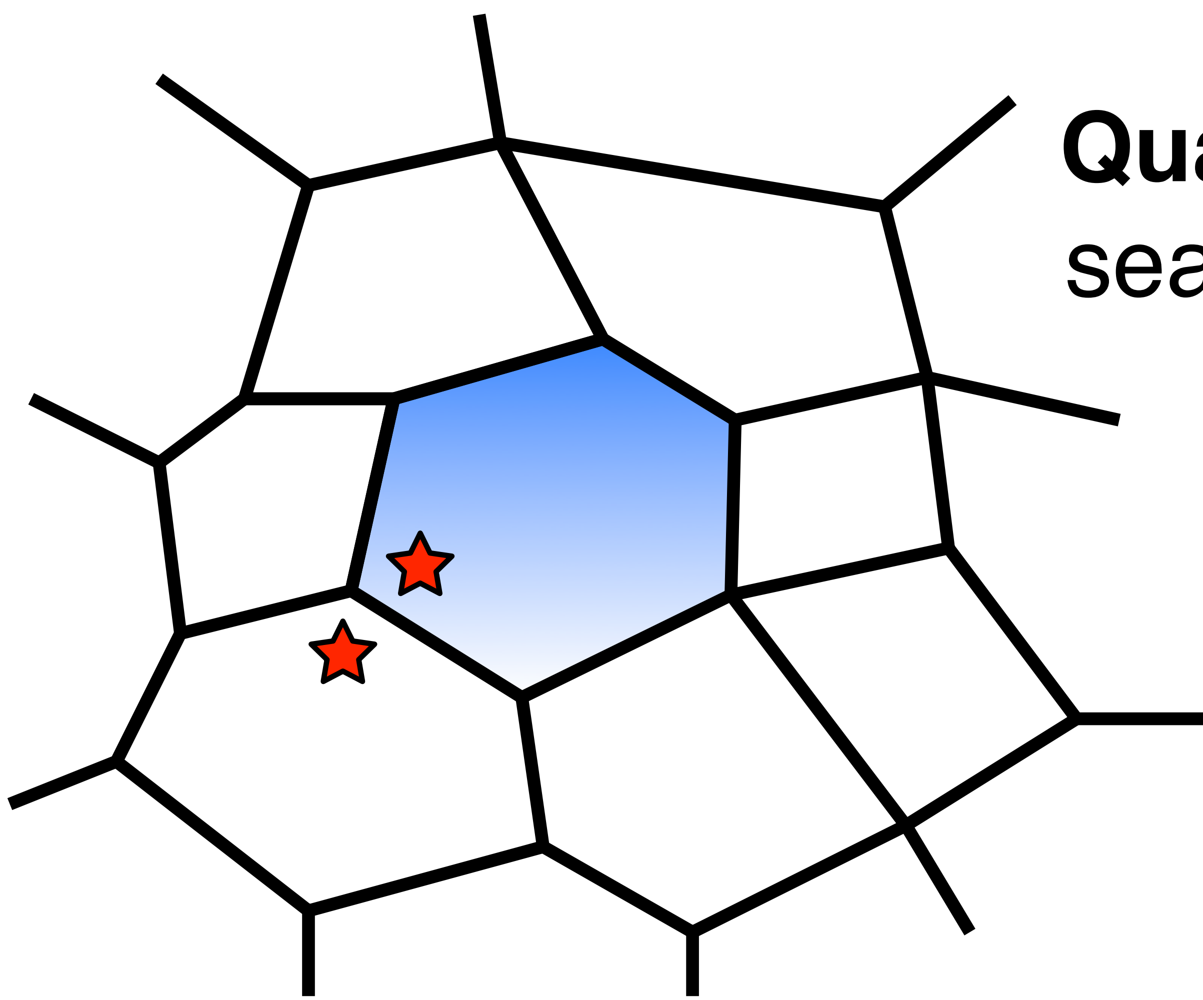
- Idea: Re-use matches from voting stage
- Problem: Not enough correspondences



Quantized Matching: Restrict search to visual word [Sattler,ICCV'11]

Correspondence Selection

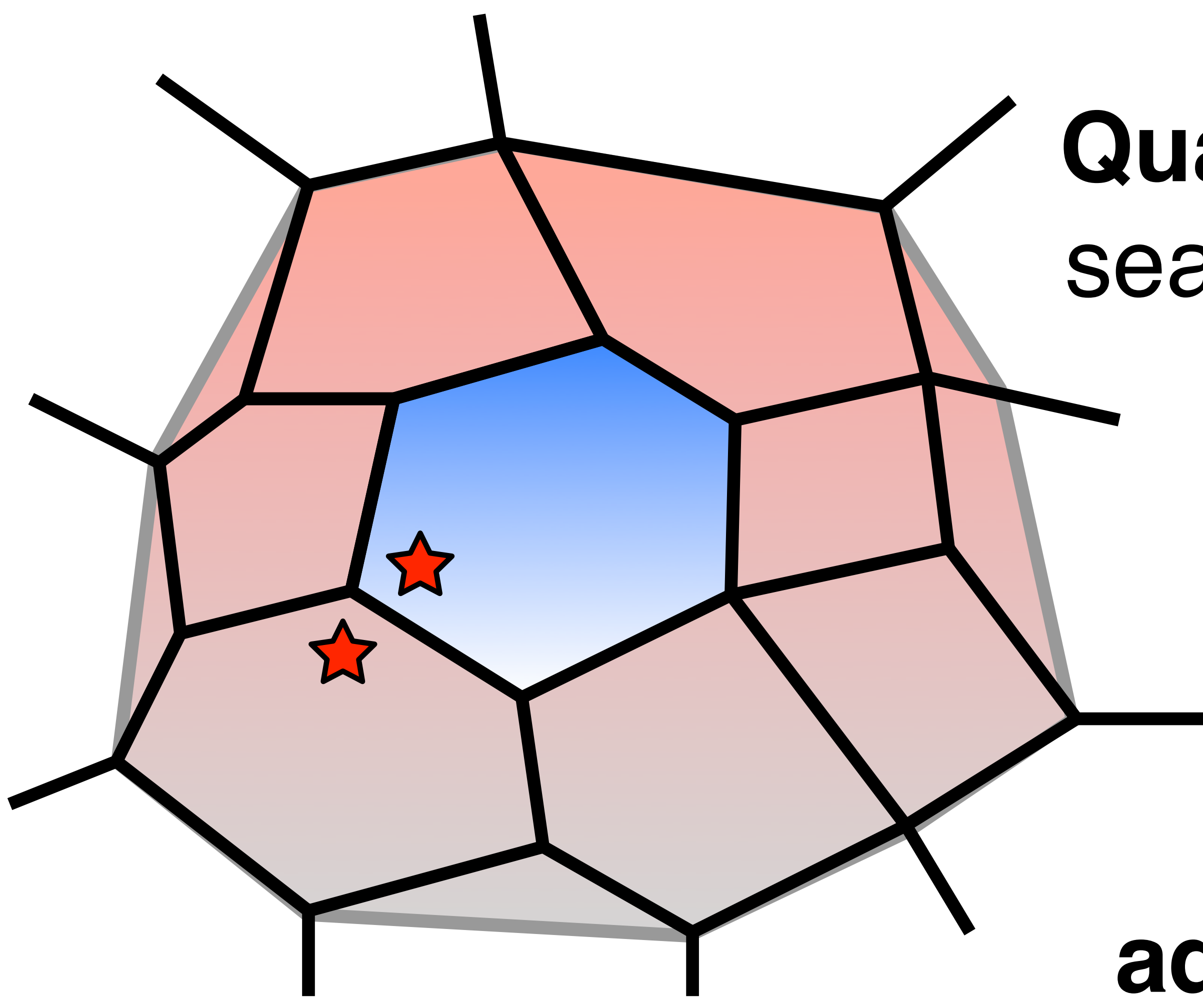
- Idea: Re-use matches from voting stage
- Problem: Not enough correspondences



Quantized Matching: Restrict search to visual word [Sattler,ICCV'11]

Correspondence Selection

- Idea: Re-use matches from voting stage
- Problem: Not enough correspondences



Quantized Matching: Restrict search to visual word [Sattler,ICCV'11]

Coarser vocabulary from hierarchical clustering, **no additional assignment costs**

Correspondence Selection

Quantized Matching: Restrict nearest neighbor search to same visual word

Matching Method	Voc. Size	# Images Registered	Correspondence Search [ms]	RANSAC ok [ms]	RANSAC err [ms]
Regular SIFT	-	320 (87%)	300.3	0.9	0.0
Quantized SIFT	100	319 (86%)	14.5	3.1	155.3
Quantized Hamming (64-bit)	100	307 (83%)	3.6	141.6	2825.0

median timings per query image - database image pair

Correspondence Selection

Quantized Matching: Restrict nearest neighbor search to same visual word

Matching Method	Voc. Size	# Images Registered	Correspondence Search [ms]	RANSAC ok [ms]	RANSAC err [ms]
Regular SIFT	-	320 (87%)	300.3	0.9	0.0
Quantized SIFT	100	319 (86%)	14.5	3.1	155.3
Quantized Hamming (64-bit)	100	307 (83%)	3.6	141.6	2825.0

median timings per query image - database image pair

Correspondence Selection

Quantized Matching: Restrict nearest neighbor search to same visual word

Matching Method	Voc. Size	# Images Registered	Correspondence Search [ms]	RANSAC ok [ms]	RANSAC err [ms]
Regular SIFT	-	320 (87%)	300.3	0.9	0.0
Quantized SIFT	100	319 (86%)	14.5	3.1	155.3
Quantized Hamming (64-bit)	100	307 (83%)	3.6	141.6	2825.0

median timings per query image - database image pair

Correspondence Selection

Quantized Matching: Restrict nearest neighbor search to same visual word

Matching Method	Voc. Size	# Images Registered	Correspondence Search [ms]	RANSAC ok [ms]	RANSAC err [ms]
Regular SIFT	-	320 (87%)	300.3	0.9	0.0
Quantized SIFT	100	319 (86%)	14.5	3.1	155.3
Quantized Hamming (64-bit)	100	307 (83%)	3.6	141.6	2825.0

median timings per query image - database image pair

Correspondence Selection

Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	319 (86%)	3.1
	1k	304 (82%)	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	307 (83%)	141.6
	1k	300 (81%)	3.5
	10k	272 (74%)	0.9

Correspondence Selection

Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	$\boxed{319\ (86\%)}$	3.1
	1k	304 (82%)	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	$\boxed{307\ (83\%)}$	141.6
	1k	300 (81%)	3.5
	10k	272 (74%)	0.9

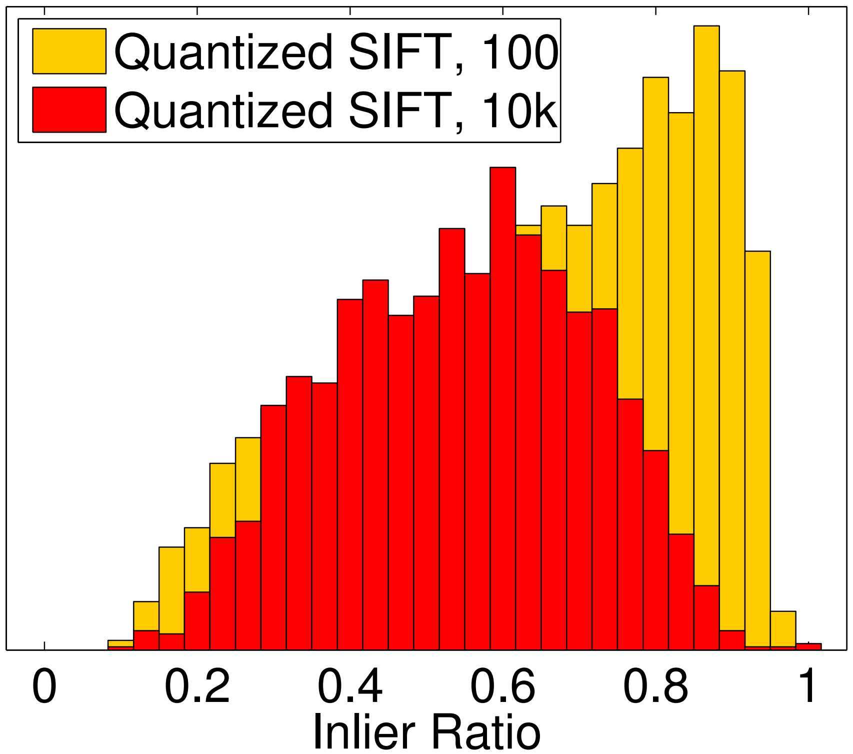
Correspondence Selection

Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	319 (86%)	3.1
	1k	$\boxed{304\ (82\%)}$	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	307 (83%)	141.6
	1k	$\boxed{300\ (81\%)}$	3.5
	10k	272 (74%)	0.9

Correspondence Selection

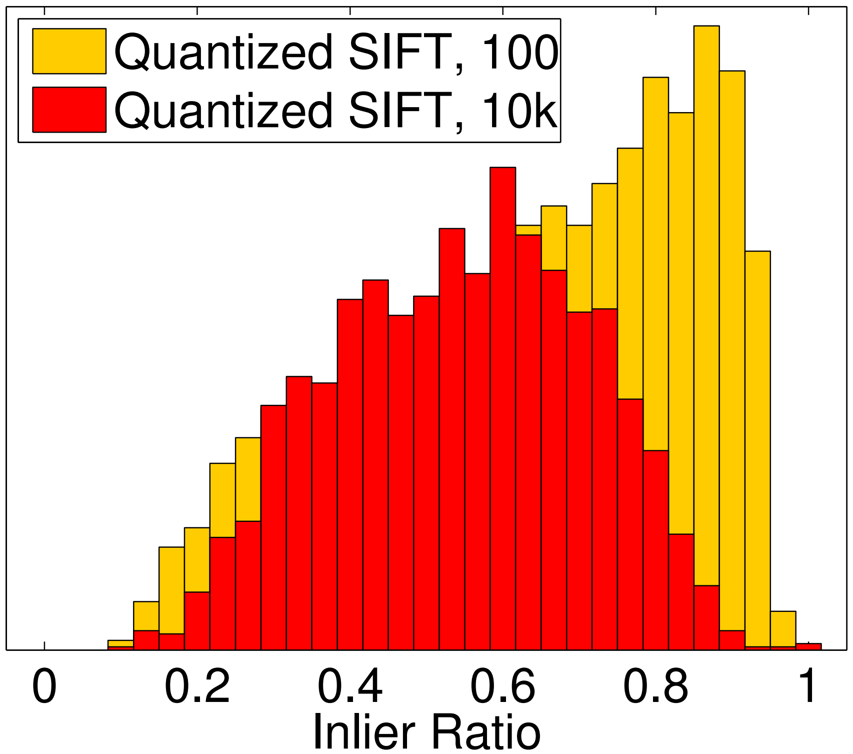
Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	319 (86%)	3.1
	1k	304 (82%)	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	307 (83%)	141.6
	1k	300 (81%)	3.5
	10k	272 (74%)	0.9

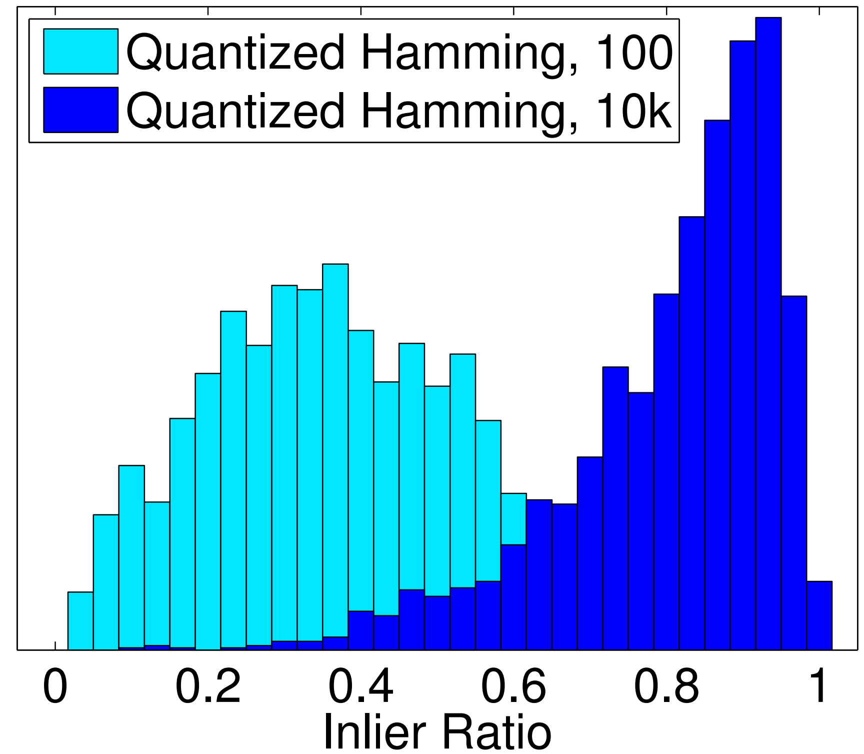
Correspondence Selection



Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	319 (86%)	3.1
	1k	304 (82%)	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	307 (83%)	141.6
	1k	300 (81%)	3.5
	10k	272 (74%)	0.9

Correspondence Selection





Matching Method	Voc. Size	# Images Registered	RANSAC ok [ms]
Regular SIFT	-	320 (87%)	0.9
Quantized SIFT	100	319 (86%)	3.1
	1k	304 (82%)	17.4
	10k	246 (67%)	10.2
Quantized Hamming (64-bit)	100	307 (83%)	141.6
	1k	300 (81%)	3.5
	10k	272 (74%)	0.9

Conclusion

- **Incorrect votes** major source of error for image retrieval-based localization
- **Hamming voting** avoids most incorrect votes at little computation & memory overhead
- Image retrieval with Hamming voting yields **scalable image-based localization**
- **Correspondence selection** can be accelerated using quantized matching