Large-Scale Approximate Nearest Neighbor Search

Jingdong Wang
Senior Researcher
Microsoft Research, Beijing, China

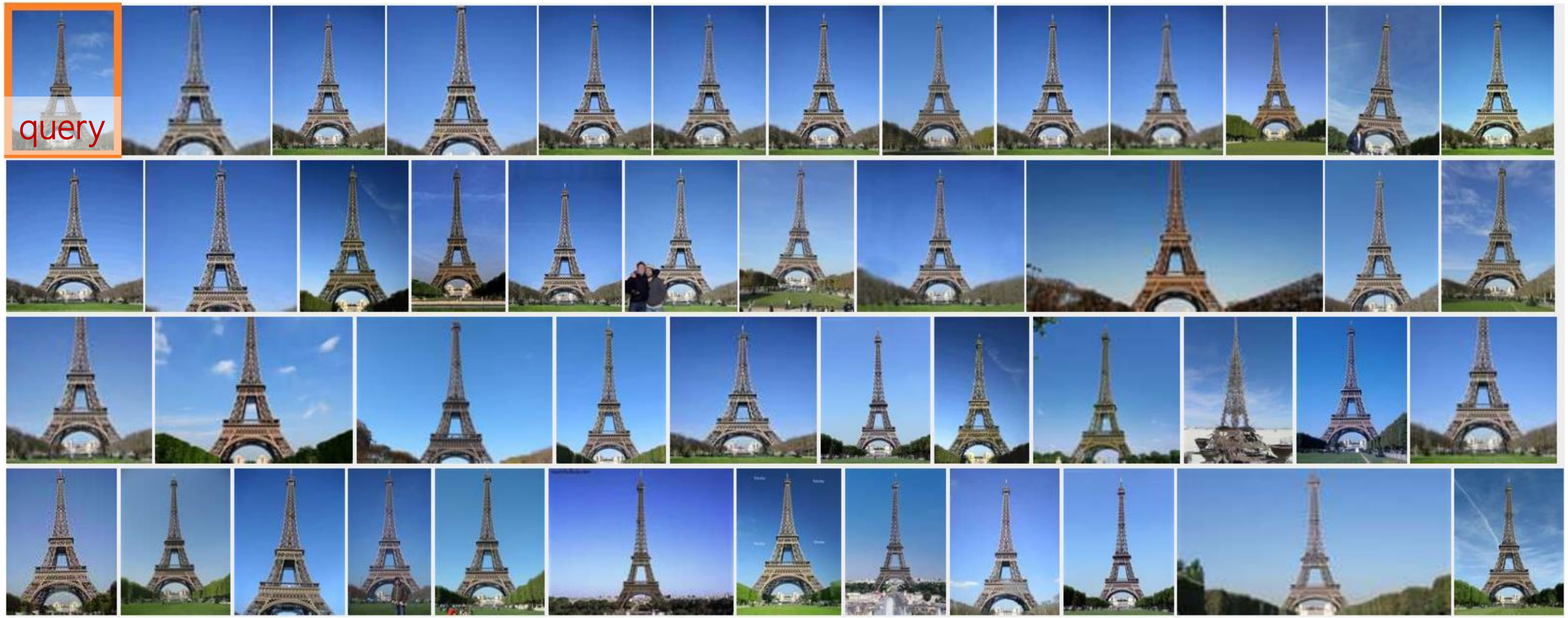
Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

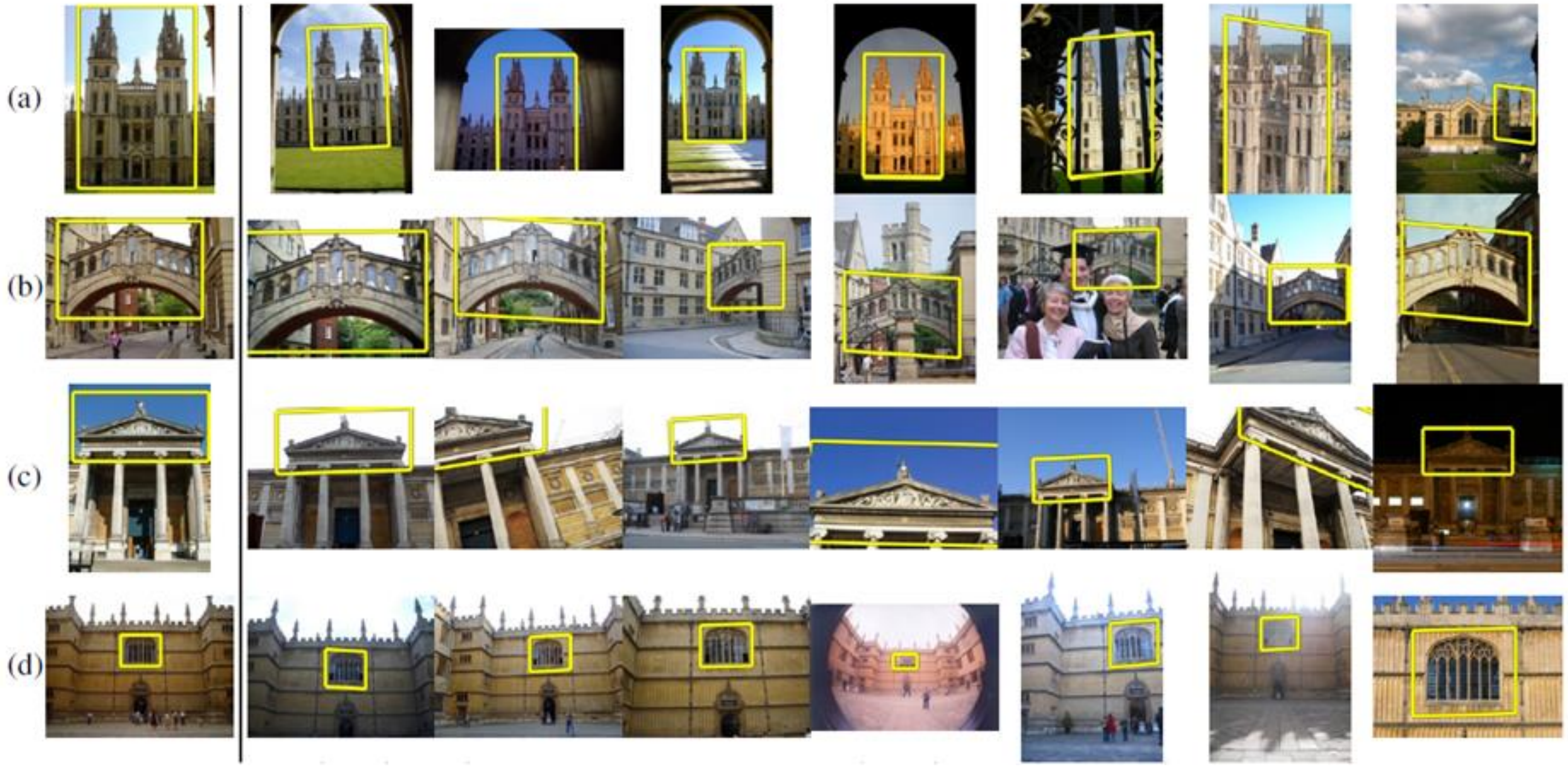
Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

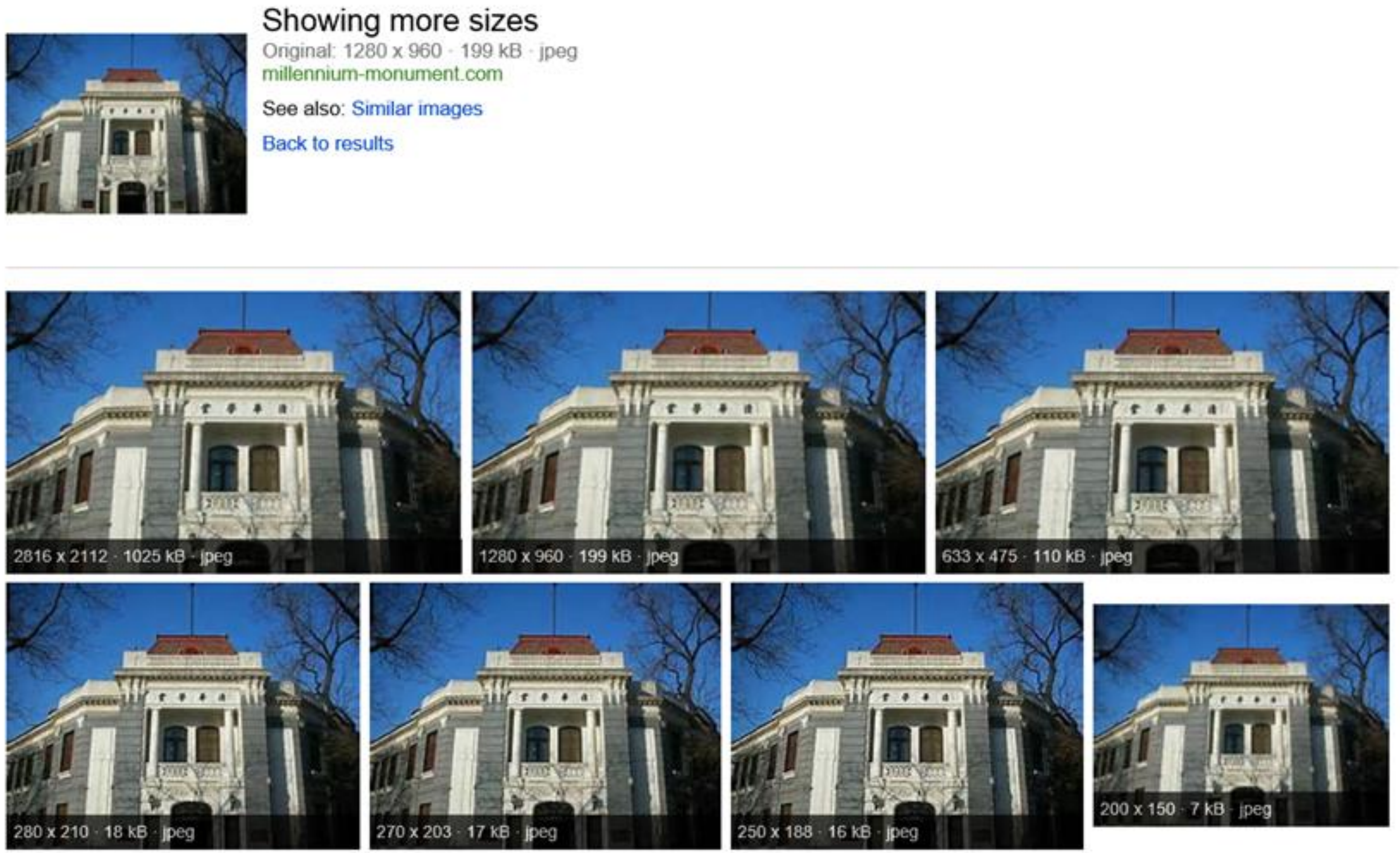
Similar Image Search



Particular Object Retrieval



Duplicate Image Search



Similarity Search

Problem definition

$$\mathrm{NN}(\mathbf{q}) = \arg\min_{\underbrace{\mathbf{x}\in\mathcal{X}}_{\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}}} \underbrace{\mathrm{dist}}_{\text{Euclidean distance}}(\mathbf{q}, \mathbf{x})$$

Time complexity: O(nd)

Principles

Reduce #(distance computations)

- Time complexity: $O(n'd)$, $n' \ll n$
- Tree, neighborhood graph, inverted index

1. High efficiency ☺
2. Large memory cost ☹

Reduce the cost of each distance computation

- Time complexity: $O(nd')$, $d' \ll d$
- Compact codes (hashing, quantization)

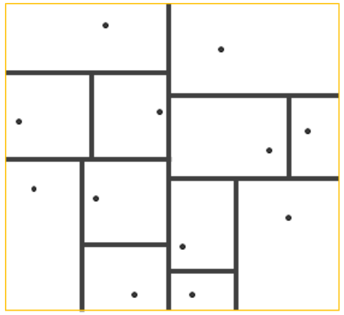
1. Small memory cost ☺
2. Low efficiency ☹

Three key factors: Time, accuracy, memory

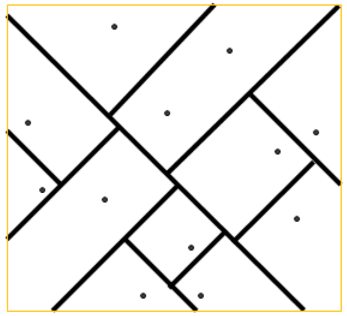
Our work

Index Structure

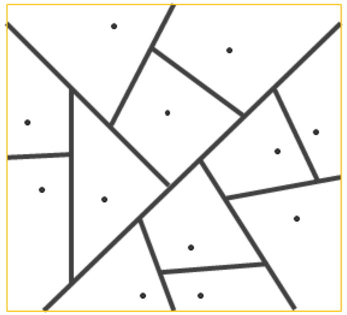
- Trinary-projection tree (CVPR10, TPAMI14)



KD tree: Single coordinate axis

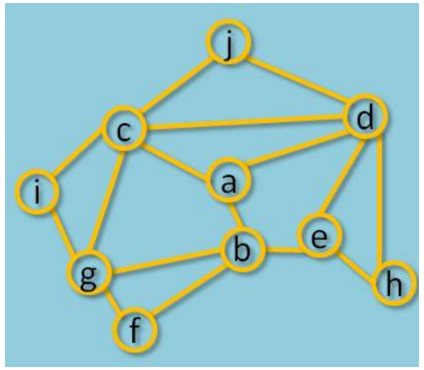


TP tree: Trinary combination of coordinate axes



PCA tree arbitrary axes

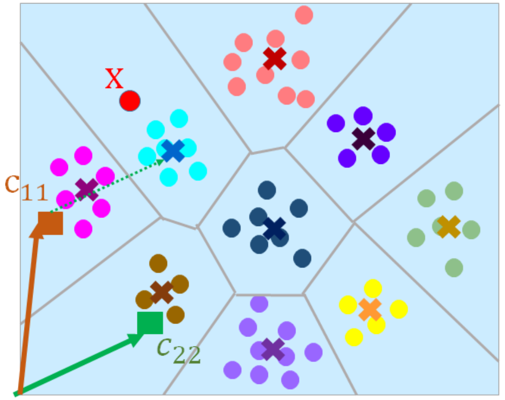
- Neighborhood graph construction and search (ACMMM12, CVPR12, ICCV13)



Compact Coding

- Complementary hashing (ICCV11)
- Order preserving hashing (ACMMM13)
- Optimized distance for binary code ranking (ACMMM14)
- *Composite quantization (ICML14, TPAMI)*
- *Sparse composite quantization (CVPR15)*
- *Collaborative Quantization for Cross-Modal Similarity Search (CVPR16)*
- *Supervised Quantization for Similarity Search (CVPR16)*
- A survey on learning to hash (2015, TPAMI)

x1	010101010110111
x2	101010101010101
x3	110110101010010
x4	001101011010110
x5	101011010101010
x6	101010101010101
x7	110101011010110



Outline

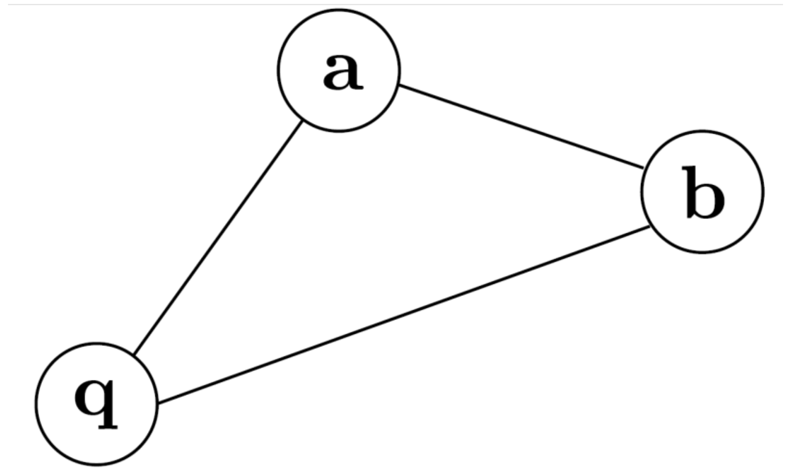
- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Neighborhood graph search

Point b is a nearest neighbor of point a

Point a is near to query q

⇨ b is most likely to be near to q



$$\|\mathbf{q}-\mathbf{b}\|_2 \leq \|\mathbf{q}-\mathbf{a}\|_2 + \|\mathbf{a}-\mathbf{b}\|_2$$

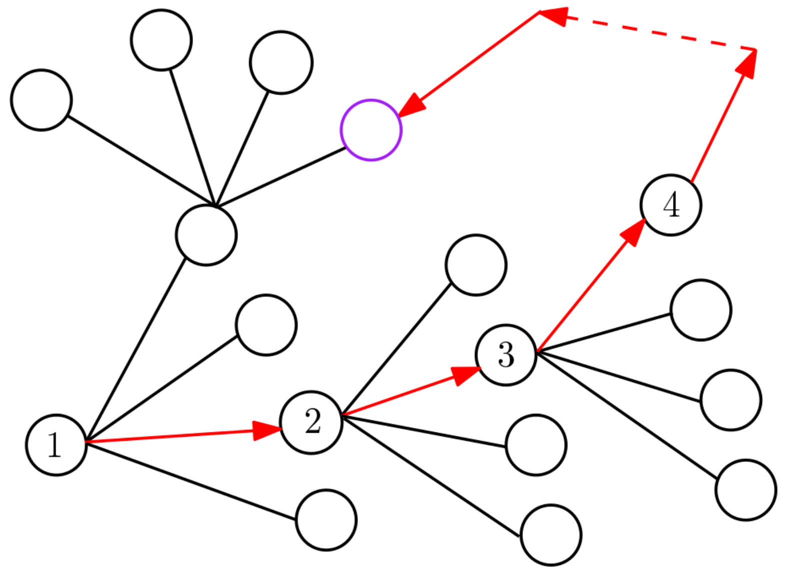
Neighborhood graph as index structure:

Each database point is connected with its k-nearest points

Search

Greedy search

1. Visit the neighborhood points of point *v*
2. Select the point as *v* that is the nearest to the query from the neighborhood points
3. Iterate step 1 and step 2

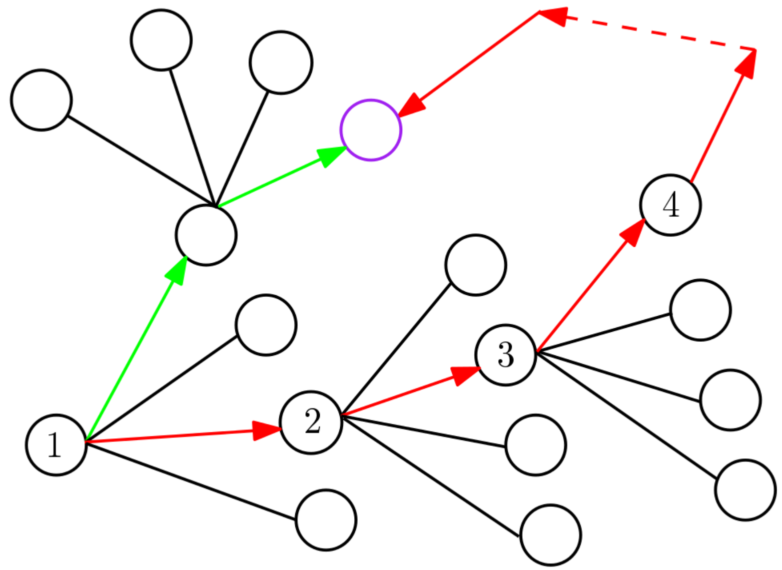


Too greedy to make good use of the other points in the neighborhood except the point nearest to the query

Search

Greedy search

1. Visit the neighborhood points of point *v*
2. Select the point as *v* that is the nearest to the query from the neighborhood points
3. Iterate step 1 and step 2

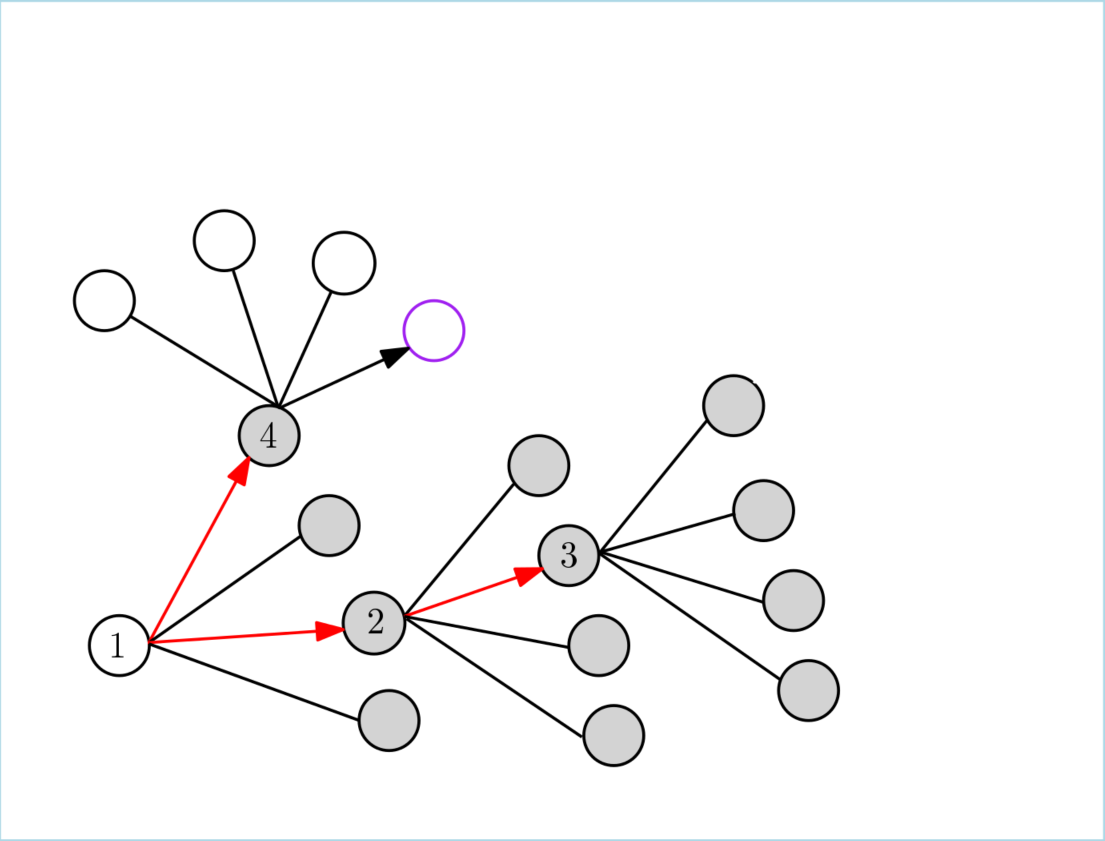


Too greedy to make good use of the other points in the neighborhood except the point nearest to the query

Best-First Search

Best-first search

1. Push the neighborhood points of *v* into a priority queue
2. Pop out the best point *v* from the queue
3. Iterate step 1 and step 2



A priority queue storing the points in the neighborhood

Easy Implementation

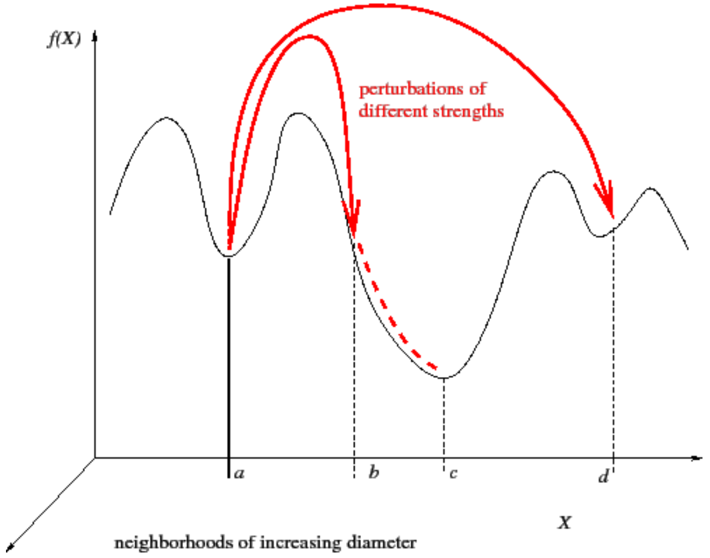
Modify few codes of depth-first search

```
procedure DFS(G,v):
    let S be a stack


    S.push(v)
    while S is not empty
          v = S.pop()
          label v as discovered
          for all edges from v to w in G.adjacentEdges(v) do
              if w is not labeled as discovered:


                     S.push(w)
```

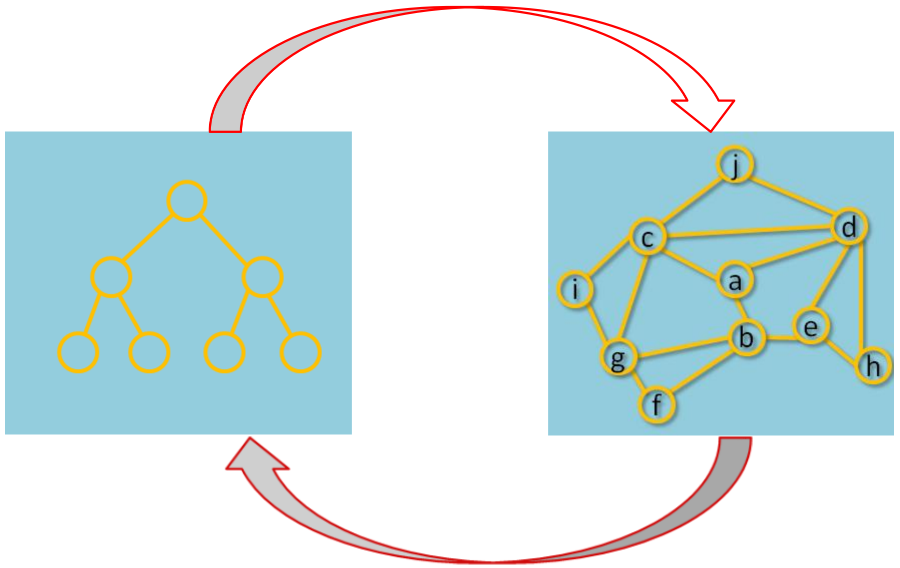
```
procedure BFS(G,s,q,R):
    let S be a priority queue
    s.key = dist(s,q)
    R.push(s)
    S.push(s)
    while S is not empty
          v = S.pop()
          label v as discovered
          for all edges from v to w in G.adjacentEdges(v) do
              if w is not labeled as discovered:
                     w.key = dist(w,q)
                     R.push(w)
                     S.push(w)
```

Nonlocal Search

- Best-first search is local
  - Stuck at a locally optimal point
  - Exhaustive neighborhood expansions

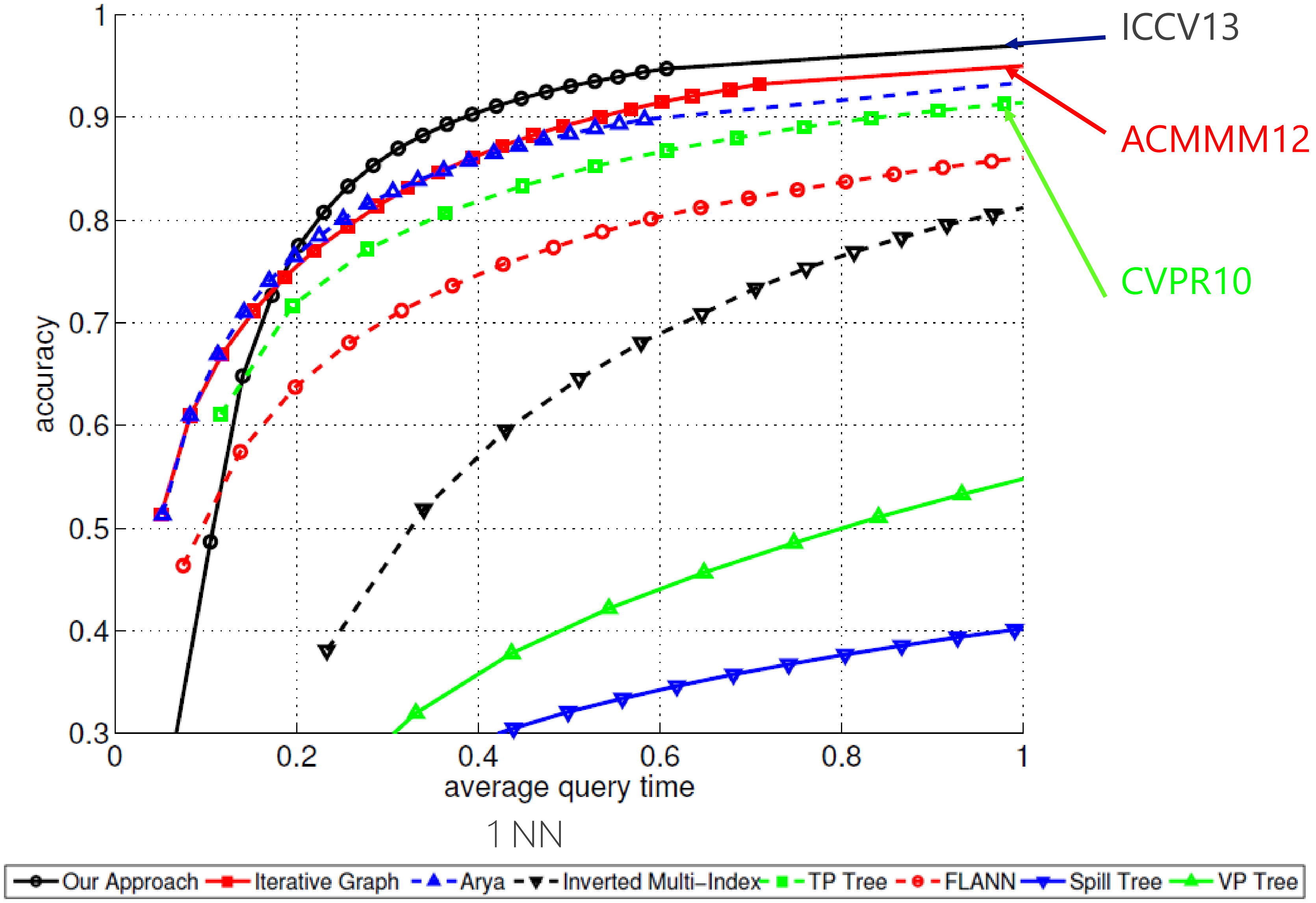


- Iterative local search
  - Iterative query-driven new starting point generation when the local search cannot find better points
  - Generate new starting points using kd-trees (ACMMM12), product quantization (ICCV13)



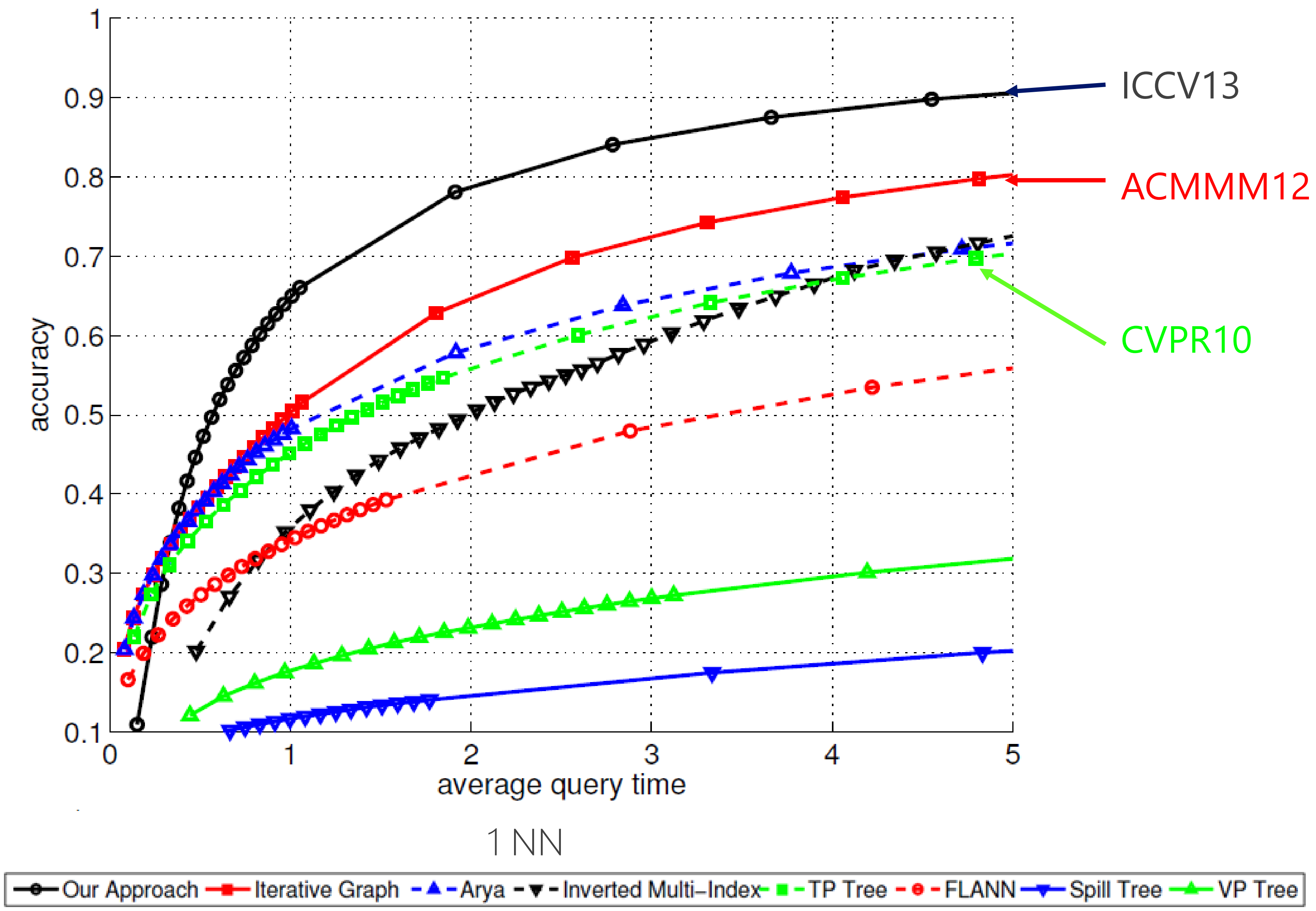
[1] Query-driven iterated neighborhood graph search for large scale indexing. Jingdong Wang, Shipeng Li. ACMNMM 2012
[2] Fast Neighborhood Graph Search Using Cartesian Concatenation. Jing Wang, Jingdong Wang, Gang Zeng, Rui Gan, Shipeng Li, Baining Guo. ICCV 2013.

SIFT 1M



ICCV13
ACMMM12
CVPR10
accuracy
average query time
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
0.2
0.4
0.6
0.8
1
1 NN
Our Approach
Iterative Graph
Arya
Inverted Multi–Index
TP Tree
FLANN
Spill Tree
VP Tree

GIST 1M



ICCV13
ACMMM12
CVPR10
accuracy
average query time
0
1
2
3
4
5
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Our Approach
Iterative Graph
Arya
Inverted Multi–Index
TP Tree
FLANN
Spill Tree
VP Tree

1 NN

Evaluation on 20M 100D CNN Features

Algorithm	Query time (ms)	Precision@10
Hamming distance + ITQ	83	0.93
FLANN (OpenCV)	46	0.93
FLANN (ours)	29	0.93
Composite NN Search	13.5	0.93
Neighborhood Graph Search	**7.4**	**0.95**

20M 100D float in database, 75K queries

[1] Trinary-Projection Trees for Approximate Nearest Neighbor Search. Jingdong Wang, Naiyan Wang, You Jia, Jian Li, Gang Zeng, Hongbin Zha, Xian-Sheng Hua. TPAMI 2014.
[2] Fast Neighborhood Graph Search Using Cartesian Concatenation. Jing Wang, Jingdong Wang, Gang Zeng, Rui Gan, Shipeng Li and Baining Guo. ICCV13.
[3] Query-driven iterated neighborhood graph search for large scale indexing. Jingdong Wang, and Shipeng Li. ACM Multimedia 2012
[4] Scalable k-NN graph construction for visual descriptors. Jing Wang, Jingdong Wang, Gang Zeng, Zhuowen Tu, Rui Gan, Shipeng Li. CVPR 2012.
[5] Zhansheng Jiang, Lingxi Xie, Xiaotie Deng, Weiwei Xu, Jingdong Wang: Fast Nearest Neighbor Search in the Hamming Space. MMM 2016.

Summary

- Neighborhood graph is better than KD-tree, Hierarchical k-means tree
- Memory cost is large

- Construction
  - Trinary-Projection Trees for Approximate Nearest Neighbor Search. Jingdong Wang, Naiyan Wang, You Jia, Jian Li, Gang Zeng, Hongbin Zha, Xian-Sheng Hua. TPAMI 2014.
  - Scalable k-NN graph construction for visual descriptors. Jing Wang, Jingdong Wang, Gang Zeng, Zhuowen Tu, Rui Gan, Shipeng Li. CVPR 2012.
  - Improvement by *relative neighborhood graph*

Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Composite Quantization for Approximate Nearest Neighbor Search. Ting Zhang, Chao Du, Jingdong Wang. ICML 2014, TPAMI 2018

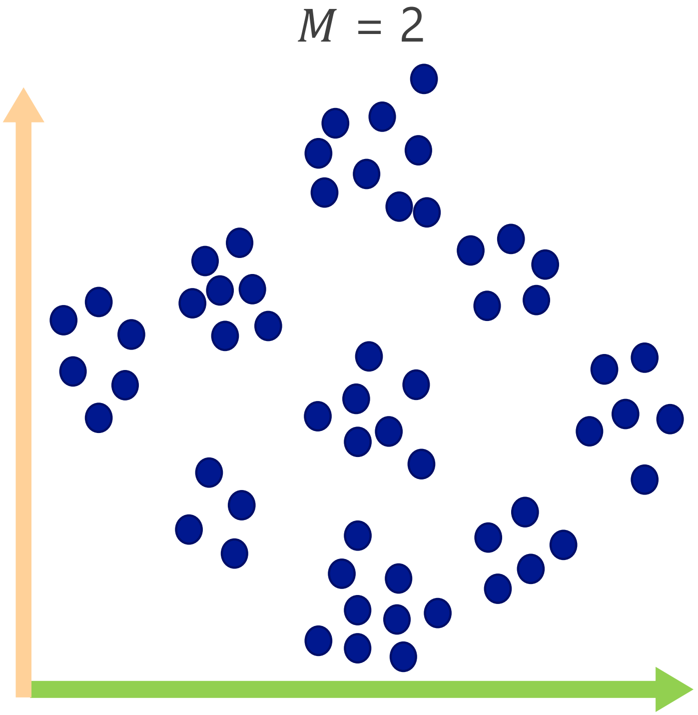
Product quantization

- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Code presentation: $(i_1, i_2, \ldots, i_M)$
- Distance computation:
  - $d(\mathrm{q}, \bar{\mathrm{x}})^2 = d\left(\mathrm{q}_1, \mathrm{p}_{1i_1}\right)^2 + d\left(\mathrm{q}_2, \mathrm{p}_{2i_2}\right)^2 + \cdots + d\left(\mathrm{q}_M, \mathrm{p}_{Mi_M}\right)^2$
  - M additions using a pre-computed distance table

$$
\mathrm{x} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_M \end{bmatrix}
\qquad
\begin{array}{llll}
\mathrm{p}_{1i_1} & \longleftarrow & d(\{\mathrm{p}_{11}, \mathrm{p}_{12}, \cdots, \mathrm{p}_{1K}\}, \mathrm{q}_1) & \triangleq \{d(\mathrm{q}_1, \mathrm{p}_{11}), d(\mathrm{q}_1, \mathrm{p}_{12}), \cdots, d(\mathrm{q}_1, \mathrm{p}_{1K})\} \\
\mathrm{p}_{2i_2} & \longleftarrow & d(\{\mathrm{p}_{21}, \mathrm{p}_{22}, \cdots, \mathrm{p}_{2K}\}, \mathrm{q}_2) & \\
\vdots & & & \\
\mathrm{p}_{Mi_M} & \longleftarrow & d(\{\mathrm{p}_{M1}, \mathrm{p}_{M2}, \cdots, \mathrm{p}_{MK}\}, \mathrm{q}_M) &
\end{array}
$$

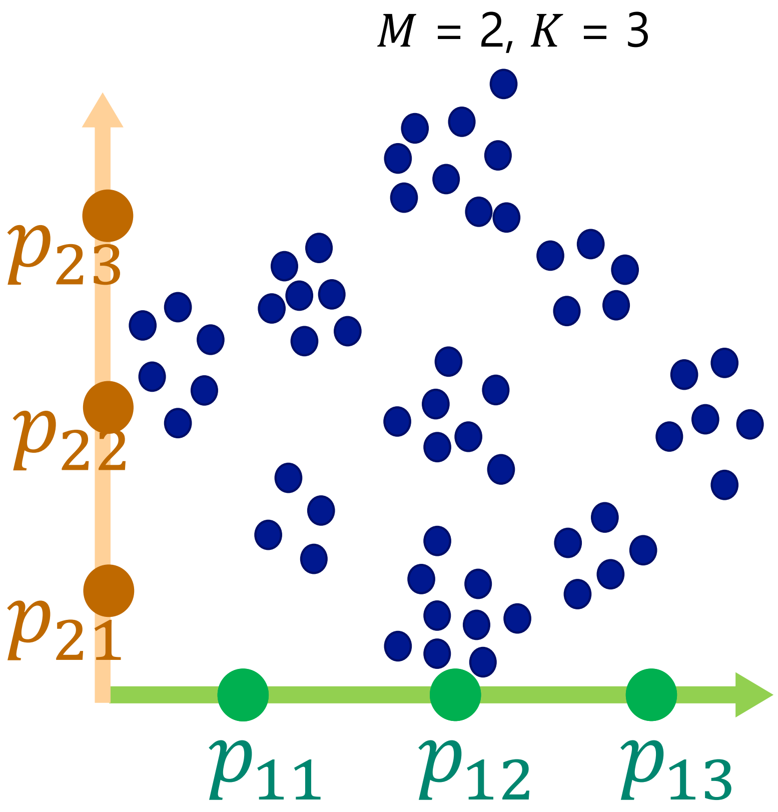
Product quantization

- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Codebook generation
  - Do k-means for each subspace



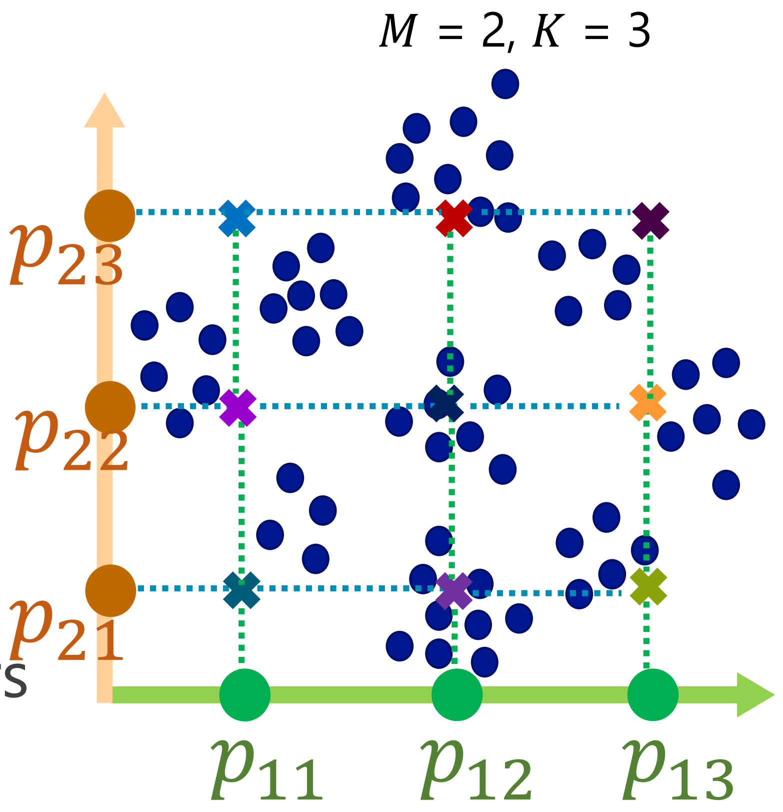
Product quantization

- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Codebook generation
  - Do k-means for each subspace



Product quantization

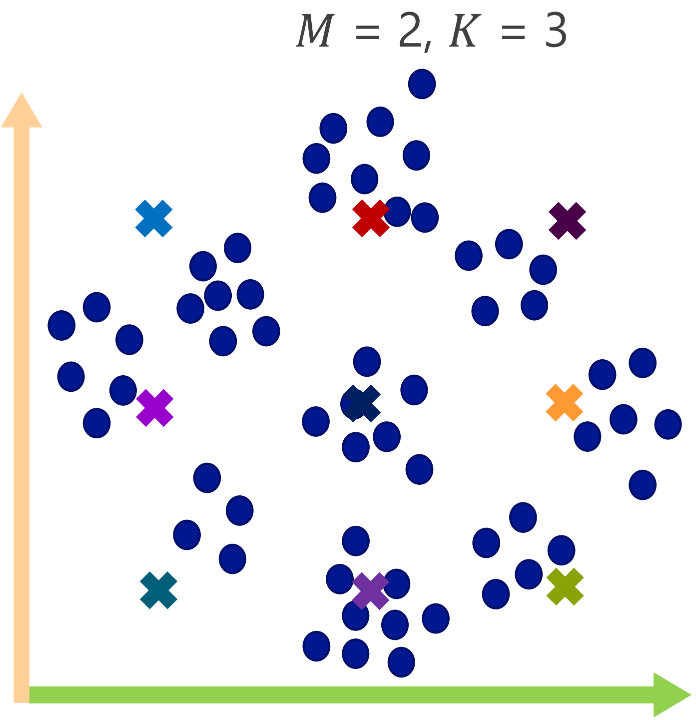
- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Codebook generation
  - Do k-means for each subspace
- Result in K^M groups
  - The center of each is the concatenation of M subvectors



$$\times = \begin{bmatrix} p_{11} \\ p_{21} \end{bmatrix}$$

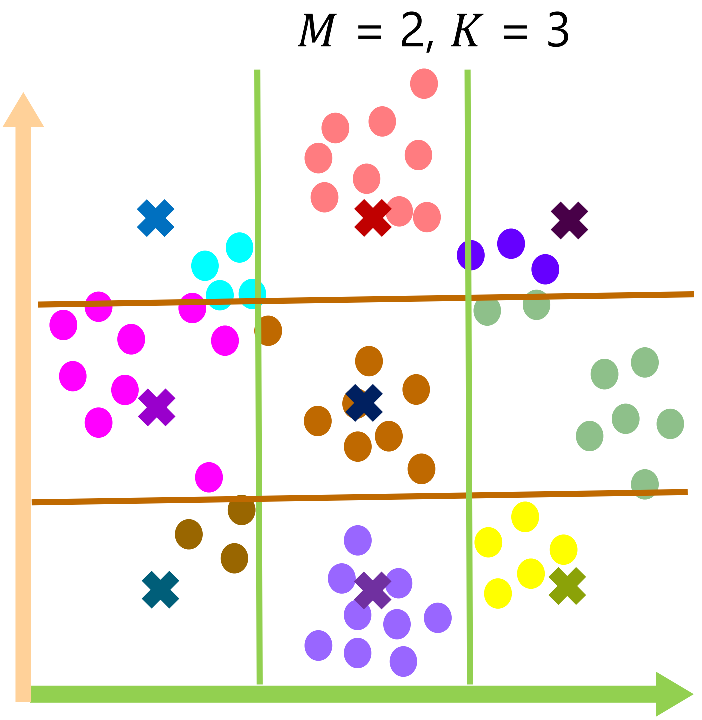
Product quantization

- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Codebook generation
  - Do k-means for each subspace
- Result in K^M groups
  - The center of each is the concatenation of M subvectors



Product quantization

- Approximate $\mathbf{x}$ by the concatenation of M subvectors
- Codebook generation
  - Do k-means for each subspace
- Result in K^M groups
  - The center of each is the concatenation of M subvectors



Cartesian K-means

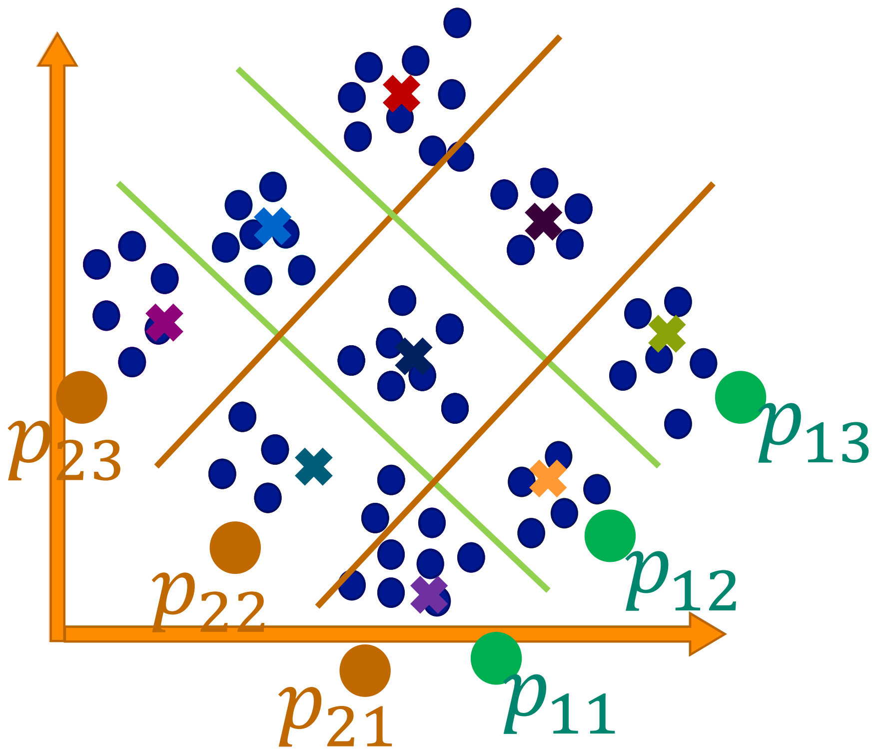
- Extended product quantization
  - Optimal space rotation
  - Perform PQ over the rotated space

$$\mathrm{x} \approx \bar{\mathrm{x}} = \mathrm{R} \begin{bmatrix} \mathrm{p}_{1i_1} \\ \mathrm{p}_{2i_2} \\ \vdots \\ \mathrm{p}_{Mi_M} \end{bmatrix}$$

Cartesian K-means

- Extended product quantization
  - Optimal space rotation
  - Perform PQ over the rotated space

$$\mathrm{x} \approx \bar{\mathrm{x}} = \color{magenta}{\mathrm{R}} \begin{bmatrix} \mathrm{p}_{1i_1} \\ \mathrm{p}_{2i_2} \\ \vdots \\ \mathrm{p}_{Mi_M} \end{bmatrix}$$



Composite quantization (ICML 2014)

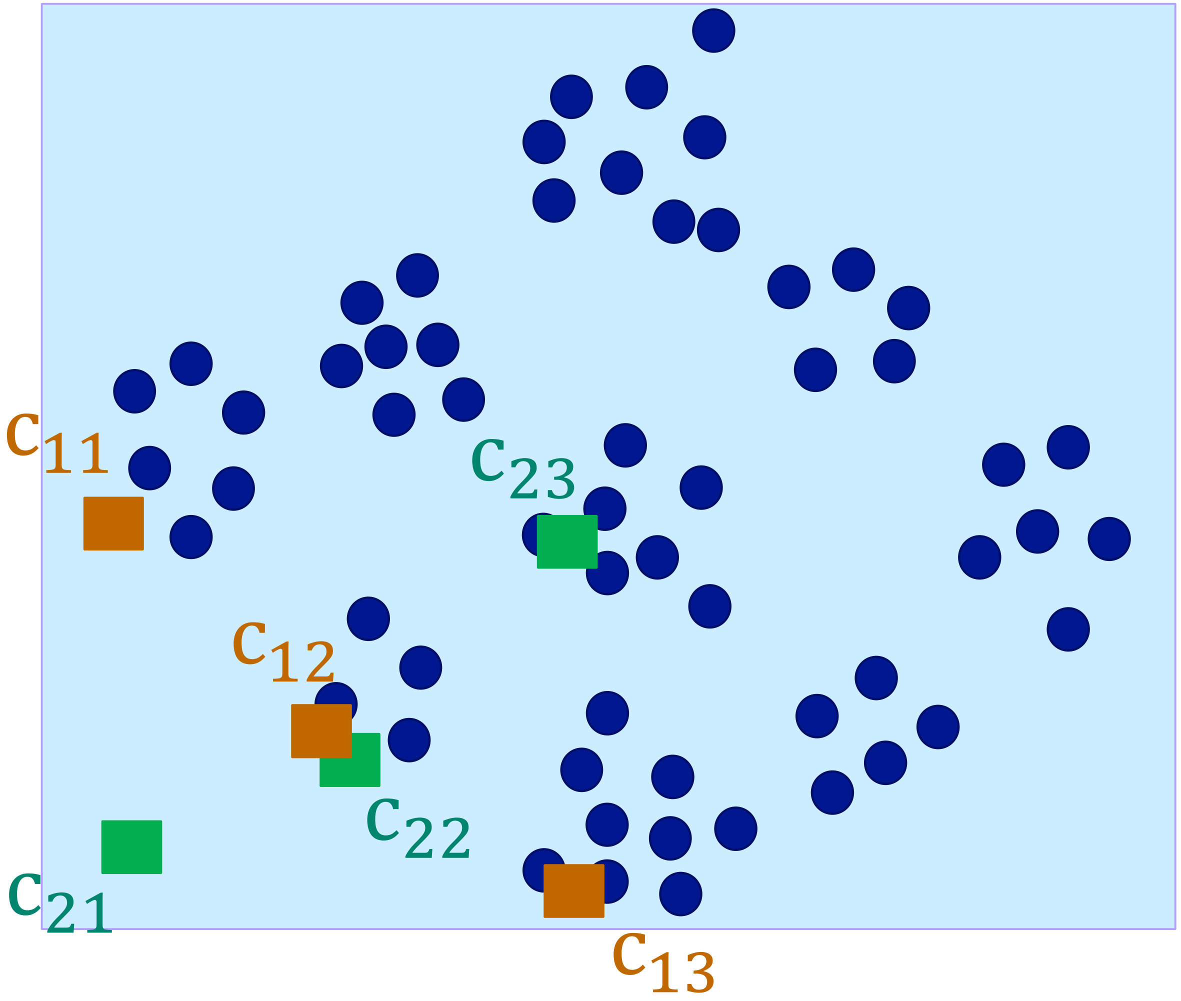
- Approximate $\mathbf{x}$ by the addition of M vectors
- Code representation: $i_1 i_2 \cdots i_M$
  - length: $M\log K$

$$\mathrm{x} \approx \bar{\mathrm{x}} = \mathrm{c}_{1i_1} + \mathrm{c}_{2i_2} + \cdots + \mathrm{c}_{Mi_M}$$

$\{\mathrm{c}_{11}, \mathrm{c}_{12}, \cdots, \mathrm{c}_{1K}\}$ $\{\mathrm{c}_{21}, \mathrm{c}_{22}, \cdots, \mathrm{c}_{2K}\}$ $\cdots$ $\{\mathrm{c}_{M1}, \mathrm{c}_{M2}, \cdots, \mathrm{c}_{MK}\}$

Source codebook 1 Source codebook 2 Source codebook M

Composite quantization

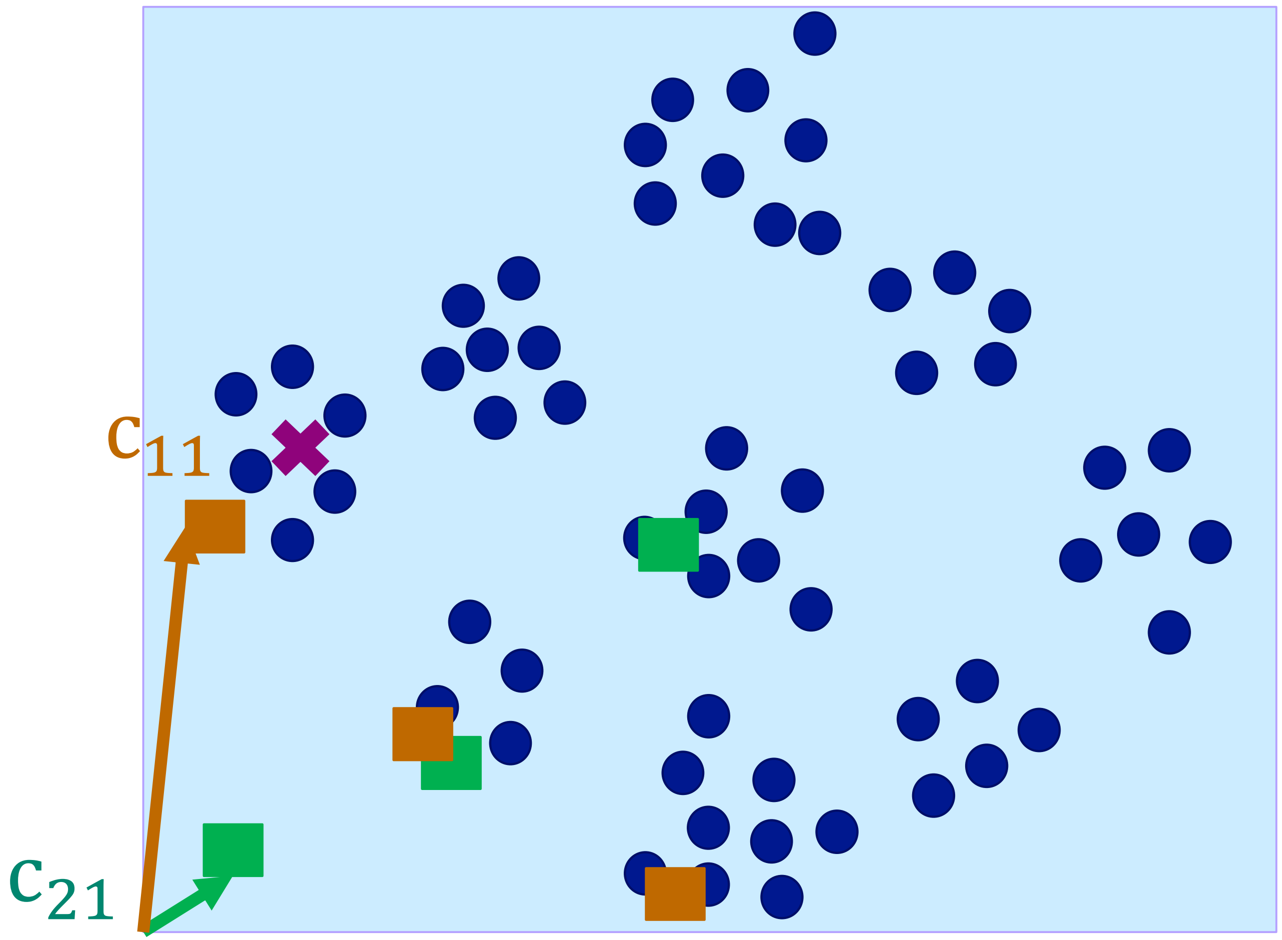


2 source codebooks:

$\{c_{11}, c_{12}, c_{13}\}$

$\{c_{21}, c_{22}, c_{23}\}$

Composite quantization



2 source codebooks:

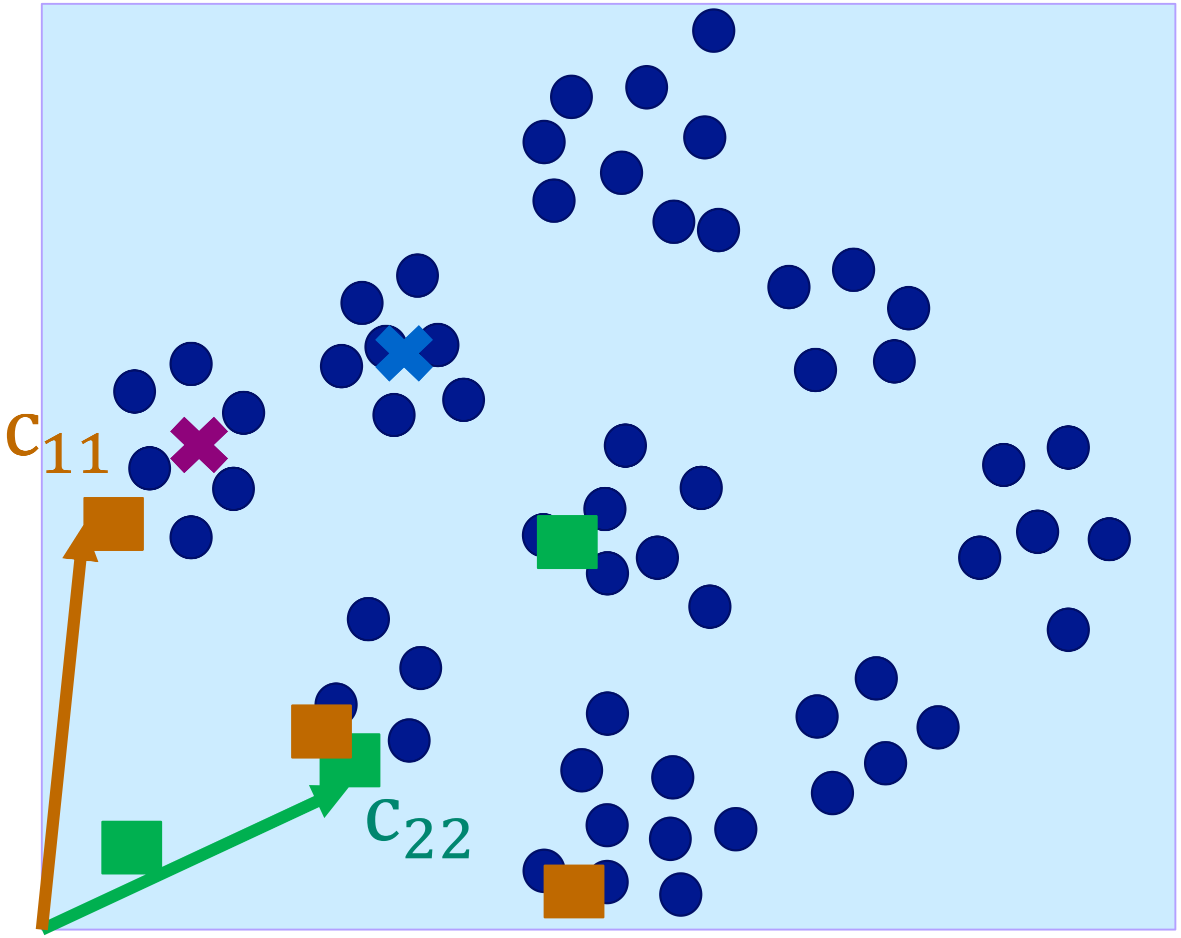
$\{c_{11}, c_{12}, c_{13}\}$

$\{c_{21}, c_{22}, c_{23}\}$

Composite center:

✖ $= c_{11} + c_{21}$

Composite quantization



2 source codebooks:

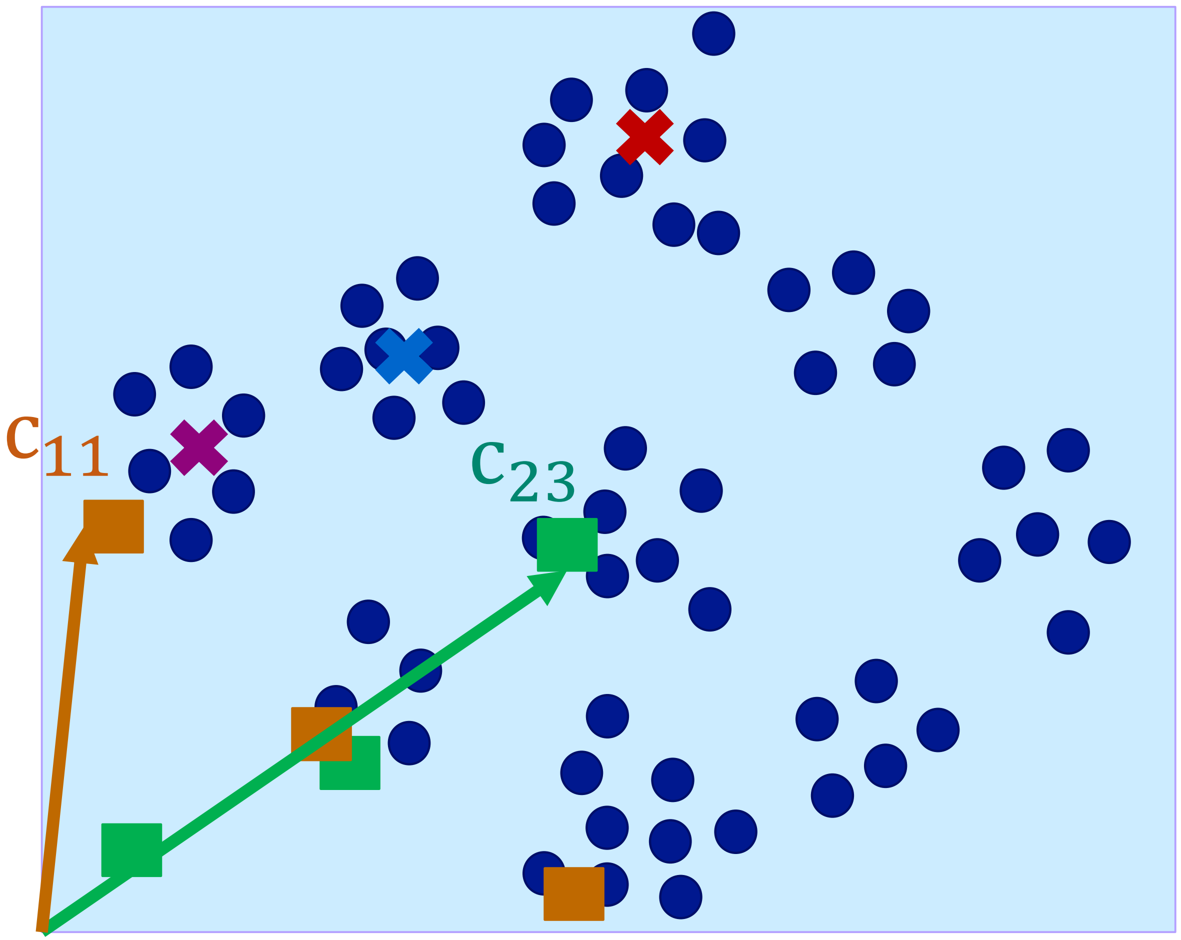
$\{c_{11}, c_{12}, c_{13}\}$

$\{c_{21}, c_{22}, c_{23}\}$

Composite center:

✖ $= c_{11} + c_{22}$

Composite quantization



2 source codebooks:

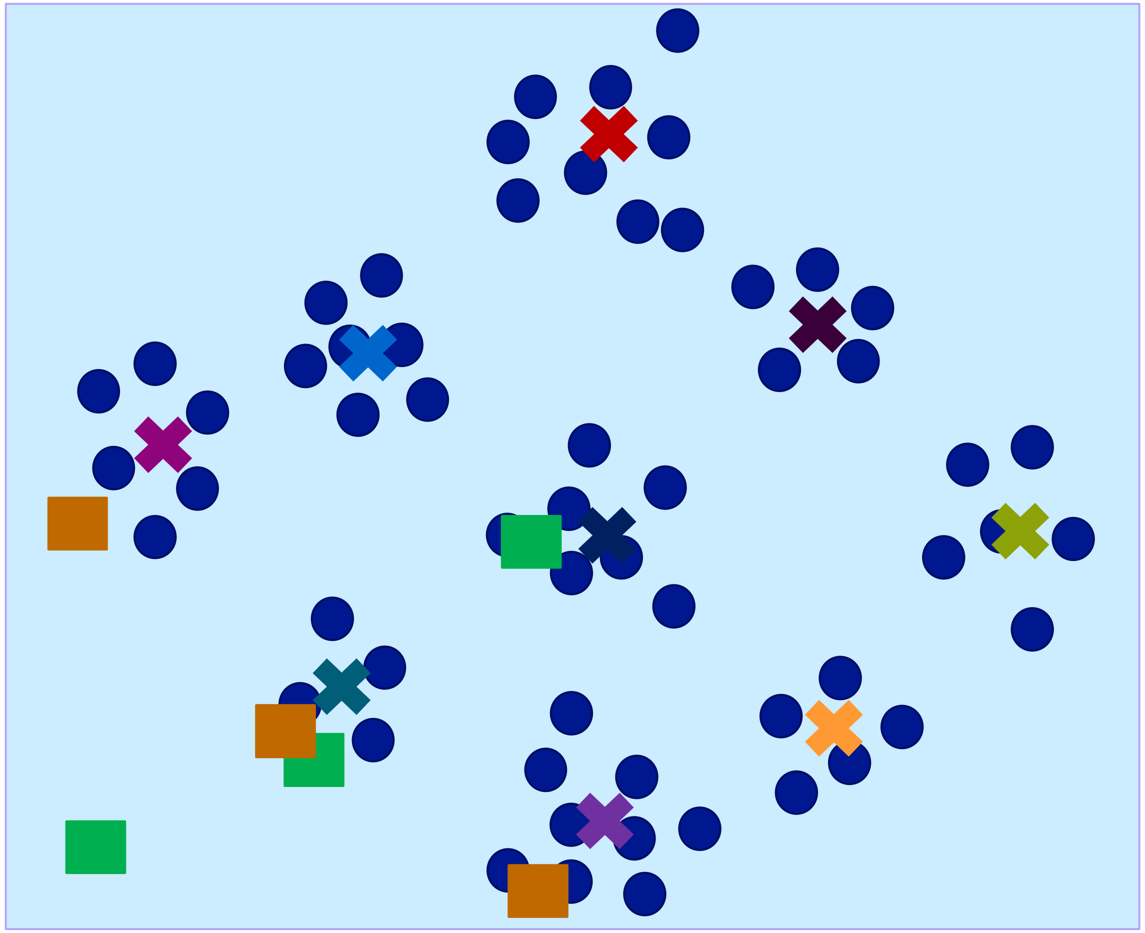
$\{c_{11}, c_{12}, c_{13}\}$

$\{c_{21}, c_{22}, c_{23}\}$

Composite center:

$= c_{11} + c_{23}$

Composite quantization



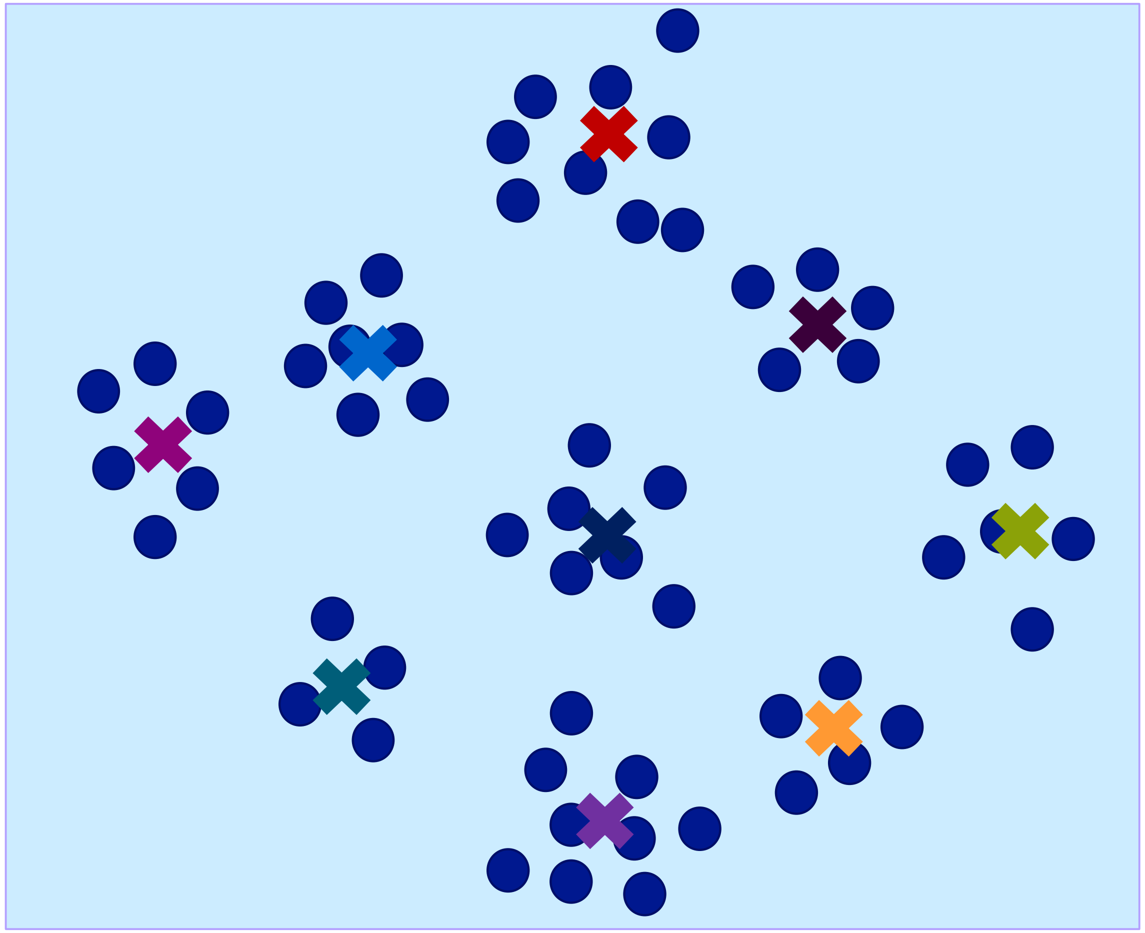
2 source codebooks:

$\{\mathbf{c}_{11}, \mathbf{c}_{12}, \mathbf{c}_{13}\}$

$\{\mathbf{c}_{21}, \mathbf{c}_{22}, \mathbf{c}_{23}\}$

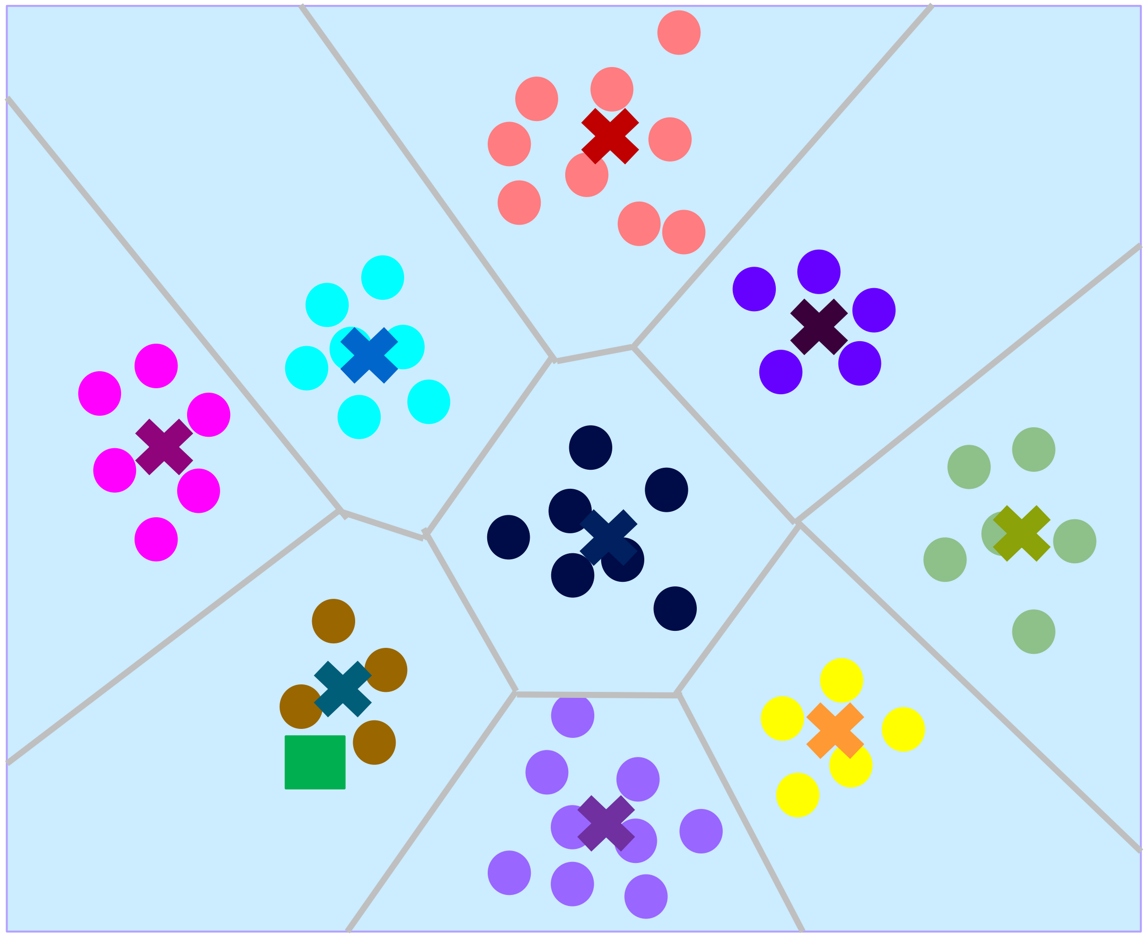
More composite centers

Composite quantization



Composite codebook:
9 composite centers

Composite quantization



Source codebook:

$\{c_{11}, c_{12}, c_{13}\}$

$\{c_{21}, c_{22}, c_{23}\}$

Space partition:

9 groups

Approximate Distance Computation

Approximate distance:

$$\|\mathrm{q}-\mathrm{x}\|_2^2 \approx \boxed{\left\|\mathrm{q}-\sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})}\right\|_2^2}$$

Time-consuming

Fast Approximate Distance Computation

$$\left\|\mathrm{q}-\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2$$

$$=\sum_{m=1}^{M}\left\|\mathrm{q}-\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2-(M-1)\|\mathrm{q}\|_2^2+\sum_{m\neq l}\mathrm{c}_{mi_m(\mathrm{x})}^T\mathrm{c}_{li_l(\mathrm{x})}$$

Fast Approximate Distance Computation

$$\left\| \mathrm{q} - \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2$$

$$= \underbrace{\sum_{m=1}^{M} \left\| \mathrm{q} - \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2}_{O(M) \text{ additions}} - \underbrace{\cancel{(M-1)\|\mathrm{q}\|_2^2}}_{\text{Constant}} + \cancel{\sum_{m \neq l} \mathrm{c}_{m i_m(\mathrm{x})}^{T} \mathrm{c}_{l i_l(\mathrm{x})}}$$

If constant

Computing this is
enough for search

Fast Approximate Distance Computation

$$\left\|\mathbf{q} - \sum_{m=1}^{M} \mathbf{c}_{mi_m(\mathbf{x})}\right\|_2^2$$

$$= \underbrace{\sum_{m=1}^{M} \left\|\mathbf{q} - \mathbf{c}_{mi_m(\mathbf{x})}\right\|_2^2}_{} - \cancel{(M-1)\|\mathbf{q}\|_2^2} + \cancel{\sum_{m \neq l} \mathbf{c}_{mi_m(\mathbf{x})}^T \mathbf{c}_{li_l(\mathbf{x})}}$$

Constant

Minimize quantization error:
$\left\|\mathbf{x} - \sum_{m=1}^{M} \mathbf{c}_{mi_m(\mathbf{x})}\right\|_2^2$

Subject to
the third term is a constant

Near-orthogonal composite quantization (NOCQ)

Formulation

- Constrained formulation

$$\min_{\{\mathrm{c}_m\},\{i_m(\mathrm{x})\},\epsilon} \sum_{\mathrm{x}} \left\| \mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2$$

$$\mathrm{s.t.} \quad \underline{\sum_{m \neq l} \mathrm{c}_{m i_m(\mathrm{x})}^{T} \mathrm{c}_{l i_l(\mathrm{x})} = \epsilon}$$

Minimize quantization error for <u>search accuracy</u>

Constant constraint for <u>search efficiency</u>

Connection

- Constrained formulation

$$\min_{\{\mathrm{c}_m\},\{i_m(\mathrm{x})\},\epsilon} \quad \sum_{\mathrm{x}} \left\| \mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2$$

$$\text{s.t.} \quad \sum_{m \neq l} \mathrm{c}_{m i_m(\mathrm{x})}^T \mathrm{c}_{l i_l(\mathrm{x})} = \epsilon$$

Minimize quantization error for <u>search accuracy</u>

Constant constraint for <u>search efficiency</u>

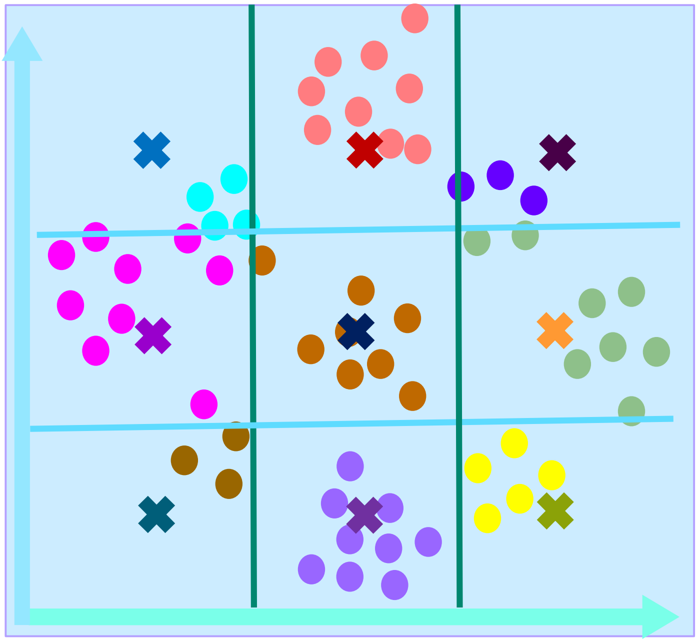
- Product quantization and Cartesian k-means: suboptimal solutions of our approach (NOCQ)

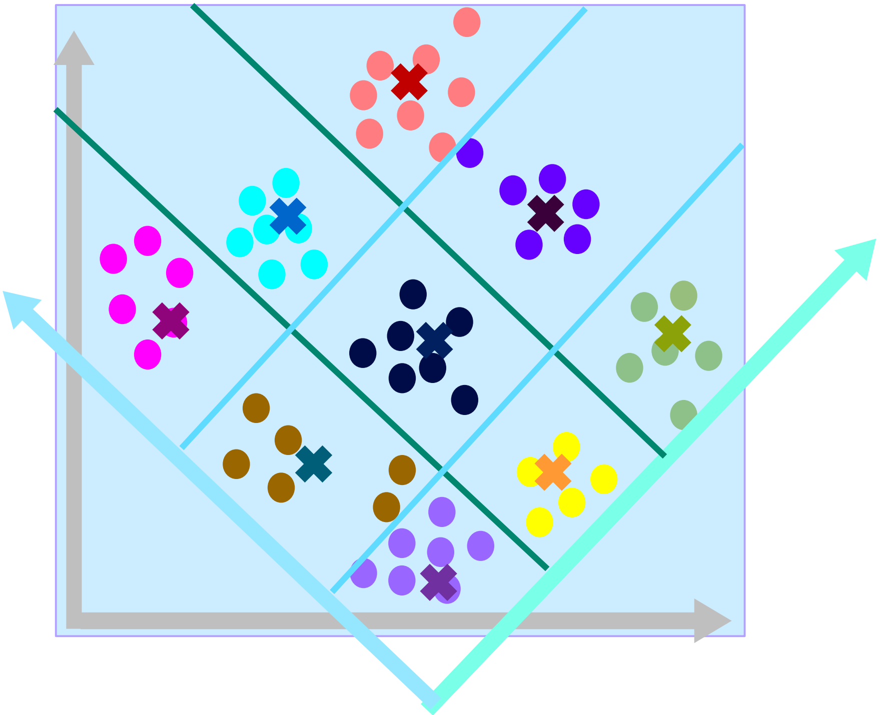
Non-overlapped space partitioning → Codebooks are mutually orthogonal → $\sum_{m \neq l} \mathrm{c}_{m i_m(\mathrm{x})}^T \mathrm{c}_{l i_l(\mathrm{x})} = \epsilon$

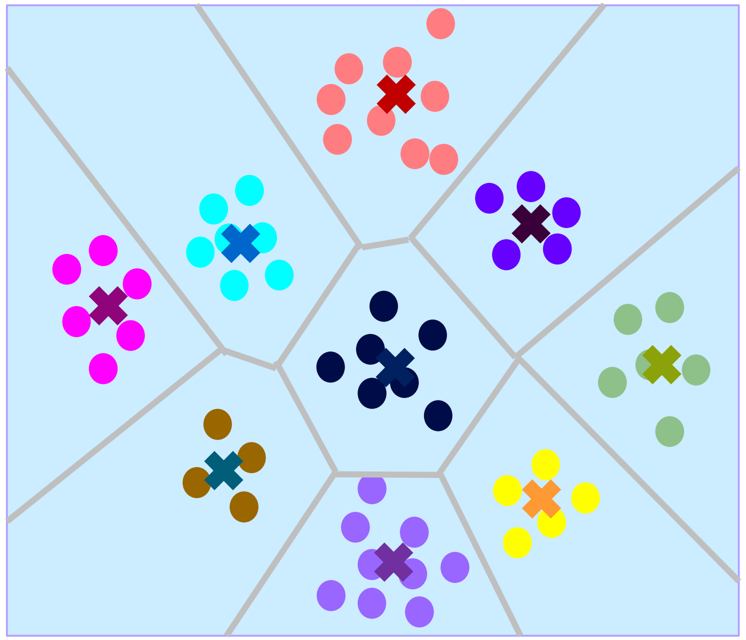
Product quantization and Cartesian k-means

Connection

Near-orthogonal composite quantization generalizes product quantization and Cartesian k-means







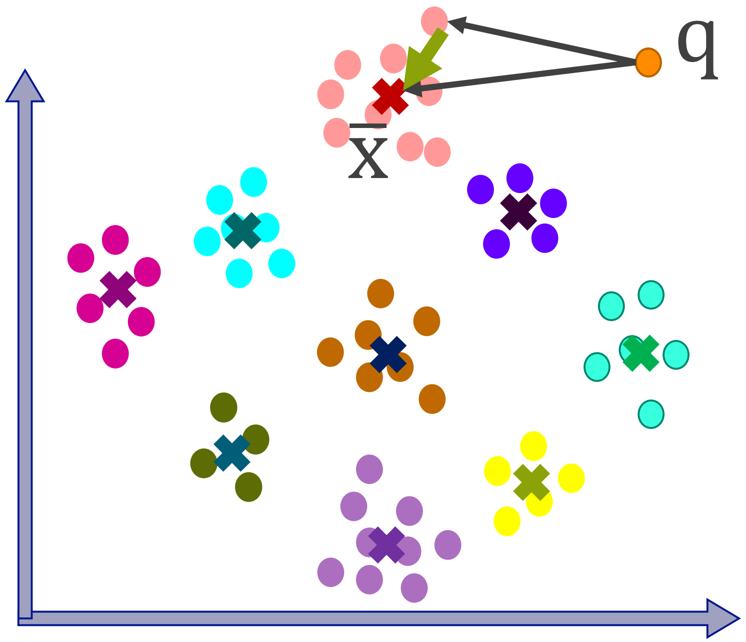
Product quantization:
Coordinate aligned space partition

Cartesian k-means:
Rotated coordinate aligned space partition

Near-orthogonal composite quantization:
Flexible space partition

A Distance Preserving View of Quantization

- Quantization
  - Data approximation: $\bar{x} \approx x$
- Better search
  - If better distance preserving: $\|q - \bar{x}\|_2 \approx \|q - x\|_2$
- Distance preserving view
  - Triangle inequality: $\left|\|q - x\|_2 - \|q - \bar{x}\|_2\right| \le \|x - \bar{x}\|_2$
  - Minimize the upper bound: $\sum_x \|x - \bar{x}\|_2^2$



A Joint Minimization View

Generalized triangle inequality

$$\left|\tilde{d}(\mathrm{q},\bar{\mathrm{x}}) - \hat{d}(\mathrm{q},\mathrm{x})\right| \le \underbrace{\|\mathrm{x}-\bar{\mathrm{x}}\|_2}_{\text{Distortion}} + \underbrace{|\delta|^{1/2}}_{\text{Efficiency}}$$

Triangle inequality

$$\left|\|\mathrm{q}-\mathrm{x}\|_2 - \|\mathrm{q}-\bar{\mathrm{x}}\|_2\right| \le \|\mathrm{x}-\bar{\mathrm{x}}\|_2$$

$$\tilde{d}(\mathrm{q},\bar{\mathrm{x}}) = \left(\Sigma_{m=1}^{M}\left\|\mathrm{q}-\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2\right)^{1/2}$$

$$\hat{d}(\mathrm{q},\mathrm{x}) = \left(\|\mathrm{q}-\mathrm{x}\|_2^2 + (M-1)\|\mathrm{q}\|_2^2\right)^{1/2}$$

$$\delta = \Sigma_{m\neq l}\mathrm{c}_{mi_m(\mathrm{x})}^{T}\mathrm{c}_{li_l(\mathrm{x})}$$

$$\bar{\mathrm{x}} = \Sigma_{m=1}^{M}\mathrm{c}_{mi_m(x)}$$

A Joint Minimization View

Generalized triangle inequality

$$\left|\tilde{d}(\mathrm{q},\bar{\mathrm{x}}) - \hat{d}(\mathrm{q},\mathrm{x})\right| \leq \underbrace{\|\mathrm{x}-\bar{\mathrm{x}}\|_2}_{\text{Distortion}} + \underbrace{|\delta|^{1/2}}_{\text{Efficiency}}$$

Our formulation

$$\min_{\{\mathrm{C}_m\},\{i_m(\mathrm{x})\},\epsilon} \quad \Sigma_{\mathrm{x}}\|\mathrm{x}-\bar{\mathrm{x}}\|_2^2$$

$$s.t. \quad \delta = \epsilon$$

Minimize distortion for search accuracy

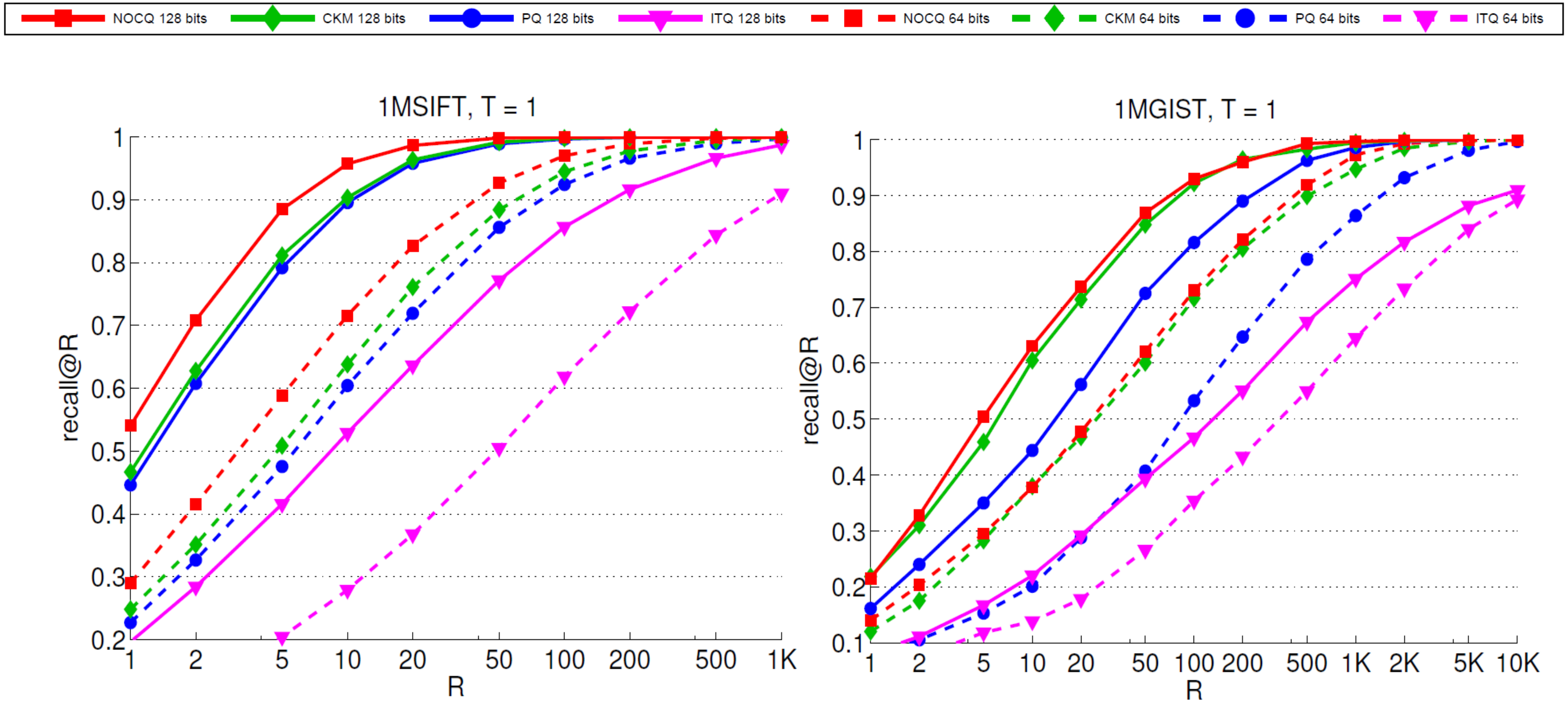
Constant constraint for search efficiency

Experiments

- Datasets
  - 1 million of 128D SIFT vectors, 10000 queries
  - 1 million of 960D GIST vectors, 1000 queries
  - 1 billion of 128D SIFT vectors, 1000 queries

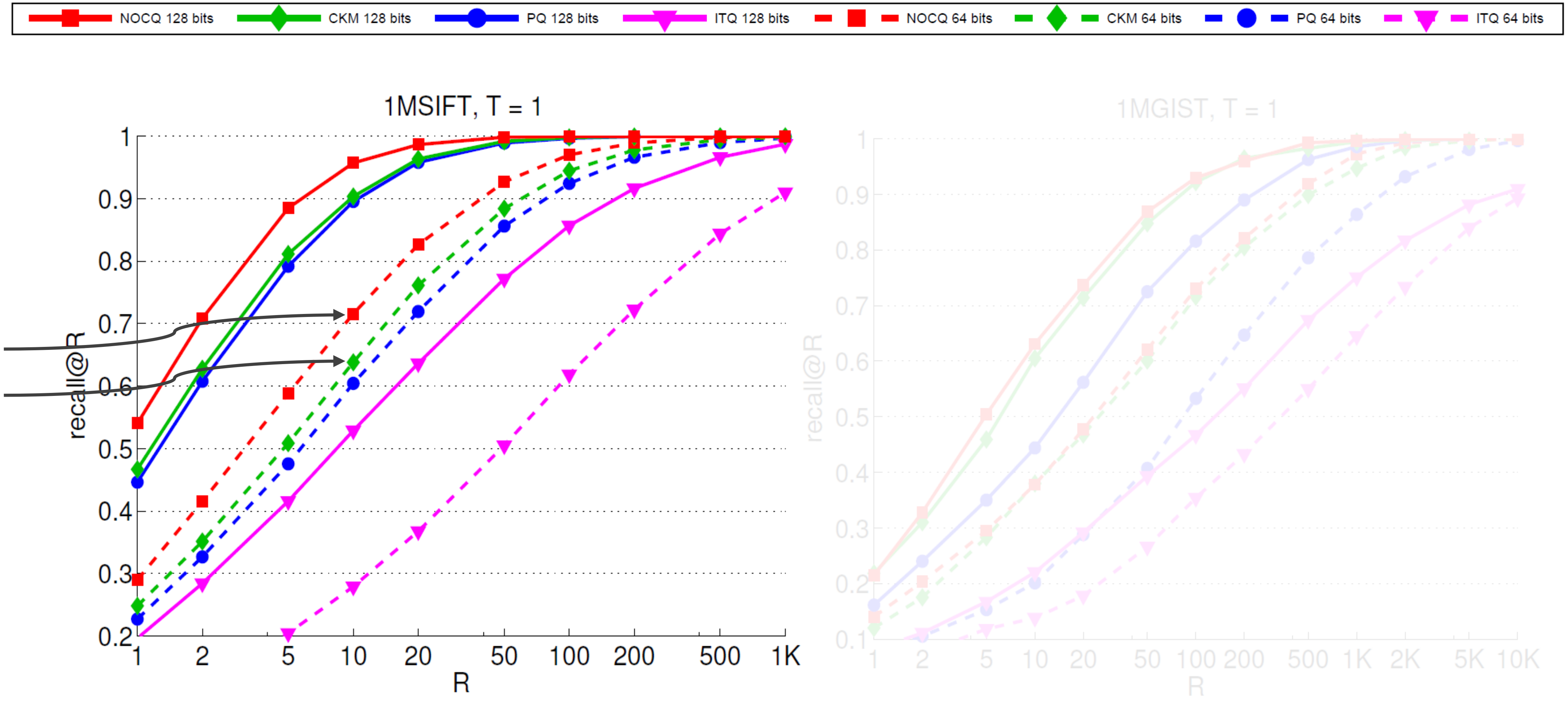
- Evaluation
  - Recall@R
    - the fraction of queries for which the ground-truth Euclidean nearest neighbor is in the R retrieved items

Comparison on 1M SIFT and 1M GIST



Comparison on 1M SIFT and 1M GIST

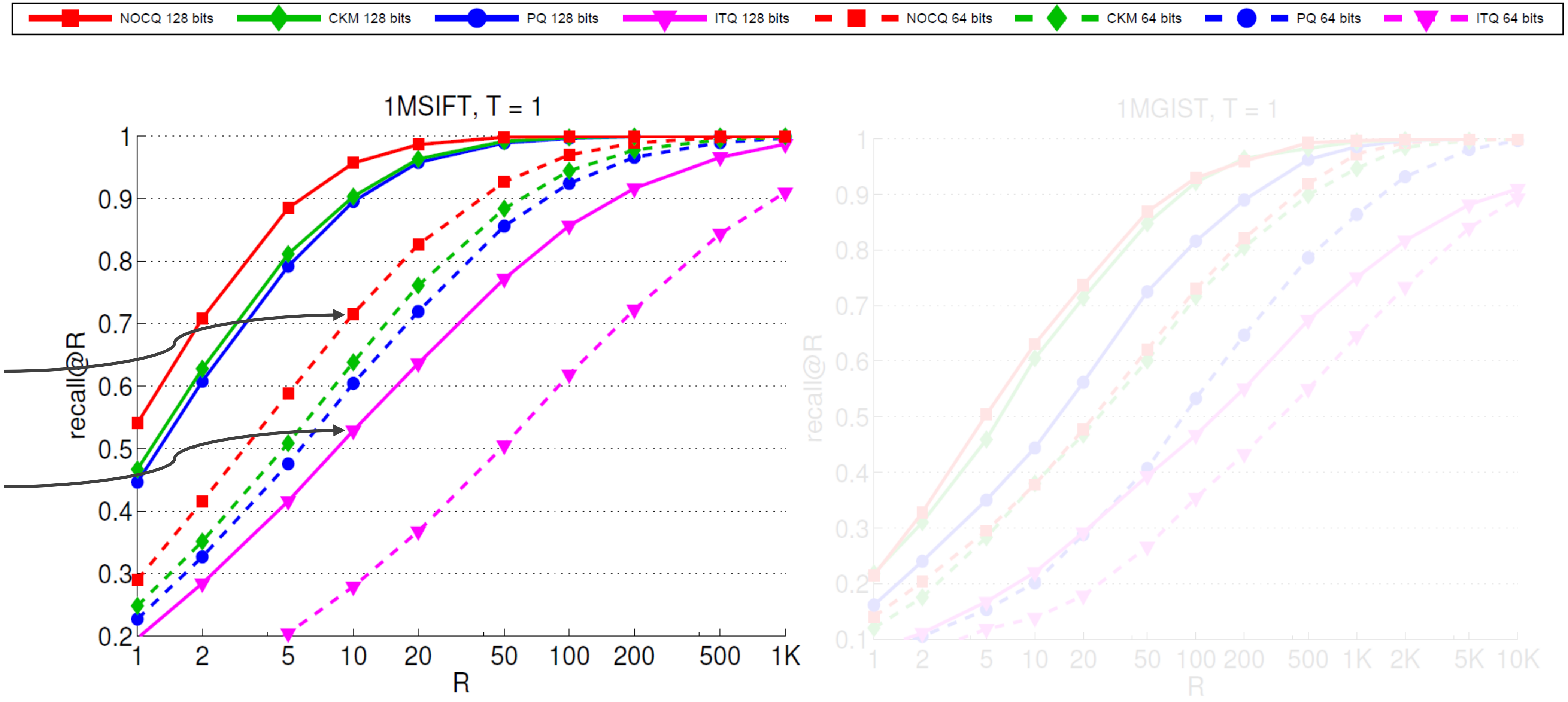
64 btis
Recall@10:
Our:71.59%
CKM:63.83%



Comparison on 1M SIFT and 1M GIST

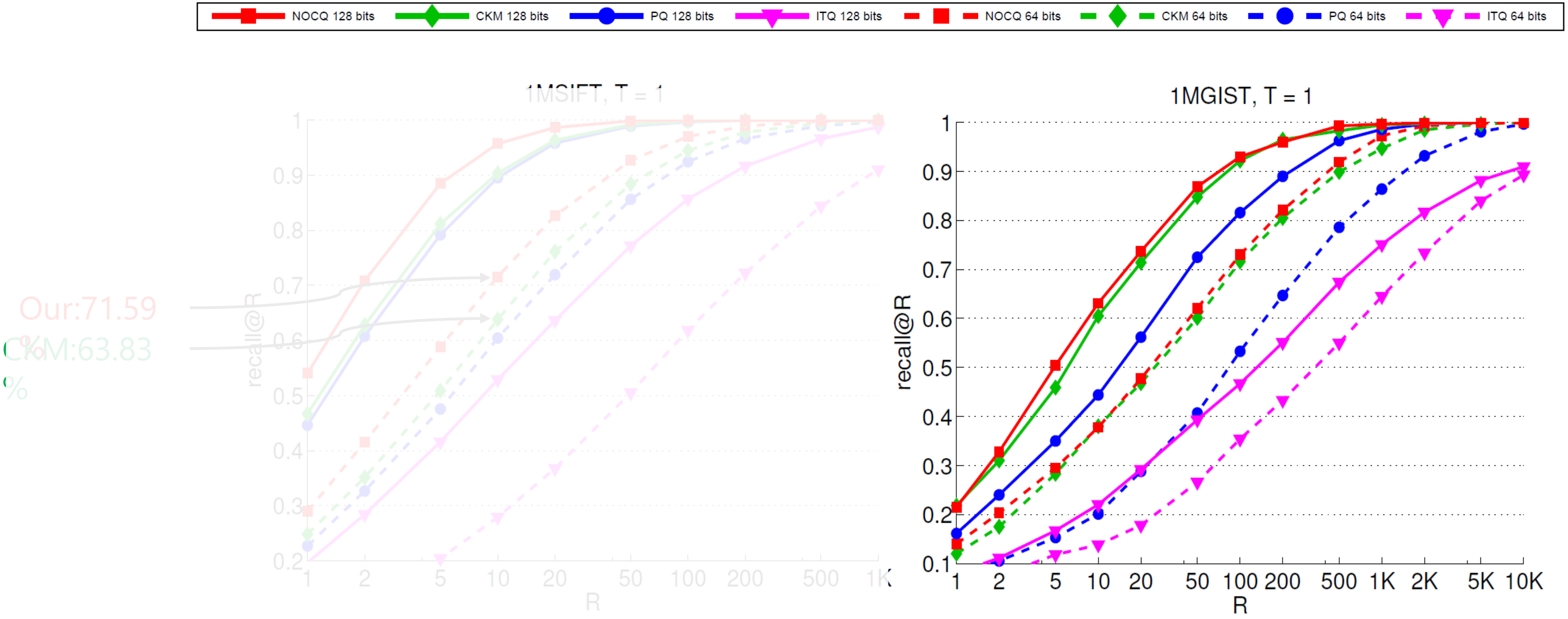
Our:71.59%
64 bits

ITQ: 53.95%
128 bits



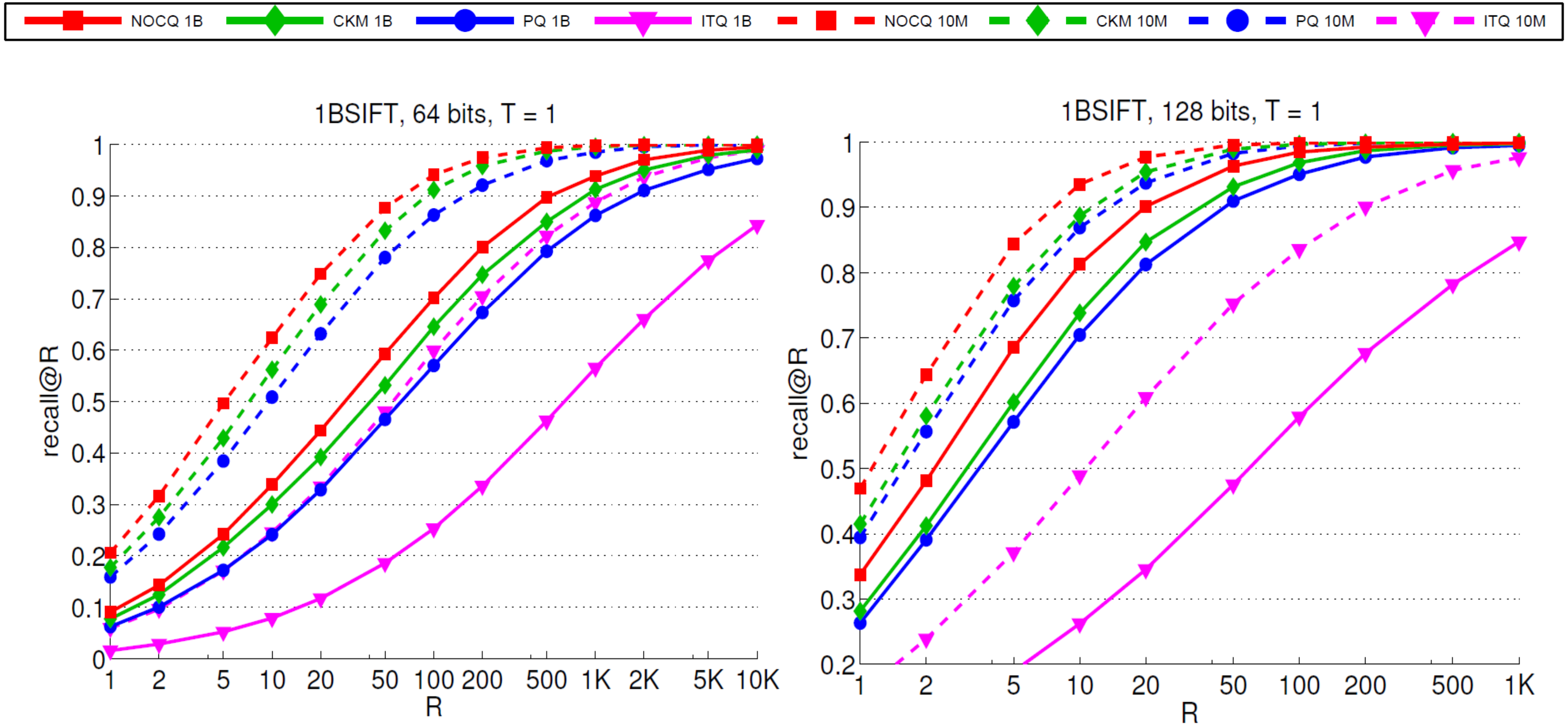
ITQ without asymmetric distance underperformed ITQ with asymmetric distance
Our approach with 64 bits outperforms (A) ITQ with 128 bits, with slightly smaller search cost

Comparison on 1M SIFT and 1M GIST

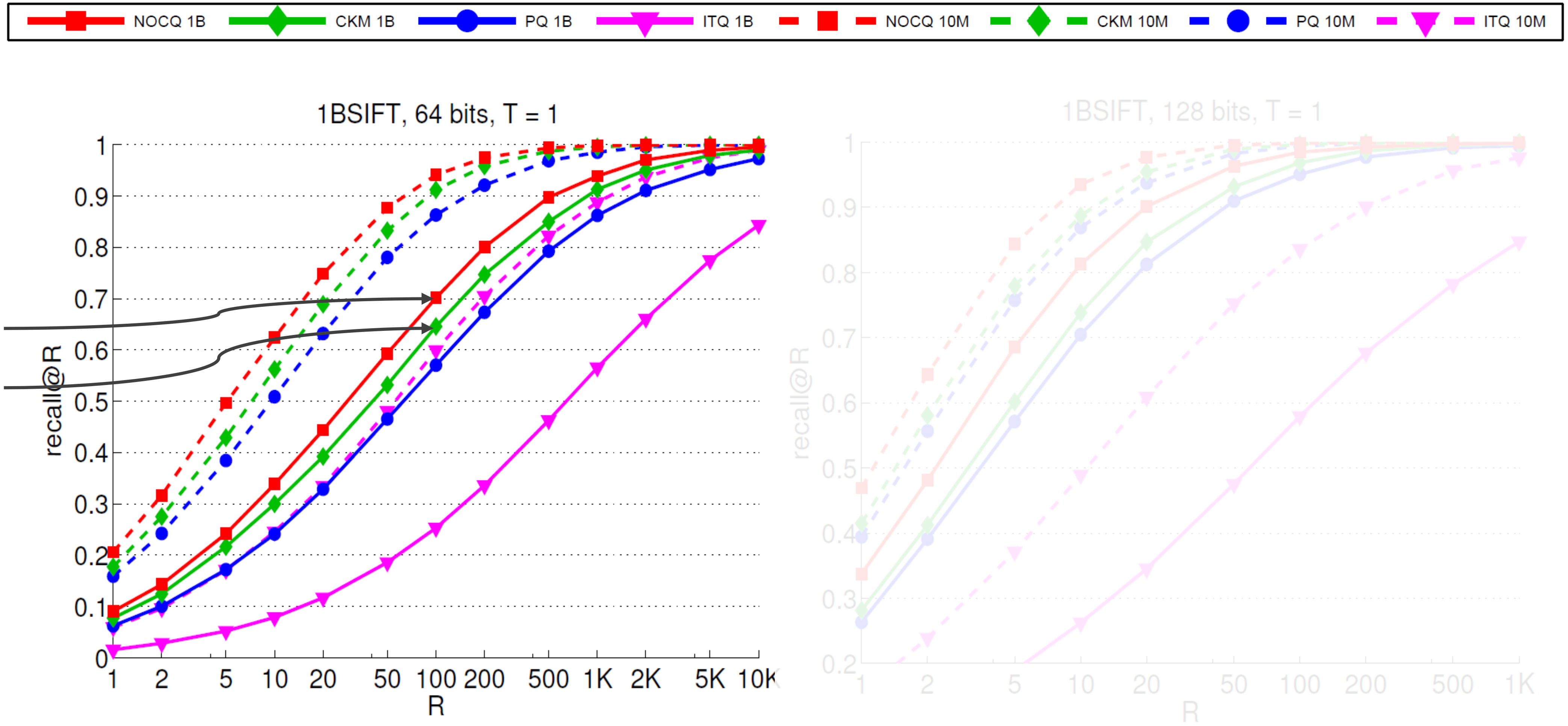


Relatively small improvement on 1M GIST might be that CKM has already achieved large improvement

Comparison on 1B SIFT



Comparison on 1B SIFT

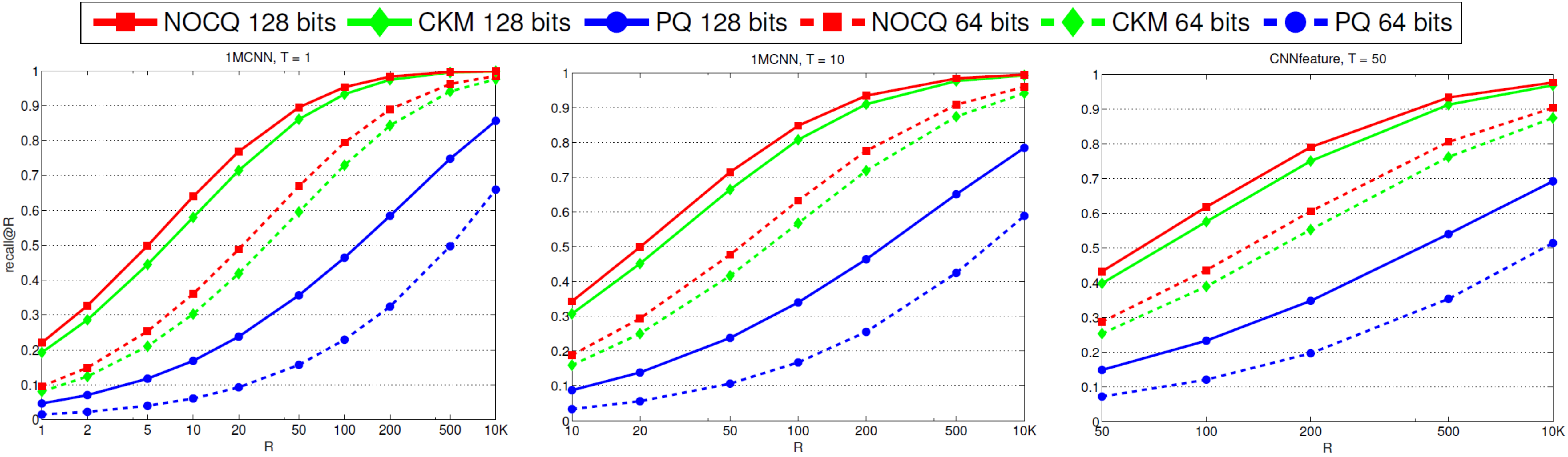


Recall@100:

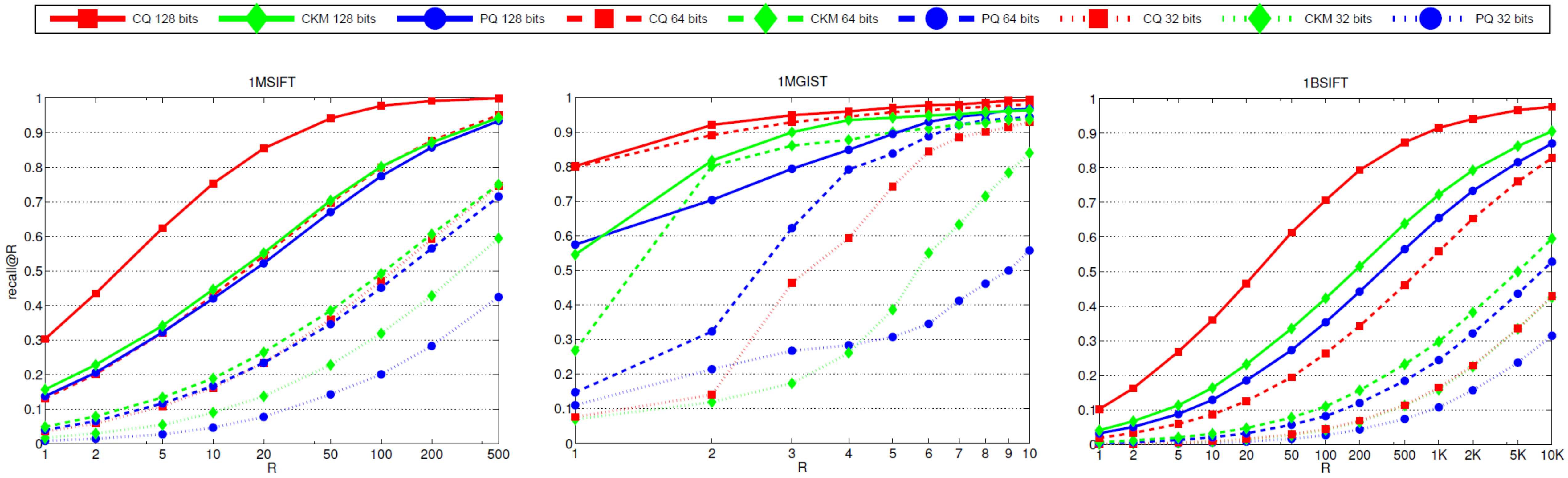
Our:70.12%

CKM:64.57%

Comparison on 1M CNN



Inner product search



Sparse Composite Quantization (CVPR 2015)

- Distance computation between a query and the dictionary element

$$\|\mathrm{q}-\mathrm{c}\|^2 = \|\mathrm{q}\|^2 - 2\mathrm{q}^T\mathrm{c} + \|\mathrm{c}\|^2$$

If known

If c is a sparse vector

- The cost of distance computation: $O(\|\mathrm{c}\|_0)$
- Our idea: sparsifying the dictionary elements!

CQ + Quantize the third term (TPAMI 2018)

$$\left\|\mathrm{q}-\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2$$

$$=\sum_{m=1}^{M}\left\|\mathrm{q}-\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2-(M-1)\|\mathrm{q}\|_2^2+\underline{\sum_{m\neq l}\mathrm{c}_{mi_m(\mathrm{x})}^{T}\mathrm{c}_{li_l(\mathrm{x})}}$$

Encode with 1 byte

- Perform similarly to NOCQ for long codes
- Worse for short codes

Euclidean distance to inner product

$$\|\mathbf{q}-\mathbf{x}\|_2^2 = \|\mathbf{q}\|_2^2 - 2\mathbf{q}^\top\mathbf{x} + \boxed{\|\mathbf{x}\|_2^2}$$

Encode the norm [1]

Worse than NOCQ

$$= \|\mathbf{q}\|_2^2 + [\mathbf{q}^\top \ 1]\boxed{\begin{bmatrix} -2\mathbf{x} \\ \gamma\|\mathbf{x}\|_2^2 \end{bmatrix}}$$

CQ (TPAMI)

Need to tune γ carefully

[1] Artem Babenko, Victor S. Lempitsky: Additive Quantization for Extreme Vector Compression. CVPR 2014: 931-938

Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Supervised quantization for similarity search. Xiaojuan Wang, Ting Zhang, Guo-Jun Qi, Jinhui Tang, Jingdong Wang. CVPR 2016

Semantic similarity search

- Visually similar
  - Unsupervised compact coding
  - Euclidean distance
- Semantically and visually similar
  - Supervised compact coding

Background

- Supervised hashing
  - Pairwise similarity preserving
    - Align the similarity over each pair of items with the semantic similarity
  - Multiwise similarity preserving
    - Maximize the agreement of the similarity order over more than two items
  - Classification
    - Semantically similar points belong to the same class after encoded
- Supervised Quantization
  - Relatively unexplored
  - First attempt to explore supervised quantization

Categorization

Method	Hash/quantization	Pairwise	Multiwise	Classification
LDA hashing	Hash	▲		
Minimal loss hashing	hash	▲		
Binary reconstructive embedding	Hash	▲		
Two-step hashing	Hash	▲		
FastHash	Hash	▲		
Supervised deep hashing	Hash	▲		
Order preserving hashing	Hash		▲	
Triplet loss hashing	Hash		▲	
Listwise supervision hashing	Hash		▲	
Deep semantic ranking based hashing	Hash		▲	
SDH	Hash			▲
Our approach	Quantization			▲

Supervised quantization

- Perform composite quantization in a discriminative space learned by a transformation matrix P
  - $\left\|\mathrm{P}^T\mathrm{x} - \bar{\mathrm{x}}\right\|_2^2 = \left\|\mathrm{P}^T\mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2$

Supervised quantization

- Perform composite quantization in a discriminative space learned by a transformation matrix P
  - $\left\|\mathrm{P}^T\mathrm{x} - \bar{\mathrm{x}}\right\|_2^2 = \left\|\mathrm{P}^T\mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2$
- The encoded points belonging to the same class lie in a cluster

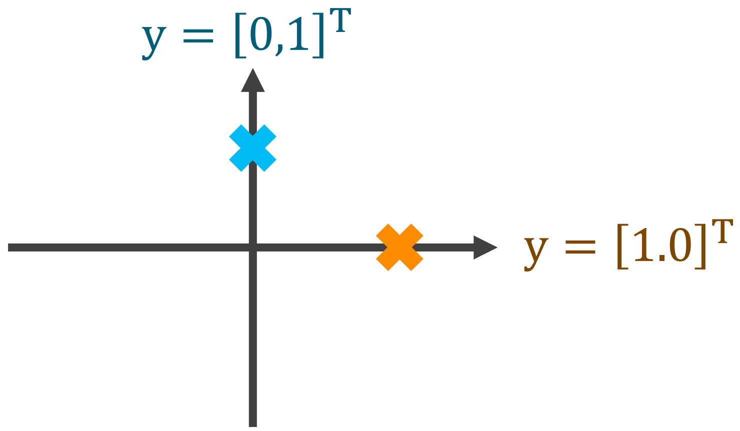
Supervised quantization

- Perform composite quantization in a discriminative space learned by a transformation matrix P
  - $\left\| \mathrm{P}^T \mathrm{x} - \bar{\mathrm{x}} \right\|_2^2 = \left\| \mathrm{P}^T \mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2$
- The encoded points belonging to the same class lie in a cluster
  - The center of cluster is defined by the label vector $\mathrm{y} \in \{0,1\}^C$
  - $\left\| \mathrm{y} - \mathrm{W}^T \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2 + \lambda \| \mathrm{W} \|_F^2$

Supervised quantization

- Perform composite quantization in a discriminative space learned by a transformation matrix P
  - $\left\|\mathrm{P}^T\mathrm{x} - \bar{\mathrm{x}}\right\|_2^2 = \left\|\mathrm{P}^T\mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2$
- The encoded points belonging to the same class lie in a cluster
  - The center of cluster is defined by the label vector $\mathrm{y} \in \{0,1\}^C$
  - $\left\|\mathrm{y} - \mathrm{W}^T \sum_{m=1}^{M} \mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2 + \lambda\|\mathrm{W}\|_F^2$

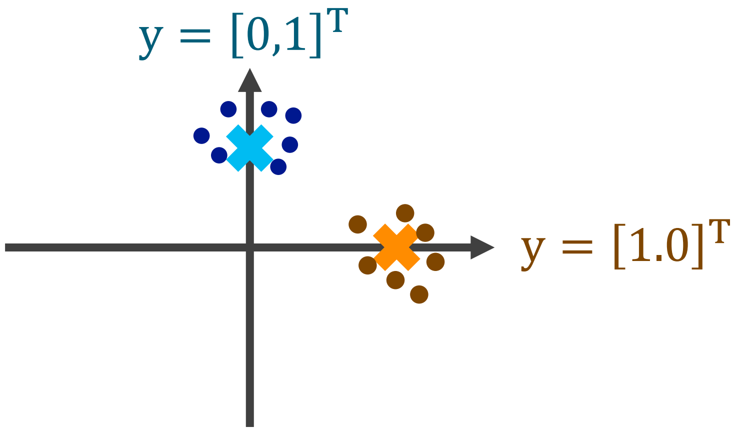
e.g., C=2



Supervised quantization

- Perform composite quantization in a discriminative space learned by a transformation matrix P
  - $\left\| \mathrm{P}^T \mathrm{x} - \bar{\mathrm{x}} \right\|_2^2 = \left\| \mathrm{P}^T \mathrm{x} - \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2$
- The encoded points belonging to the same class lie in a cluster
  - The center of cluster is defined by the label vector $\mathrm{y} \in \{0,1\}^C$
  - $\left\| \mathrm{y} - \mathrm{W}^T \sum_{m=1}^{M} \mathrm{c}_{m i_m(\mathrm{x})} \right\|_2^2 + \lambda \|\mathrm{W}\|_F^2$

e.g., C=2



Supervised quantization

- Formulation

$$\min_{\mathrm{W},\mathrm{P},\{\mathrm{C}_m\},\{i_m(\mathrm{x})\},\epsilon} \underbrace{\sum_{\mathrm{x}}\left\|\mathrm{y}-\mathrm{W}^{\mathrm{T}}\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2+\lambda\|\mathrm{W}\|_F^2}_{\text{Semantic similarity preserved by classification}}+\gamma\underbrace{\sum_{\mathrm{x}}\left\|\mathrm{P}^{\mathrm{T}}\mathrm{x}-\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2}_{\text{quantization}}$$

$$\mathrm{s.t.}\quad \sum_{m\neq l}\mathrm{c}_{mi_m(\mathrm{x})}^{T}\mathrm{c}_{li_l(\mathrm{x})}=\epsilon$$

- Unconstrained formulation

$$\min_{\mathrm{W},\mathrm{P},\{\mathrm{C}_m\},\{i_m(\mathrm{x})\},\epsilon} \sum_{\mathrm{x}}\left\|\mathrm{y}-\mathrm{W}^{\mathrm{T}}\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2+\lambda\|\mathrm{W}\|_F^2+\gamma\sum_{\mathrm{x}}\left\|\mathrm{P}^{\mathrm{T}}\mathrm{x}-\sum_{m=1}^{M}\mathrm{c}_{mi_m(\mathrm{x})}\right\|_2^2+$$
$$\mu\sum_{\mathrm{x}}\left(\sum_{m\neq l}\mathrm{c}_{mi_m(\mathrm{x})}^{T}\mathrm{c}_{li_l(\mathrm{x})}-\epsilon\right)^2$$

Alternative optimization

- Update $\mathbf{W}$
  - Closed form solution, $\mathrm{W} = \left(\overline{\mathrm{X}}\overline{\mathrm{X}}^T + \lambda \mathrm{I}_r\right)^{-1}\overline{\mathrm{X}}\mathrm{Y}^T$
- Update $\mathbf{P}$
  - Closed form solution, $\mathrm{P} = \left(\mathrm{X}\mathrm{X}^T\right)^{-1}\mathrm{X}\overline{\mathrm{X}}^T$
- Update ϵ
  - Closed form solution, $\epsilon = \frac{1}{\#\{\mathrm{x}\}}\sum_{\mathrm{x}}\sum_{m\neq l}\mathrm{c}_{m i_m(\mathrm{x})}^T \mathrm{c}_{l i_l(\mathrm{x})}$
- Update $\{\mathbf{C}_m\}$
  - L-BFGS algorithm
- Update $\{i_m(\mathrm{x})\}$
  - Iteratively alternative optimization, fixing $\{i_l(\mathrm{x})\}_{\mathrm{l}\neq m}$, update $i_m(\mathrm{x})$

Experiments

- Datasets

Dataset	Dimension	Feature type	#training samples	#test samples
CIFAR-10	512	GIST feature	59,000	1,000
MNIST	784	Raw pixel feature	69,000	1,000
NUS-WIDE	500	Bag-of-words feature	191,652	2,100

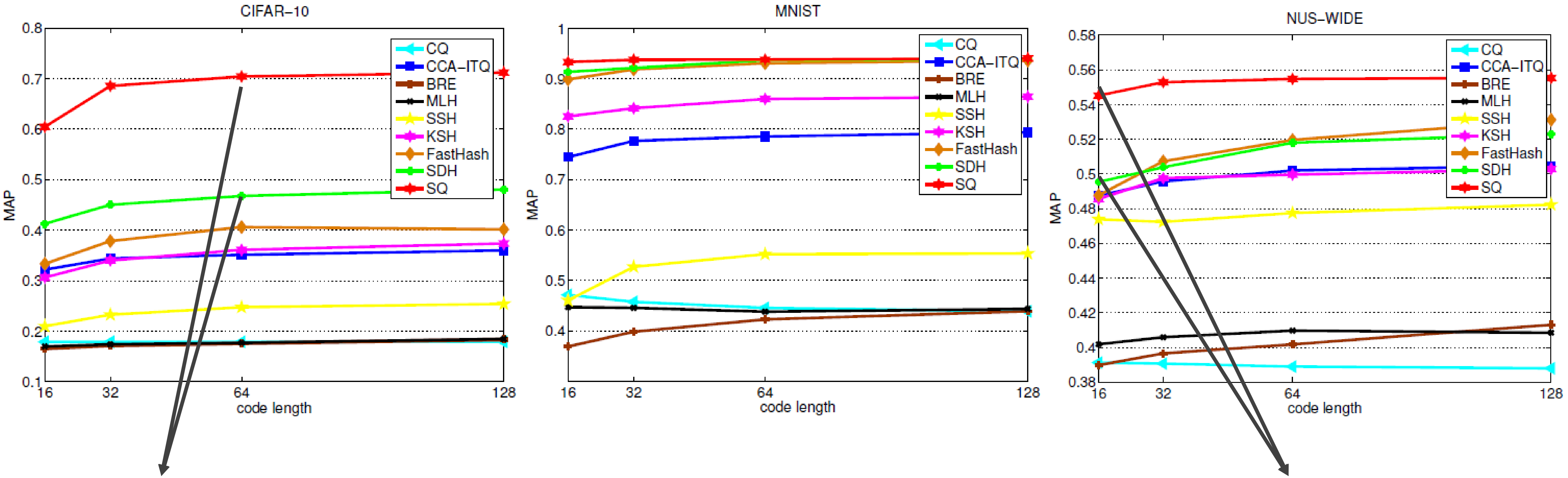
Experiments

- Compared methods
  - Supervised hashing
    - Supervised discrete hashing (SDH)
    - FastHash
    - Supervised hashing with kernels (KSH)
    - CCA-ITQ
    - Semi-supervised hashing (SSH)
    - Minimal loss hashing (MLH)
    - Binary reconstructive embedding (BRE)
  - Unsupervised quantization
    - Composite quantization (CQ)

State-of-the-art methods

- Evaluation: Mean average precision (MAP)

Comparison with SDH

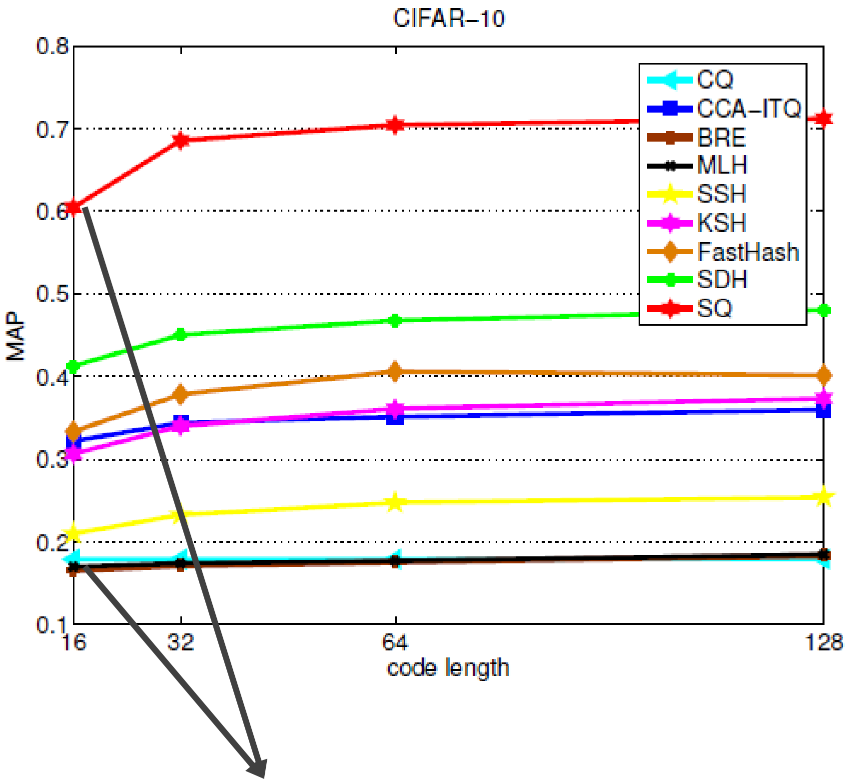


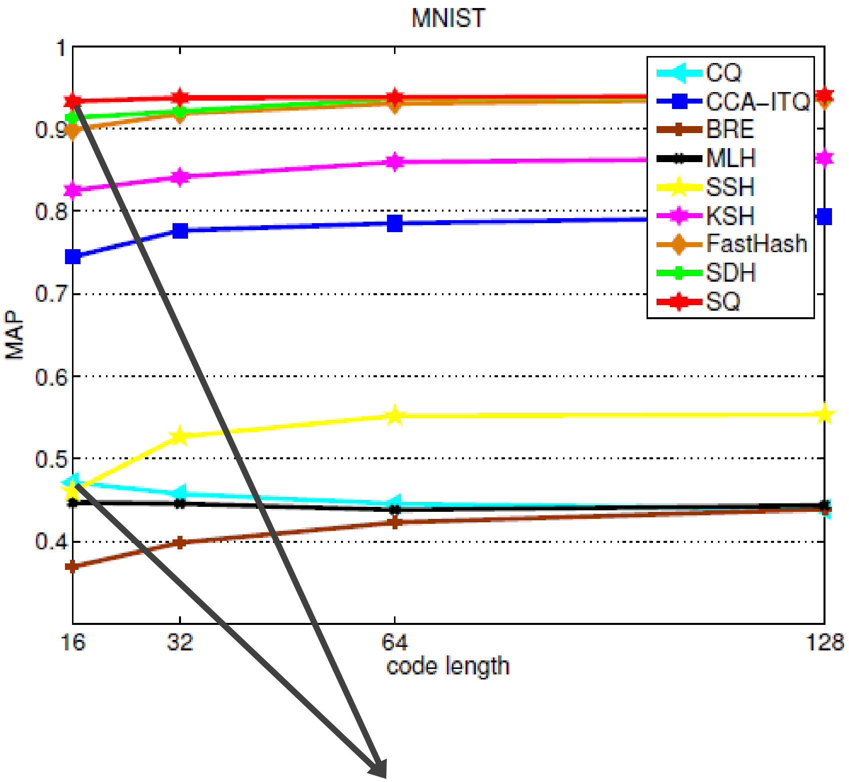
MAP Improvement on 64 bits: 23.66%

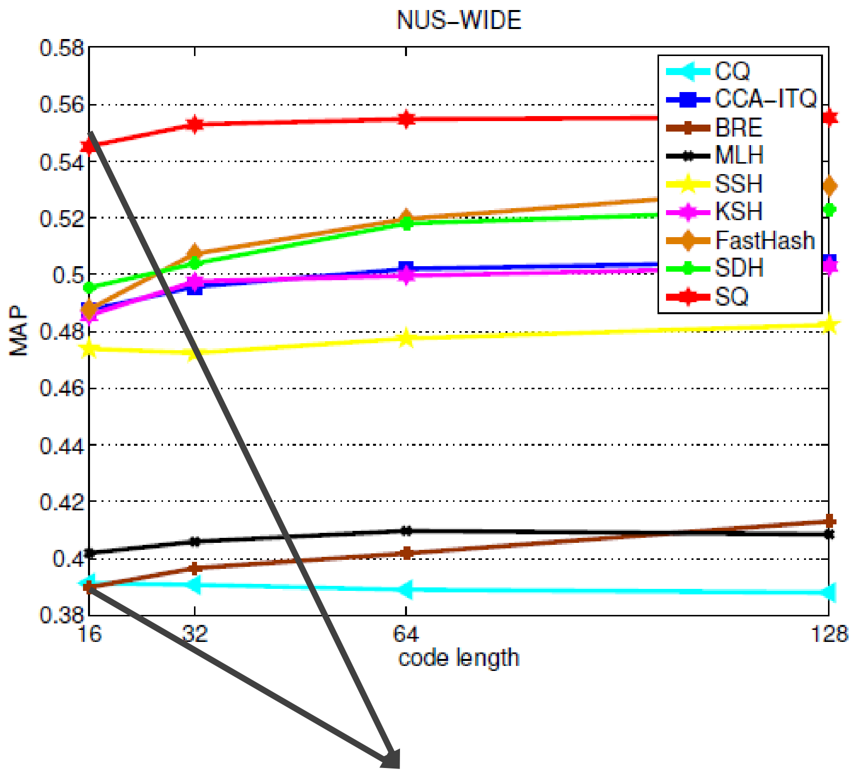
Relatively small improvement on MNIST because SDH already achieves a high performance

MAP Improvement on 16 bits: 4.65%

Comparison with CQ







MAP Improvement
on 16 bits: 42.57%

MAP Improvement
on 16 bits: 46.14%

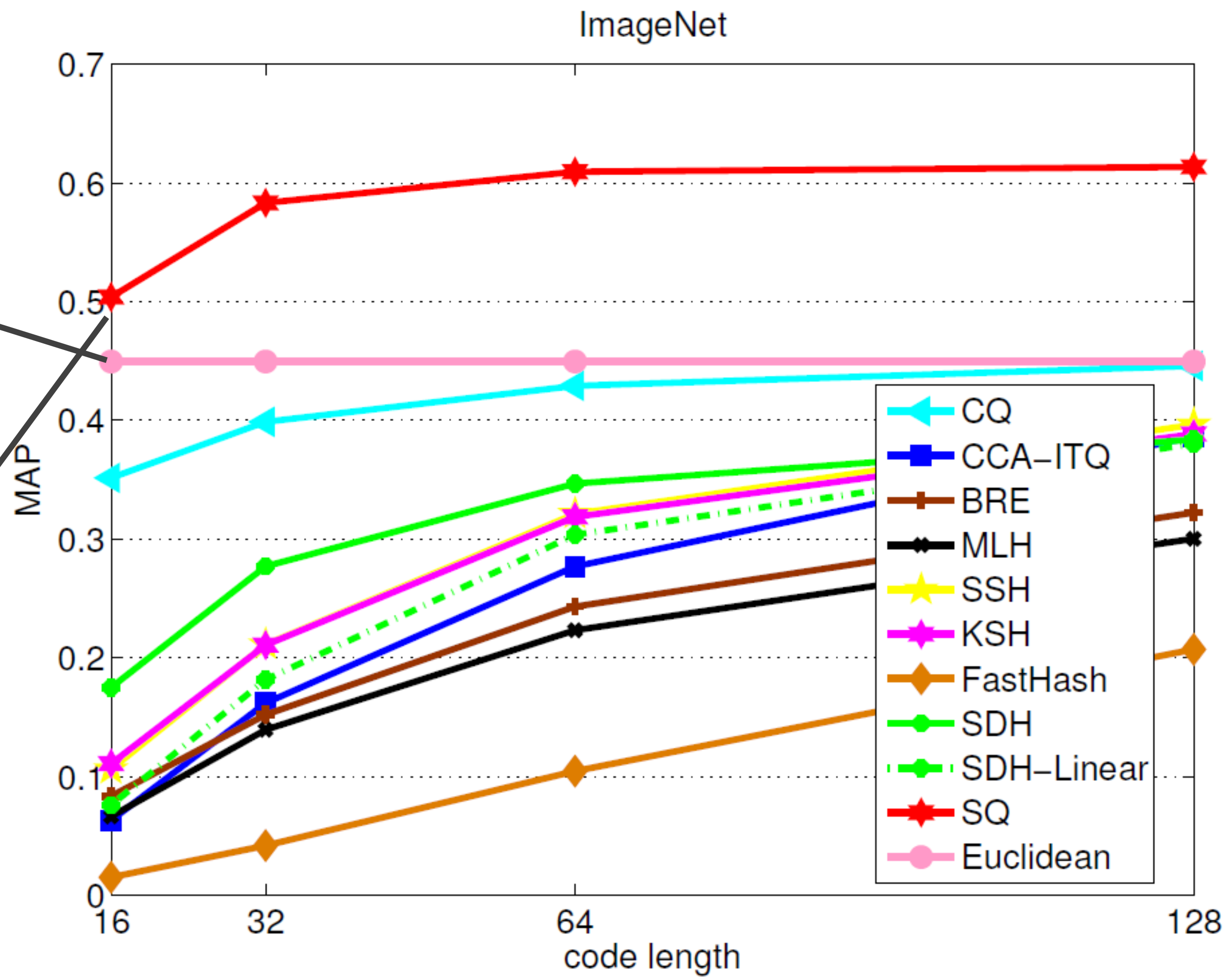
MAP Improvement
on 16 bits: 15.39%

Supervision indeed benefits the search performance

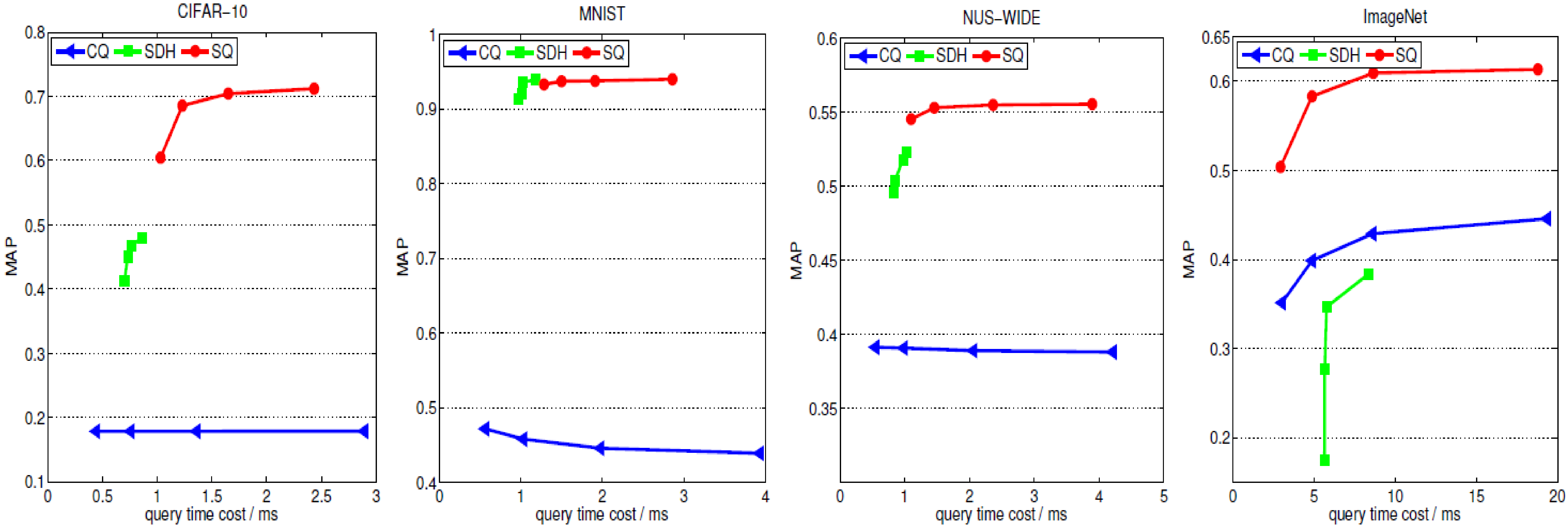
Comparison on ImageNet

Euclidean results outperforming other hashing suggests that CNN feature has powerful discriminative ability

Our approach learns better quantizer through the supervision information



Search efficiency



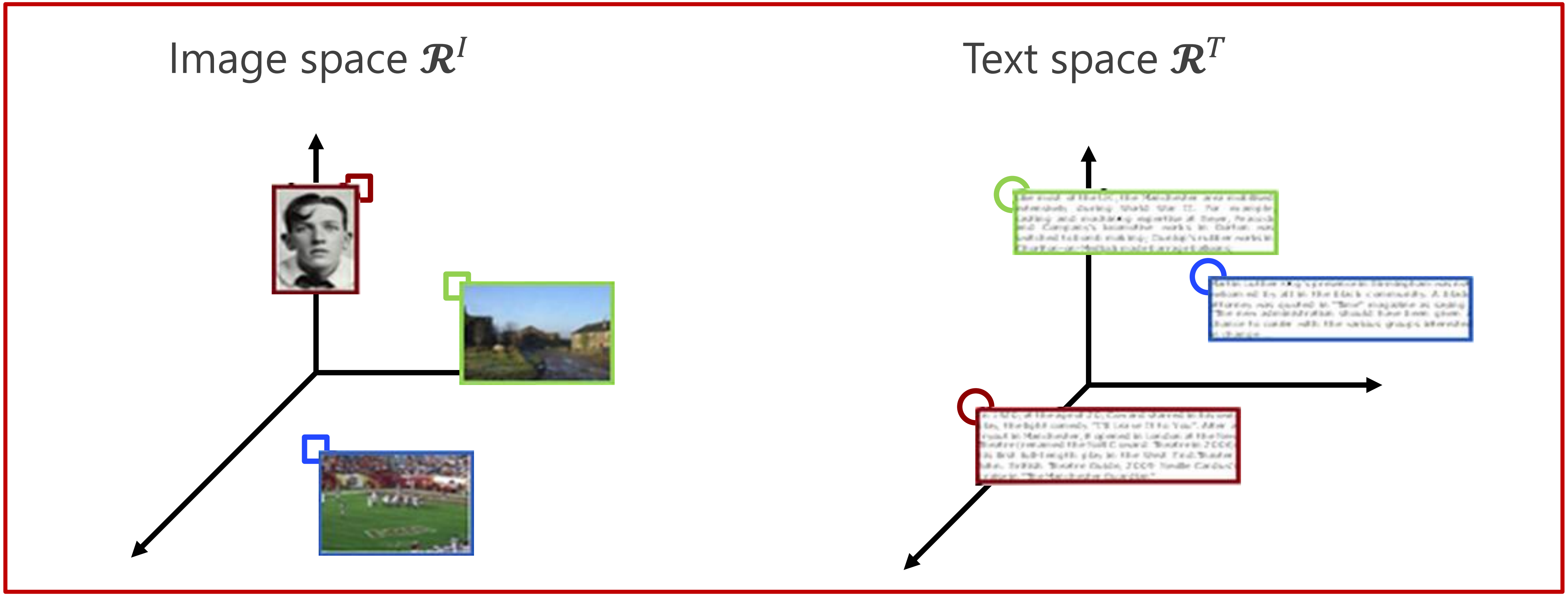
Obtain higher performance for the same query time

Outline

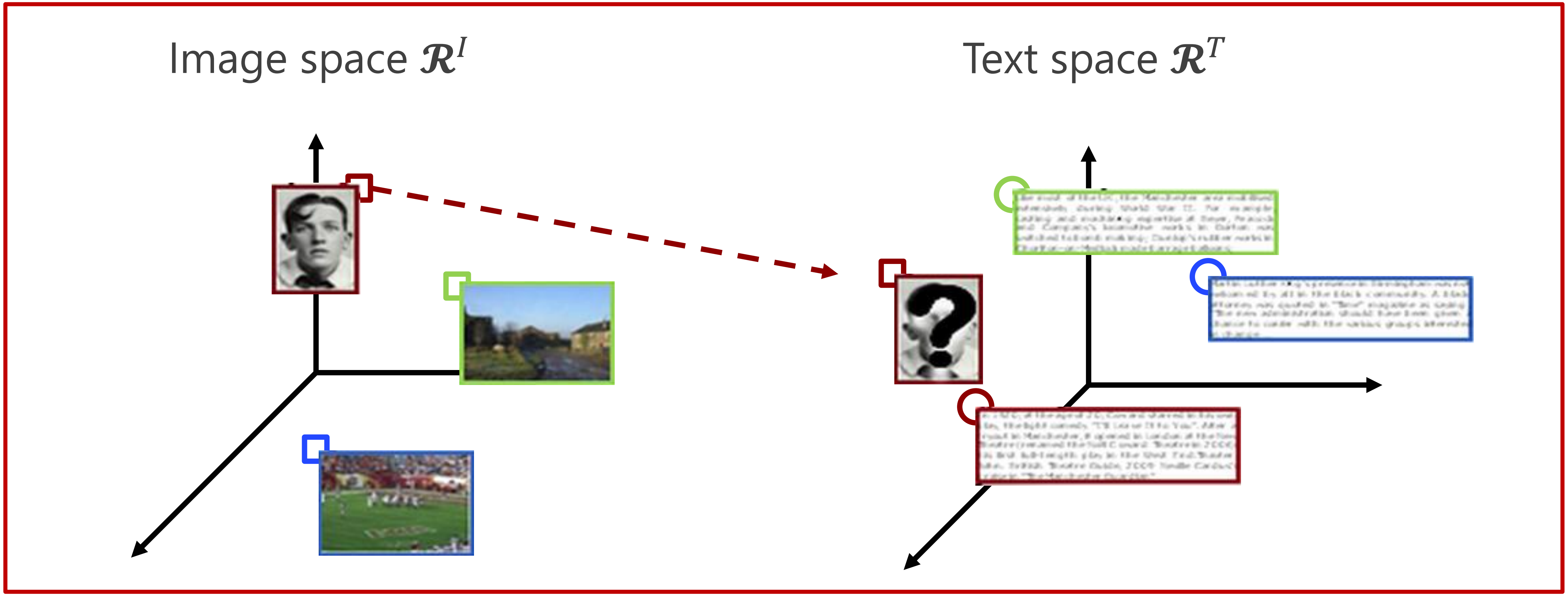
- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Collaborative quantization for cross-modal similarity search. Ting Zhang, Jingdong Wang. CVPR 2016

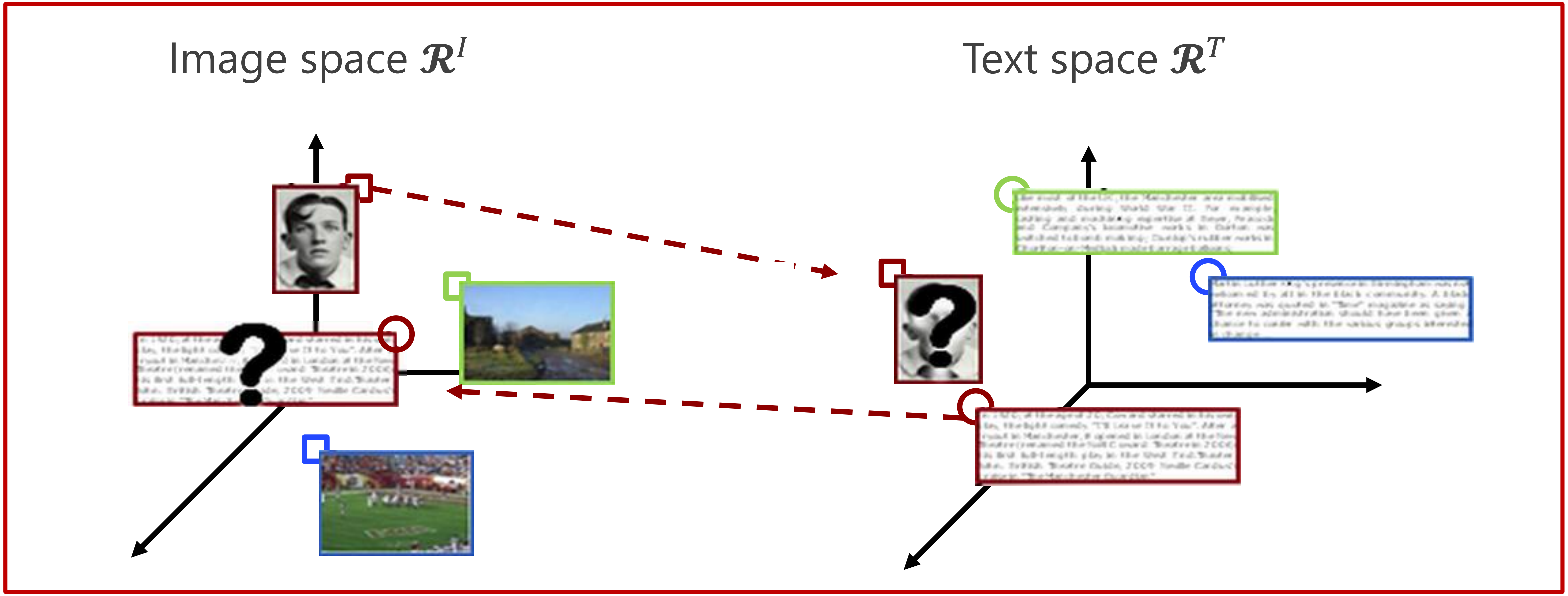
Cross-modal similarity search



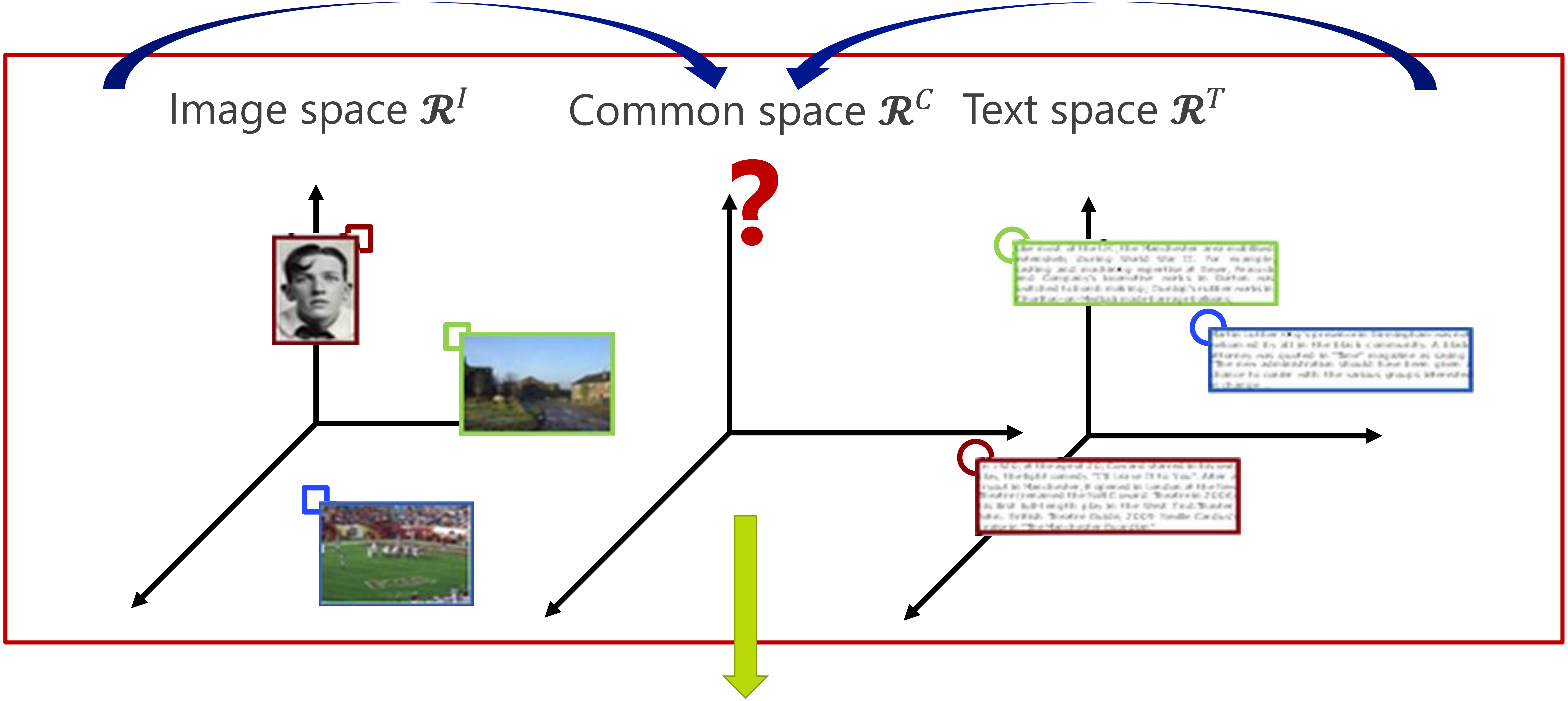
Cross-modal similarity search



Cross-modal similarity search



Compact coding approach

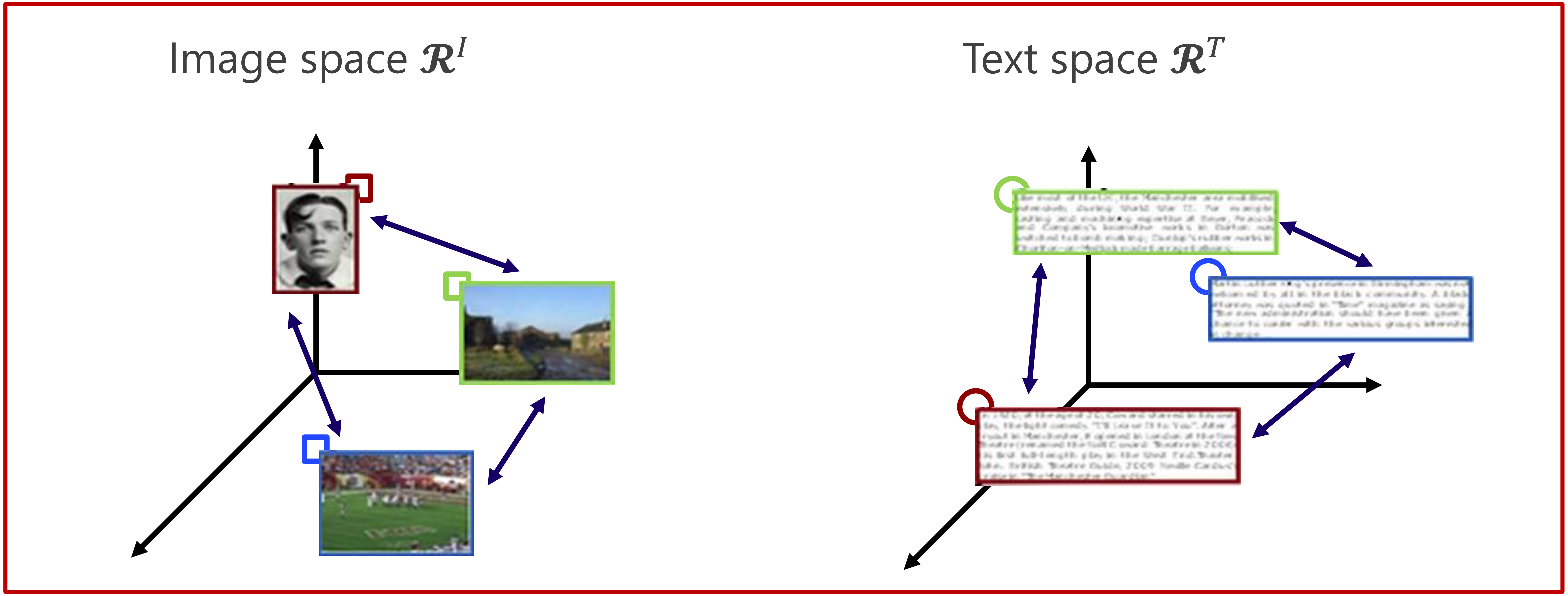


Perform hashing or quantization

Cross-modal similarity search

- Compact coding approaches
  - Map data from different modalities into a common space (by exploring the relations between the modalities)
    - Intra-modality relation (image vs. image and text vs. text)
  - Obtain codes by performing hashing or quantization

Cross-modal similarity search

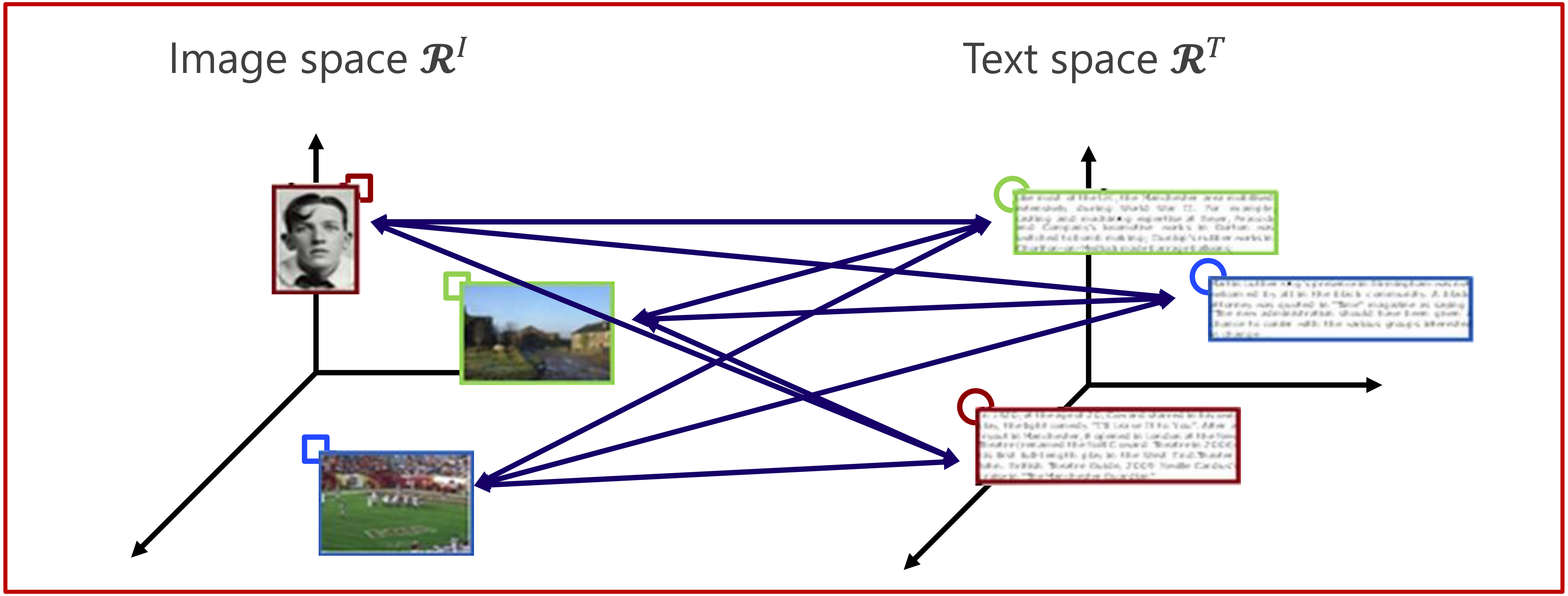


Intra-modality relation (image vs. image and text vs. text)

Cross-modal similarity search

- Compact coding approaches
  - Map data from different modalities into a common space (by exploring the relations between the modalities)
    - Intra-modality relation (image vs. image and text vs. text)
    - Inter-modality relation (image vs. text)
  - Obtain codes by performing hashing or quantization

Cross-modal similarity search



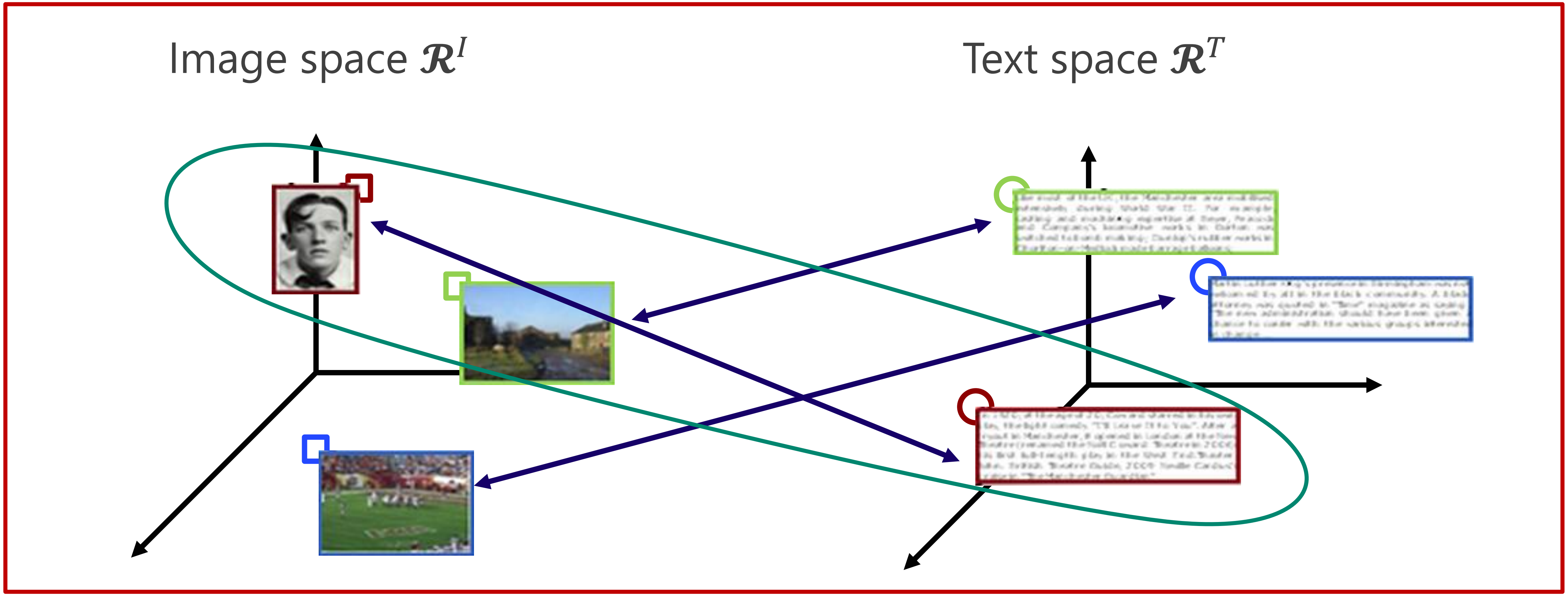
Inter-modality relation (image vs. text)

Cross-modal similarity search

- Compact coding approaches
  - Map data from different modalities into a common space (by exploring the relations between the modalities)
    - Intra-modality relation (image vs. image and text vs. text)
    - Inter-modality relation (image vs. text)
    - Intra-document relation (the correspondence of an image and a text forming a document)

  - Obtain codes by performing hashing or quantization

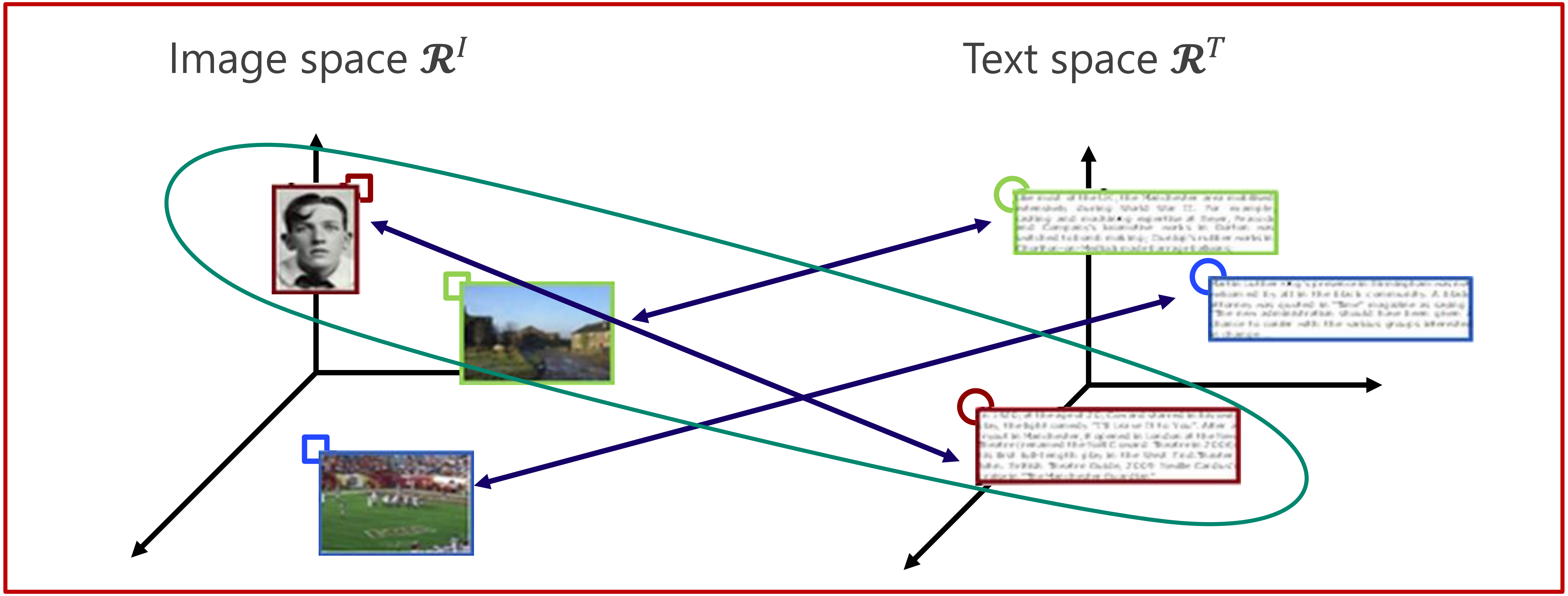
Cross-modal similarity search



Intra-document relation (the correspondence of an image and a text forming a document)

A pair of an image and a text

Cross-modal similarity search



Intra-document relation (the correspondence of an image and a text forming a document)

A pair of an image and a text

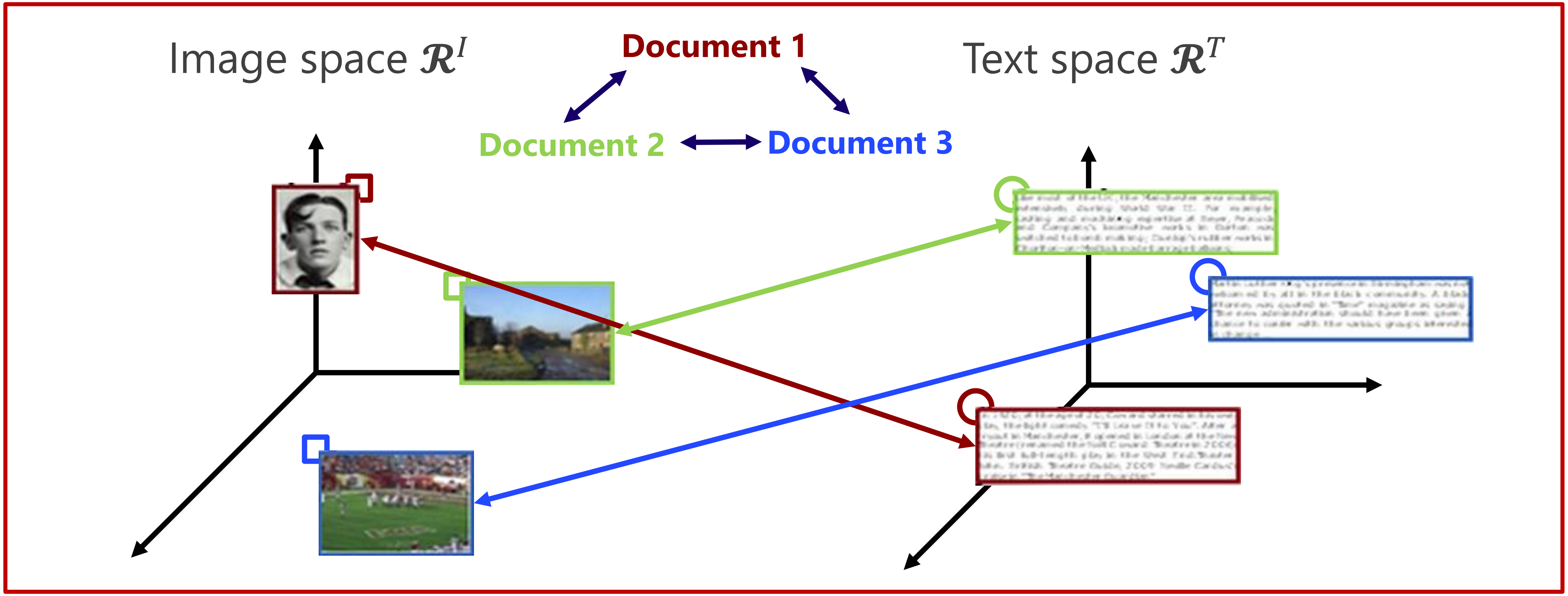
A special kind of inter-modality relation

Cross-modal similarity search

- Compact coding approaches
  - Map data from different modalities into a common space (by exploring the relations between the modalities)
    - Intra-modality relation (image vs. image and text vs. text)
    - Inter-modality relation (image vs. text)
    - Intra-document relation (the correspondence of an image and a text forming a document)
    - Inter-document relation (document vs. document)

  - Obtain codes by performing hashing or quantization

Cross-modal similarity search



Inter-document relation (document vs. document)

Cross-modal similarity search

- Compact coding approaches
  - Map data from different modalities into a common space (by exploring the relations between the modalities)
    - Intra-modality relation (image vs. image and text vs. text)
    - Inter-modality relation (image vs. text)
    - Intra-document relation (document vs. document)
    - Inter-document relation (the correspondence of an image and a text forming a document)
  - Obtain codes by performing hashing or quantization
    - One unified code for each document
    - Two separate codes, each corresponding to a modality

Categorization

Method	Multi-modal data relations				Codes		Coding method	
	Intra-modality	Inter-modality	Intra-document	Inter-document	Unified	Separate	Hash	Quantization
CMSSH		▲				▲	▲	
SCM		▲				▲	▲	
CRH		▲				▲	▲	
MMNN	▲	▲				▲	▲	
SM^{2}H	▲	▲				▲	▲	
MLBE	▲	▲				▲	▲	
IMH	▲		▲			▲	▲	
CVH			▲	▲		▲	▲	
MVSH			▲	▲		▲	▲	
SPH			▲	▲	▲		▲	
LSSH			▲		▲		▲	
CMFH			▲		▲		▲	
STMH			▲		▲		▲	
QCH		▲				▲	▲	
CCQ			▲		▲			▲
Ours			▲			▲		▲

Formulation

- Common space mapping
  - Text data Y to common space Y'
    - Matrix factorization $\|\mathrm{Y} - \mathrm{U}\mathrm{Y}'\|$

Text data Y

Y'

Formulation

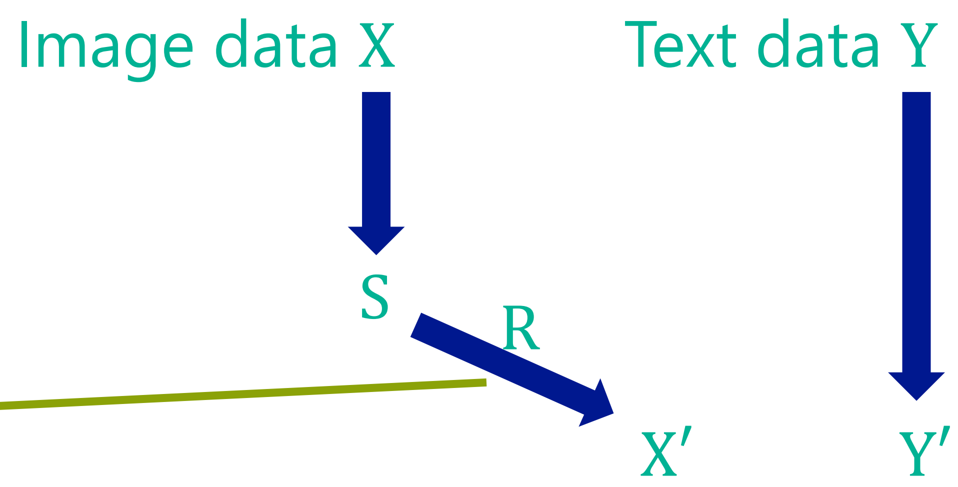
- Common space mapping
  - Image data X to common space
    - Sparse coding $\|X - BS\|_F^2 + \rho|S|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|Y - UY'\|$

Image data X $\rightarrow$ S

Text data Y $\rightarrow$ Y'

Formulation

- Common space mapping
  - Image data X to common space $\mathrm{X}' = \mathrm{RS}$
    - Sparse coding $\|\mathrm{X} - \mathrm{BS}\|_F^2 + \rho|\mathrm{S}|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|\mathrm{Y} - \mathrm{UY}'\|$

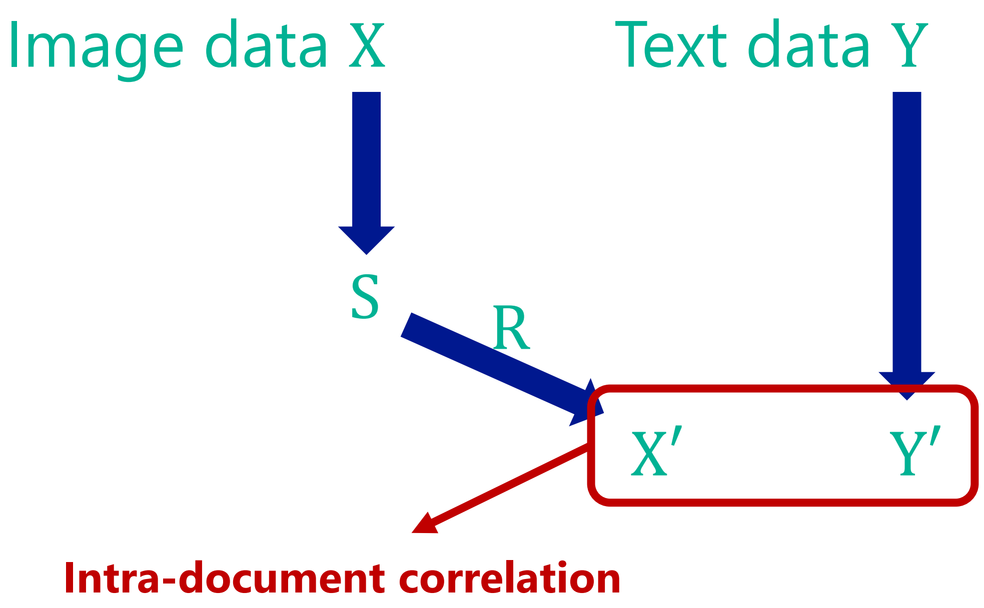


Formulation

- Common space mapping

$$\mathcal{M} = \qquad + \eta \qquad + \lambda \boxed{\|\mathrm{Y}' - \mathrm{RS}\|_F^2}$$

  - Image data X to common space $\mathrm{X}' = \mathrm{RS}$
    - Sparse coding $\|\mathrm{X} - \mathrm{BS}\|_F^2 + \rho|\mathrm{S}|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|\mathrm{Y} - \mathrm{UY}'\|$

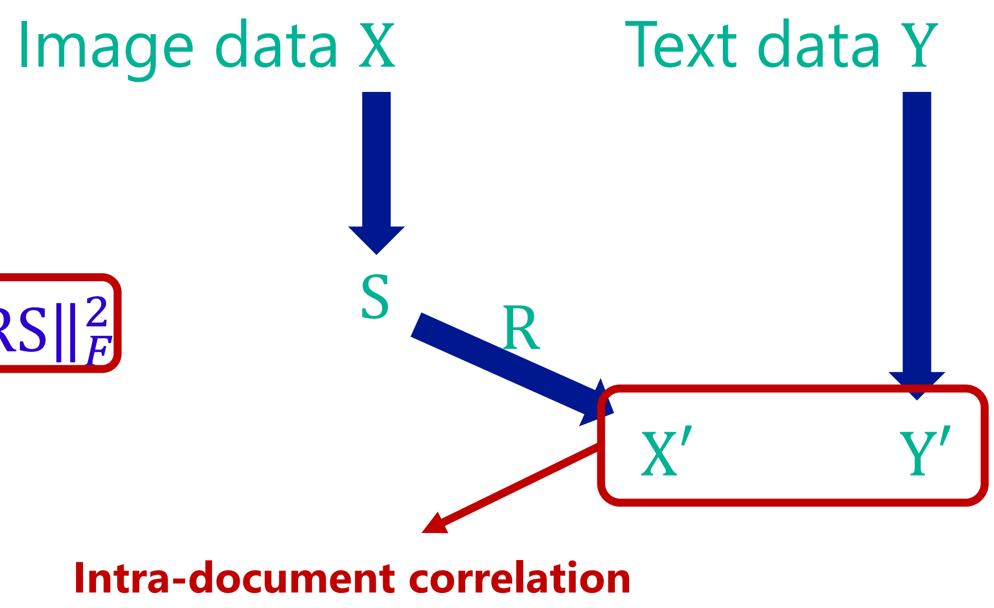


Formulation

- Common space mapping

$$\mathcal{M} = \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\boxed{\|Y' - RS\|_F^2}$$

  - Image data X to common space $X' = RS$
    - Sparse coding $\|X - BS\|_F^2 + \rho|S|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|Y - UY'\|$

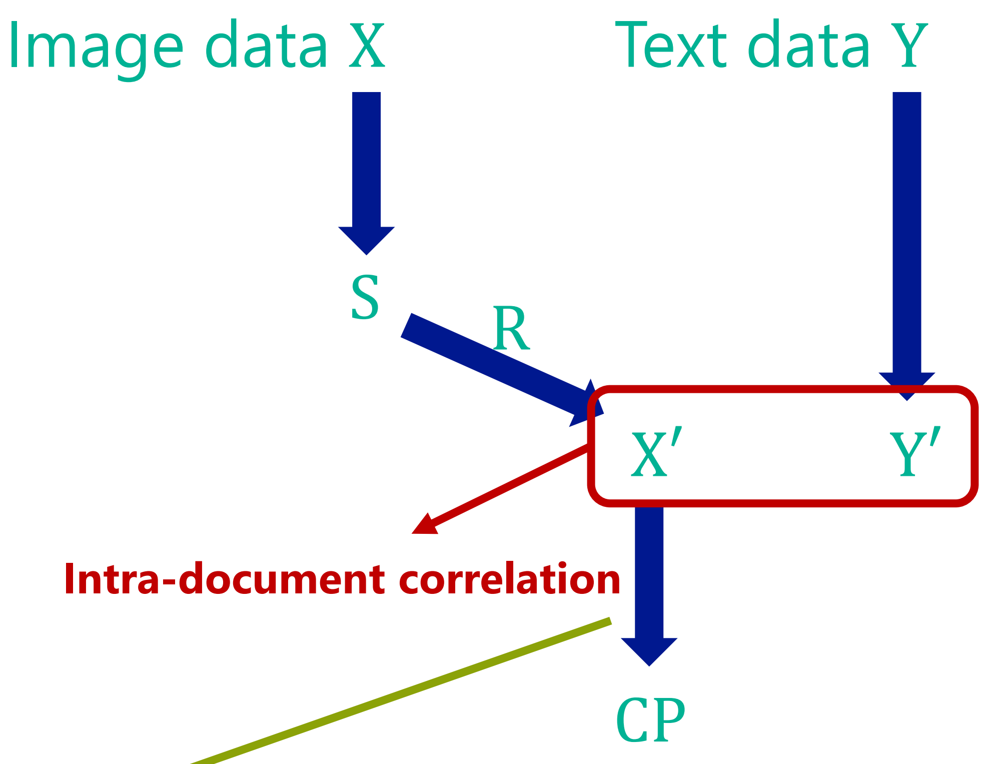


Formulation

- Common space mapping

$$\mathcal{M} = \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

  - Image data X to common space $X' = RS$
    - Sparse coding $\|X - BS\|_F^2 + \rho|S|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|Y - UY'\|$
- Collaborative quantization
  - Image quantization $\|X' - CP\|_F^2$

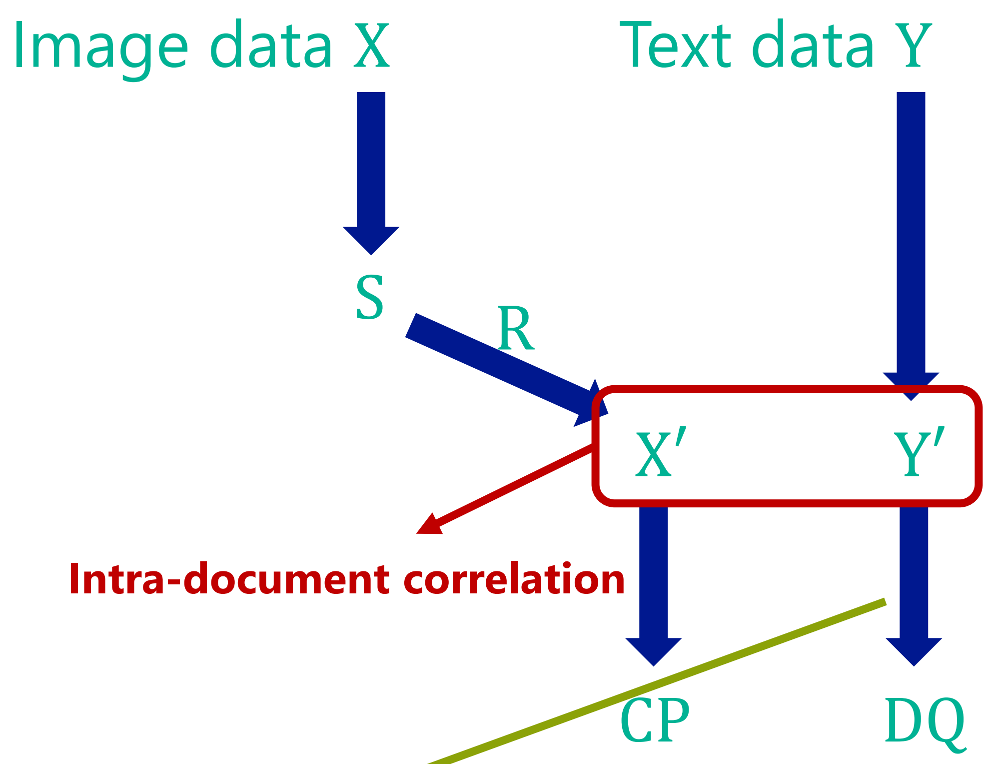


Formulation

- Common space mapping

$$\mathcal{M} = \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

  - Image data X to common space $X' = RS$
    - Sparse coding $\|X - BS\|_F^2 + \rho|S|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|Y - UY'\|$
- Collaborative quantization
  - Image quantization $\|X' - CP\|_F^2$
  - Text quantization $\|Y' - DQ\|_F^2$



Formulation

- Common space mapping

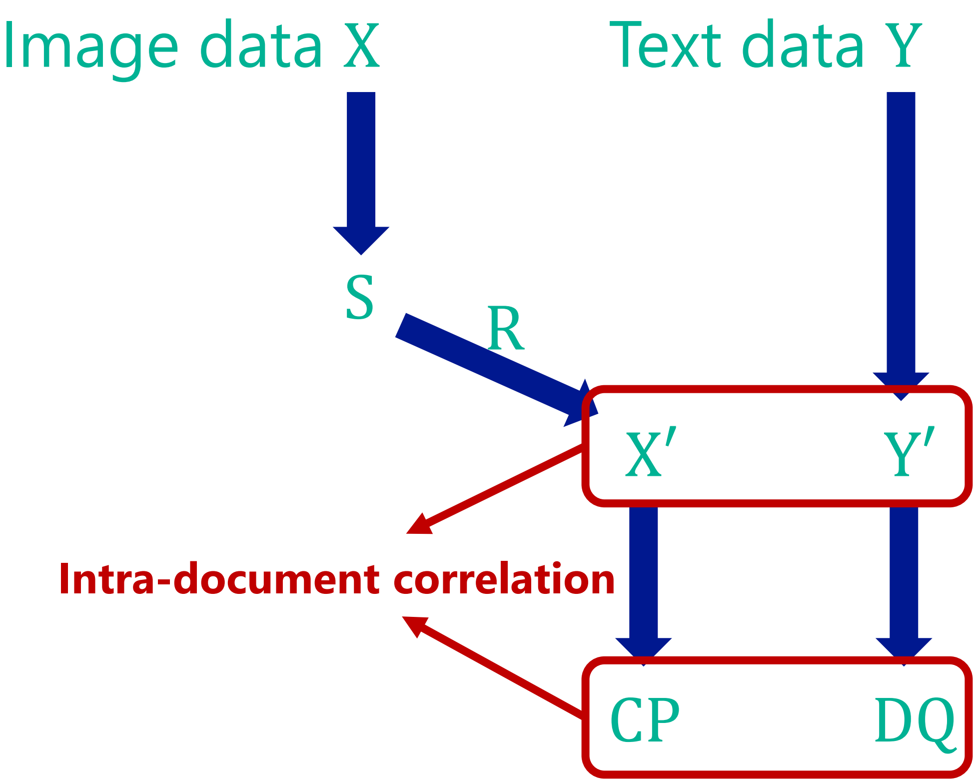
$$\mathcal{M} = \|\mathrm{X} - \mathrm{BS}\|_F^2 + \rho|\mathrm{S}|_{11} + \eta\|\mathrm{Y} - \mathrm{UY}'\|_F^2 + \lambda\|\mathrm{Y}' - \mathrm{RS}\|_F^2$$

  - Image data X to common space $\mathrm{X}' = \mathrm{RS}$
    - Sparse coding $\|\mathrm{X} - \mathrm{BS}\|_F^2 + \rho|\mathrm{S}|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|\mathrm{Y} - \mathrm{UY}'\|$

- Collaborative quantization

$$\mathcal{Q} = \qquad + \qquad + \gamma\boxed{\|\mathrm{CP} - \mathrm{DQ}\|_F^2}$$

  - Image quantization $\|\mathrm{X}' - \mathrm{CP}\|_F^2$
  - Text quantization $\|\mathrm{Y}' - \mathrm{DQ}\|_F^2$



Formulation

- Common space mapping

$\mathcal{M} = \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$

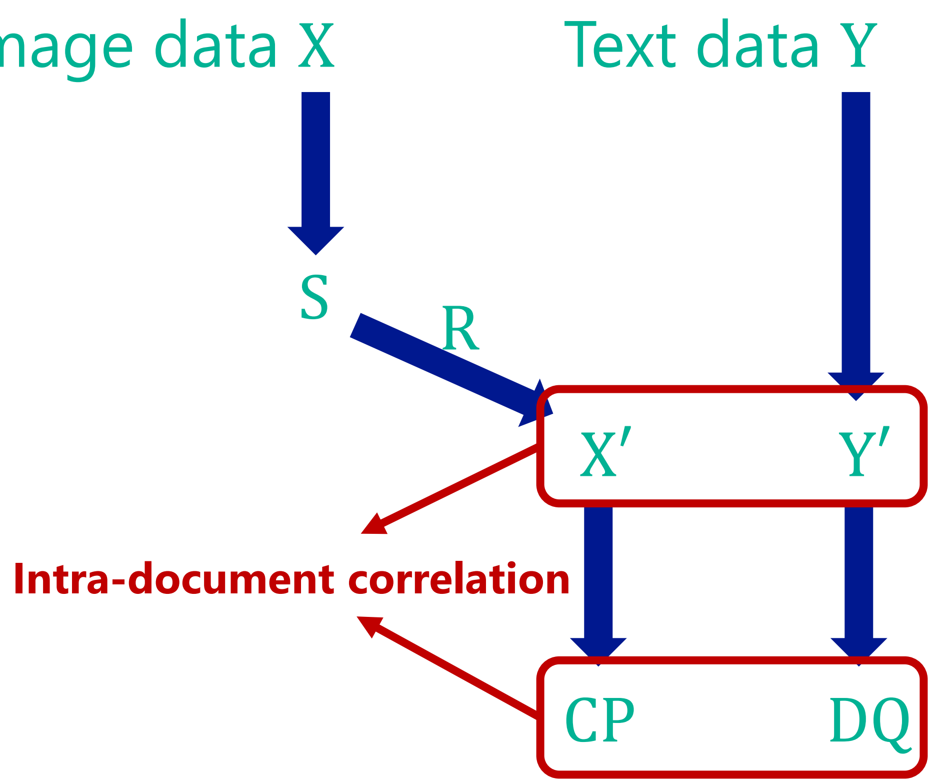
  - Image data X to common space $X' = RS$
    - Sparse coding $\|X - BS\|_F^2 + \rho|S|_{11}$
  - Text data Y to common space Y'
    - Matrix factorization $\|Y - UY'\|$

- Collaborative quantization

$\mathcal{Q} = \|X' - CP\|_F^2 + \|Y' - DQ\|_F^2 + \gamma\|CP - DQ\|_F^2$

  - Image quantization $\|X' - CP\|_F^2$
  - Text quantization $\|Y' - DQ\|_F^2$

- Overall objective function $\mathcal{F} = \mathcal{Q} + \mathcal{M}$



Alternative optimization $\mathcal{F} = \mathcal{Q} + \mathcal{M}$

- Fix $\mathcal{M}$, update $\mathcal{Q}$
  - Update each variable in $\mathcal{M}$ when fixing others

- Fix $\mathcal{Q}$, update $\mathcal{M}$
  - Update each variable in $\mathcal{Q}$ when fixing others

Experiments

- Datasets

Dataset	#Classes	Image feature type	Text feature type	Training samples	Test samples
Wiki	10	128D SIFT vectors	10D topic vectors	2173	693
FLICKR25K	38	3857D vectors	2000D vectors	22500	2500
NUS-WIDE	10	500D BoW vectors	1000D vectors	182577	4000

- Evaluation
  - Mean average precision (MAP@T)
    - Compute MAP at T retrieved items
  - Precision@T
    - Compute precision at T retrieved items

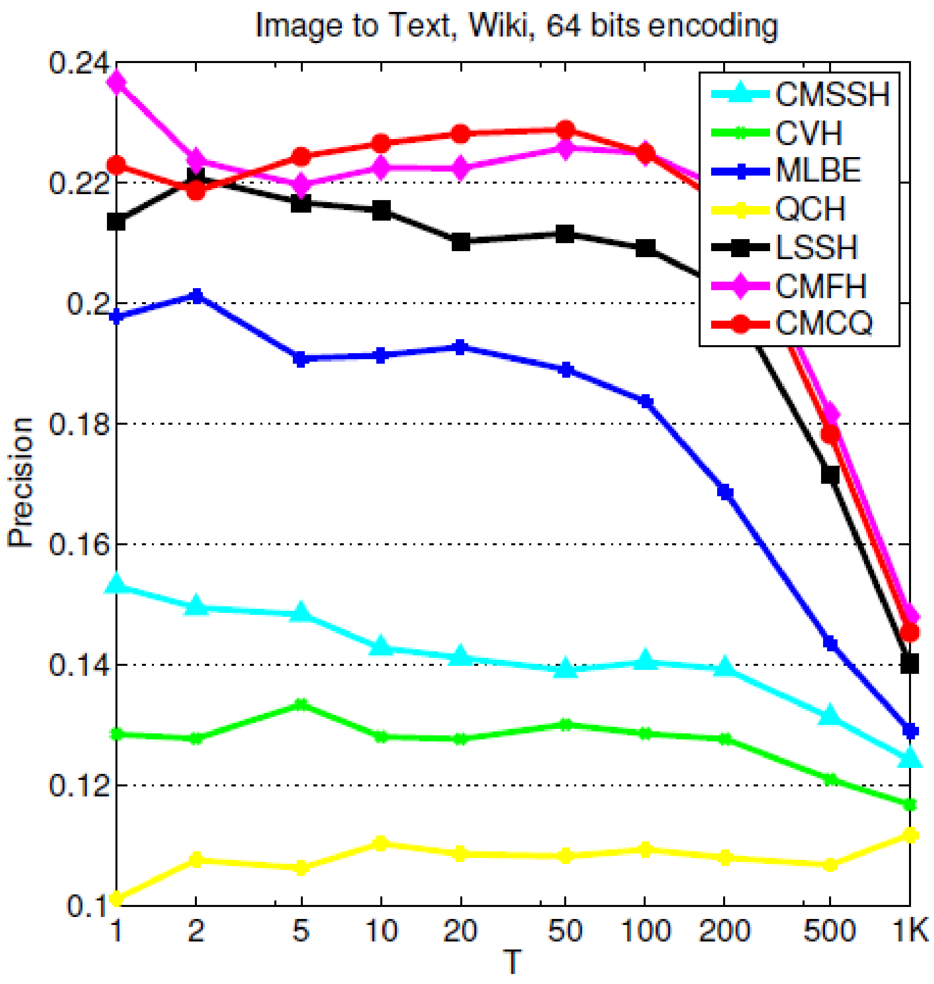
Results on Wiki

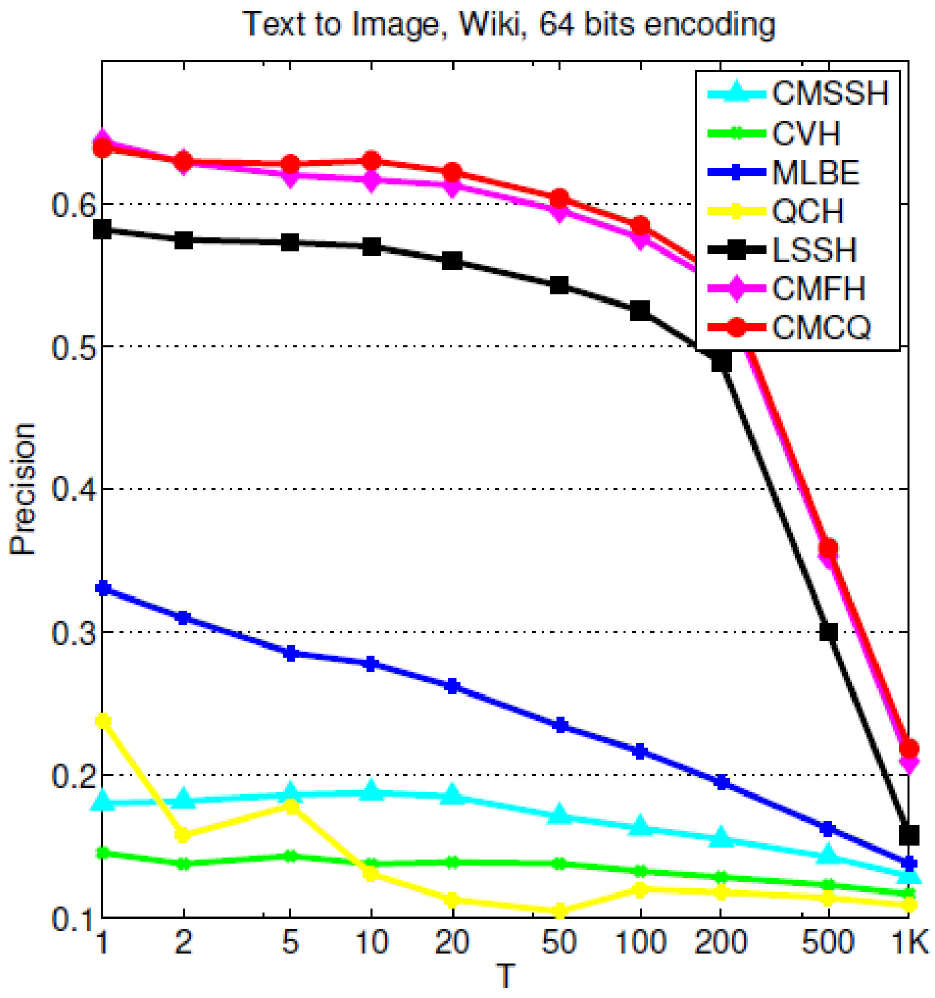
- MAP@50 comparison

Task	Method	Wiki 16 bits	Wiki 32 bits	Wiki 64 bits	Wiki 128 bits
Img to Txt	CMSSH [1]	0.2110	0.2115	0.1932	0.1909
	CVH [8]	0.1947	0.1798	0.1732	0.1912
	MLBE [29]	**0.3537**	**0.3947**	0.2599	0.2247
	QCH [23]	0.1490	0.1726	0.1621	0.1611
	LSSH [30]	0.2396	0.2336	0.2405	0.2373
	CMFH [4]	0.2548	0.2591	0.2594	**0.2651**
	(CMFH [4])	(0.2538)	(0.2582)	(**0.2619**)	(0.2648)
	(CCQ [10])	(0.2513)	(0.2529)	(0.2587)	—
	CMCQ	0.2478	0.2513	0.2567	0.2614
Txt to Img	CMSSH [1]	0.2446	0.2505	0.2387	0.2352
	CVH [8]	0.3186	0.2354	0.2046	0.2085
	MLBE [29]	0.3336	0.3993	0.4897	0.2997
	QCH [23]	0.1924	0.1561	0.1800	0.1917
	LSSH [30]	0.5776	0.5886	0.5998	0.6103
	CMFH [4]	0.6153	0.6363	0.6411	0.6504
	(CMFH [4])	(0.6116)	(0.6298)	(0.6398)	(0.6477)
	(CCQ [10])	(0.6351)	(0.6394)	(0.6405)	—
	CMCQ	**0.6397**	**0.6474**	**0.6546**	**0.6593**

Results on Wiki

- Precision@T comparison on 64 bits





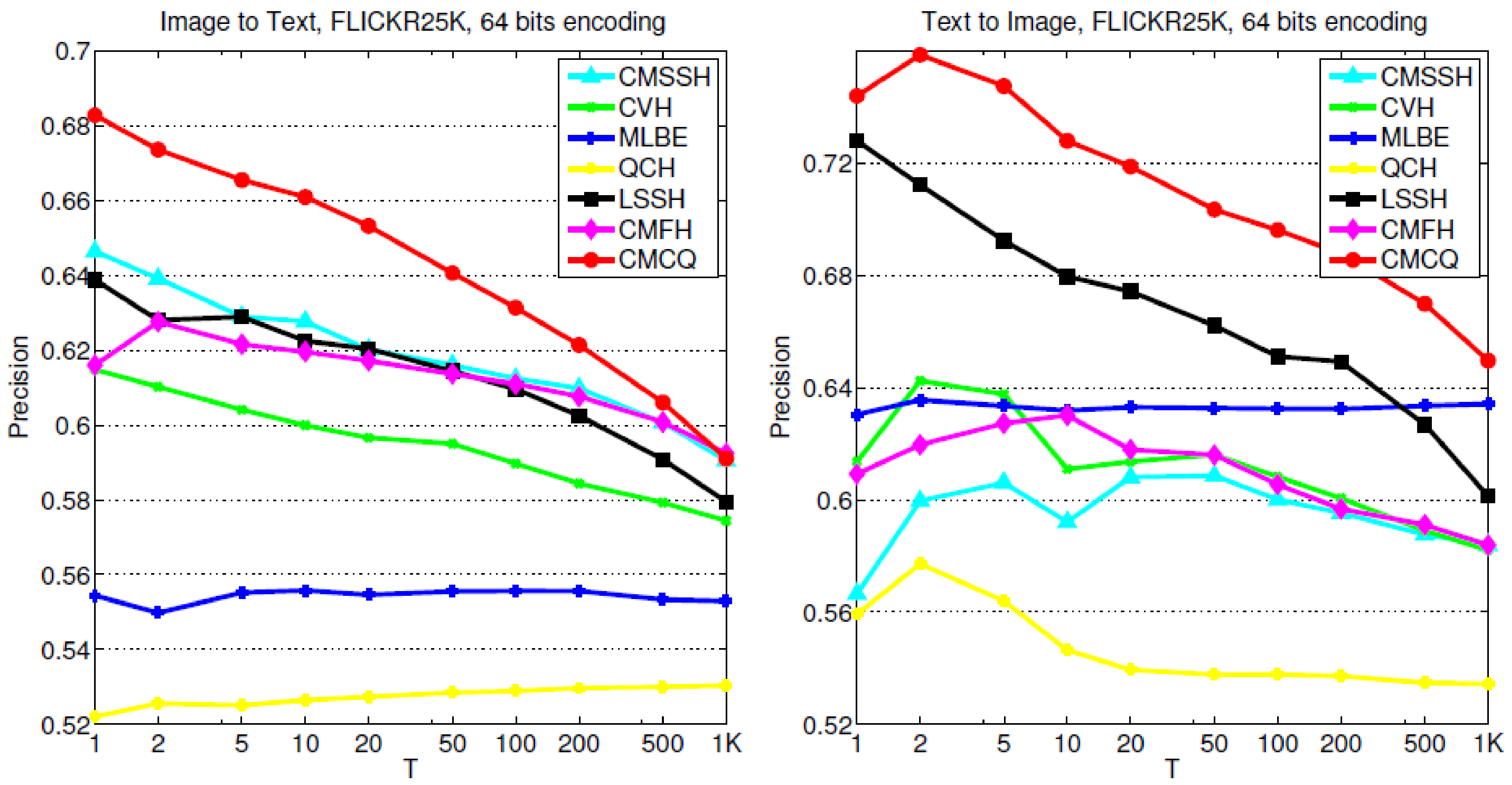
Results on FLICKR25K

- MAP@50 comparison

Task	Method	FLICKR25K			
		16 bits	32 bits	64 bits	128 bits
Img to Txt	CMSSH [1]	0.6468	0.6616	0.6681	0.6624
	CVH [8]	0.6450	0.6363	0.6273	0.6204
	MLBE [29]	0.6085	0.5866	0.5841	0.5883
	QCH [23]	0.5722	0.5780	0.5618	0.5567
	LSSH [30]	0.6328	0.6403	0.6451	0.6511
	CMFH [4]	0.5886	0.6067	0.6343	0.6550
	(CMFH [4])	—	—	—	—
	(CCQ [10])	—	—	—	—
	CMCQ	**0.6705**	**0.6716**	**0.6782**	**0.6821**
Txt to Img	CMSSH [1]	0.6123	0.6400	0.6382	0.6242
	CVH [8]	0.6595	0.6507	0.6463	0.6580
	MLBE [29]	0.5937	0.6182	0.6550	0.6392
	QCH [23]	0.5752	0.6002	0.5757	0.5723
	LSSH [30]	0.6504	0.6726	0.6965	0.7010
	CMFH [4]	0.5873	0.6019	0.6477	0.6623
	(CMFH [4])	—	—	—	—
	(CCQ [10])	—	—	—	—
	CMCQ	**0.7248**	**0.7335**	**0.7394**	**0.7550**

Results on FLICKR25K

- Precision@T comparison on 64 bits



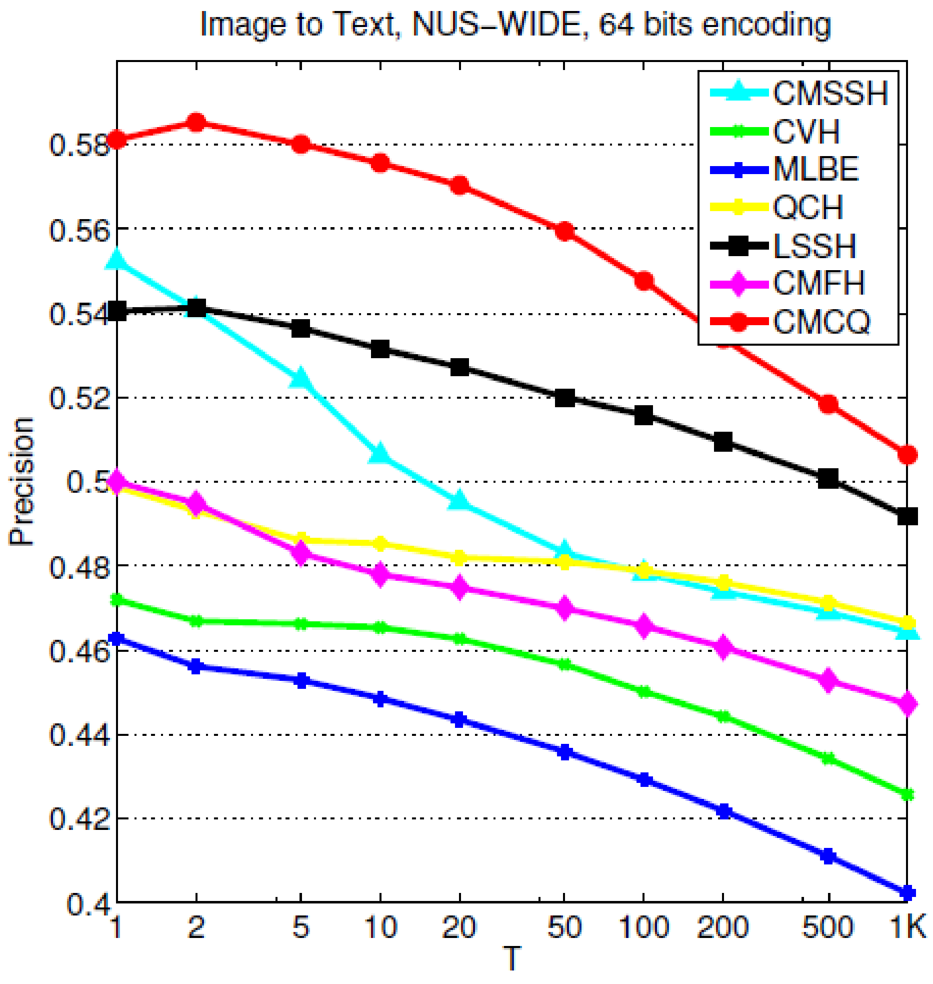
Results on NUS-WIDE

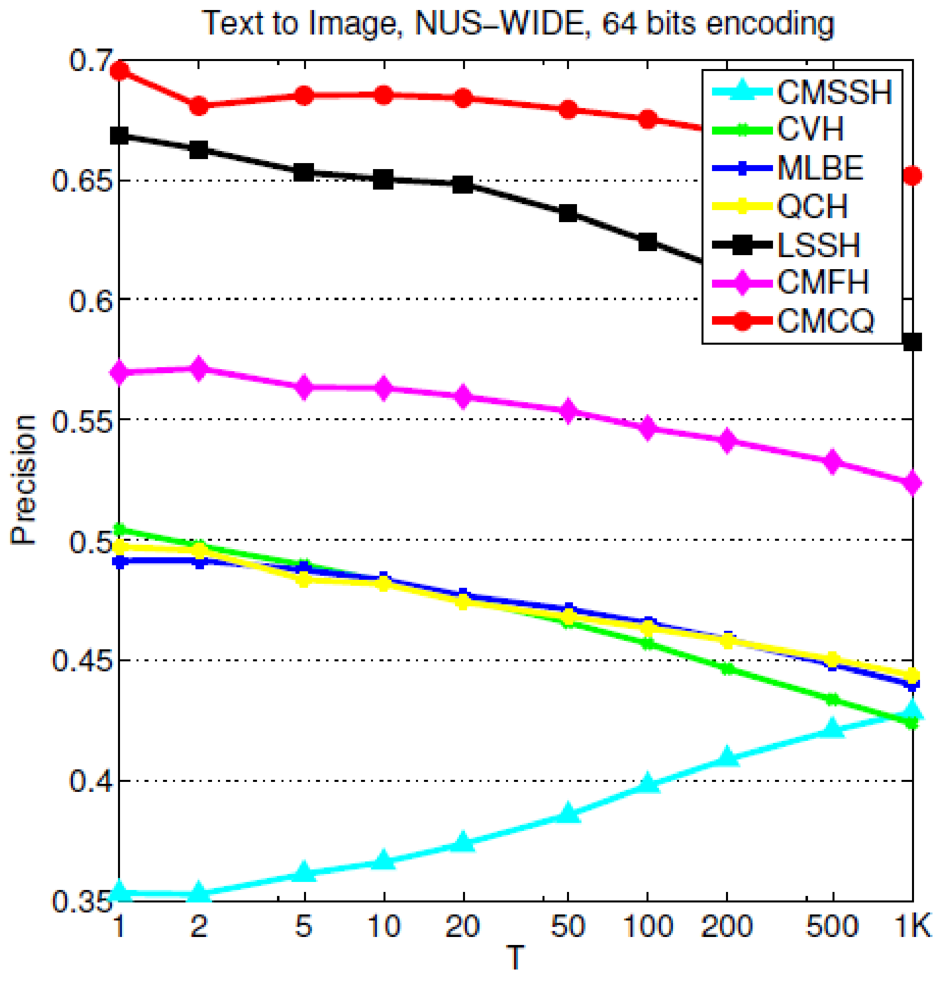
- MAP@50 comparison

Task	Method	NUS-WIDE 16 bits	32 bits	64 bits	128 bits
Img to Txt	CMSSH [1]	0.5243	0.5210	0.5211	0.4813
	CVH [8]	0.5352	0.5254	0.5011	0.4705
	MLBE [29]	0.4472	0.4540	0.4703	0.4026
	QCH [23]	0.5090	0.5270	0.5208	0.5135
	LSSH [30]	0.5368	0.5527	0.5674	0.5723
	CMFH [4]	0.4740	0.4821	0.5130	0.5068
	(CMFH [4])	—	—	—	—
	(CCQ [10])	—	—	—	—
	CMCQ	**0.5637**	**0.5902**	**0.5990**	**0.6096**
Txt to Img	CMSSH [1]	0.4177	0.4259	0.4187	0.4203
	CVH [8]	0.5601	0.5439	0.5160	0.4821
	MLBE [29]	0.4352	0.4888	0.5020	0.4425
	QCH [23]	0.5099	0.5172	0.5092	0.5089
	LSSH [30]	0.6357	0.6638	0.6820	0.6926
	CMFH [4]	0.5109	0.5643	0.5896	0.5943
	(CMFH [4])	—	—	—	—
	(CCQ [10])	—	—	—	—
	CMCQ	**0.6898**	**0.7086**	**0.7194**	**0.7254**

Results on NUS-WIDE

- Precision@T comparison on 64 bits





Empirical analysis

$$\mathcal{F} = \|X' - CP\|_F^2 + \|Y' - DQ\|_F^2 + \gamma\|CP - DQ\|_F^2 + \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

- The effect of intra-document correlation, i.e., $\gamma = 0$ vs. $\gamma \neq 0$ and $\lambda = 0$ vs. $\lambda \neq 0$

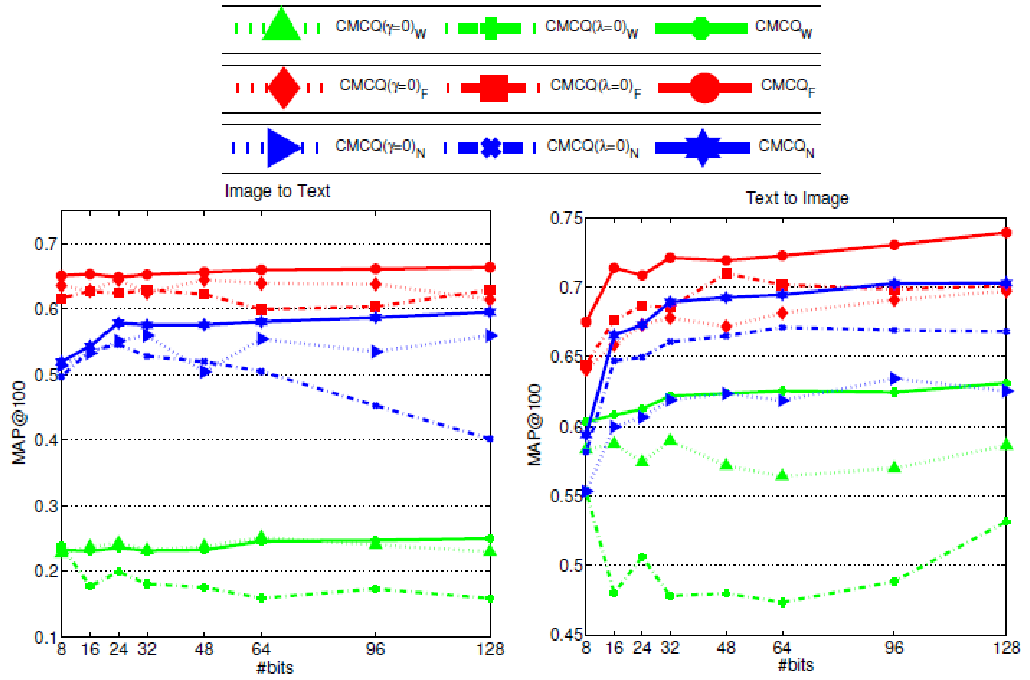


Figure 1. Illustrating the effect of the intra-document relation. The MAP is compared among CMCQ, CMCQ ($\gamma = 0$) (without correlation in the quantized space), and CMCQ ($\lambda = 0$) (without correlation in the common space) on the three datasets briefly denoted as W (Wiki), F (FLICKR25K), and N (NUS-WIDE) in the legend.

Empirical analysis

$$\mathcal{F} = \|X' - CP\|_F^2 + \|Y' - DQ\|_F^2 + \gamma\|CP - DQ\|_F^2 + \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

- The effect of intra-document correlation, i.e., $\gamma = 0$ vs. $\gamma \neq 0$ and $\lambda = 0$ vs. $\lambda \neq 0$

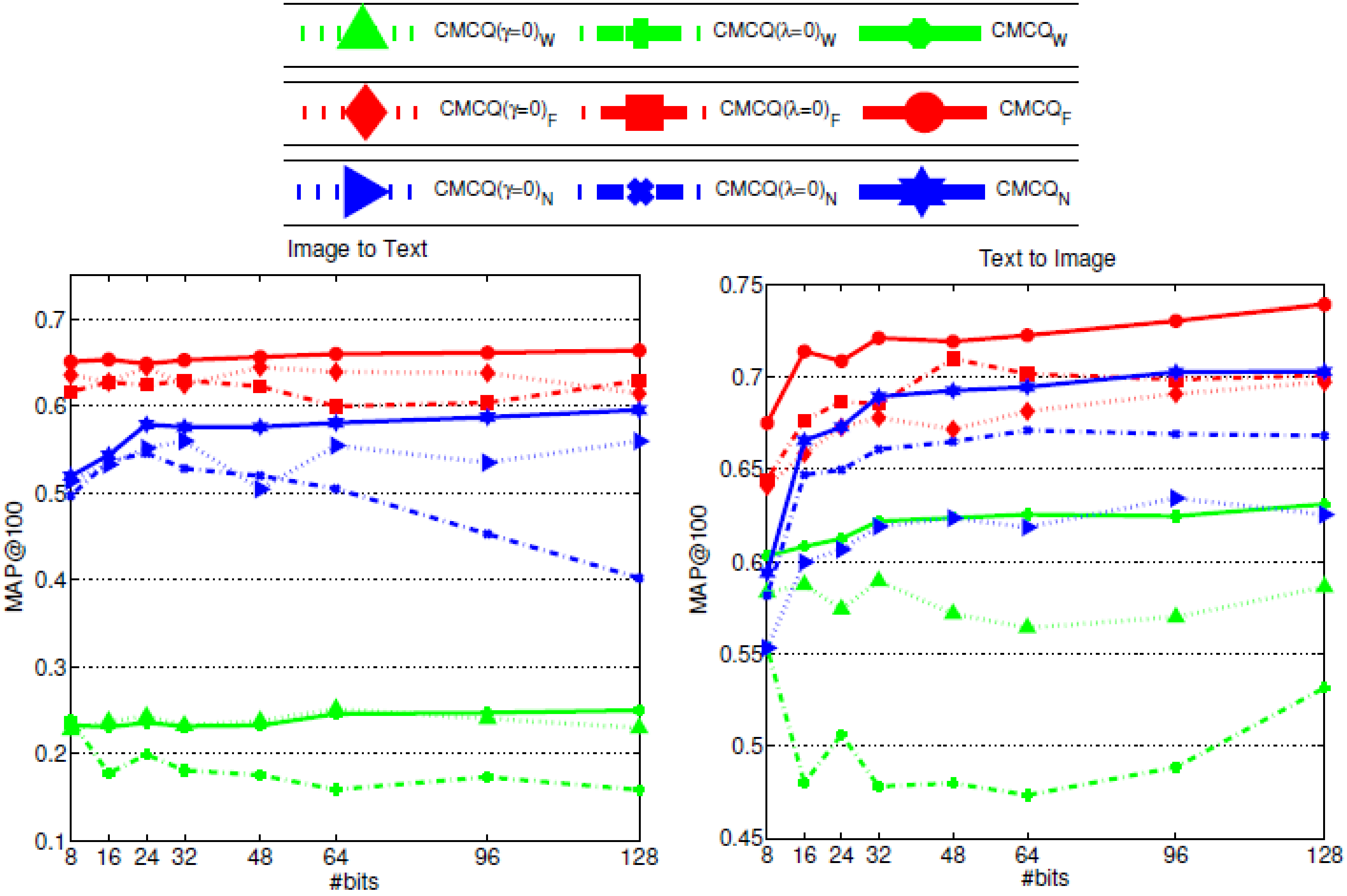


Figure 1. Illustrating the effect of the intra-document relation. The MAP is compared among CMCQ, CMCQ ($\gamma = 0$) (without correlation in the quantized space), and CMCQ ($\lambda = 0$) (without correlation in the common space) on the three datasets briefly denoted as W (Wiki), F (FLICKR25K), and N (NUS-WIDE) in the legend.

Empirical analysis

$$\mathcal{F} = \|X' - CP\|_F^2 + \|Y' - DQ\|_F^2 + \gamma\|CP - DQ\|_F^2 + \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

- The effect of dictionary, i.e., $C = D$ vs. $C \neq D$

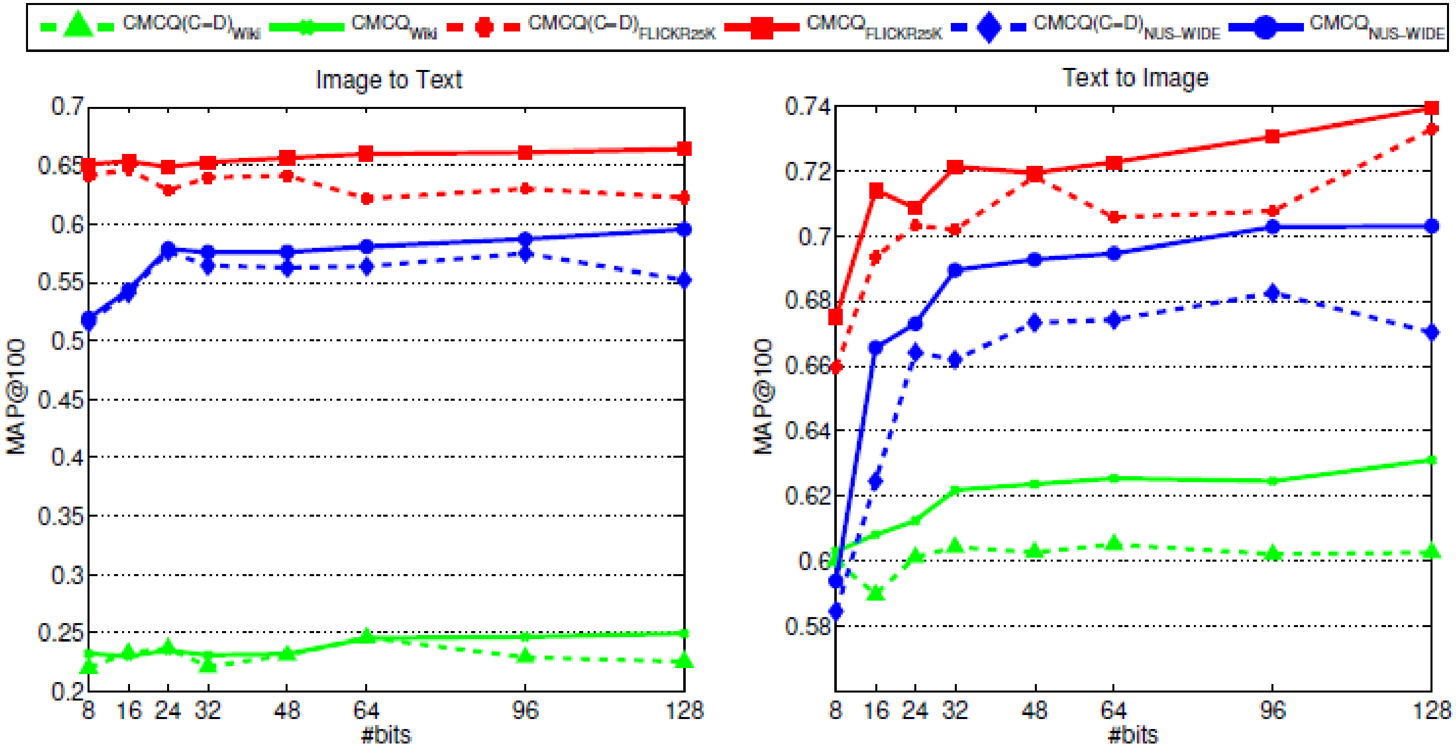


Figure 2. Illustrating the effect of the dictionary. The MAP is compared between CMCQ and CMCQ ($\mathbf{C} = \mathbf{D}$) (using one dictionary for both modalities) on the three datasets.

Empirical analysis

$$\mathcal{F} = \|X' - CP\|_F^2 + \|Y' - DQ\|_F^2 + \gamma\|CP - DQ\|_F^2 + \|X - BS\|_F^2 + \rho|S|_{11} + \eta\|Y - UY'\|_F^2 + \lambda\|Y' - RS\|_F^2$$

- Parameter sensitive analysis

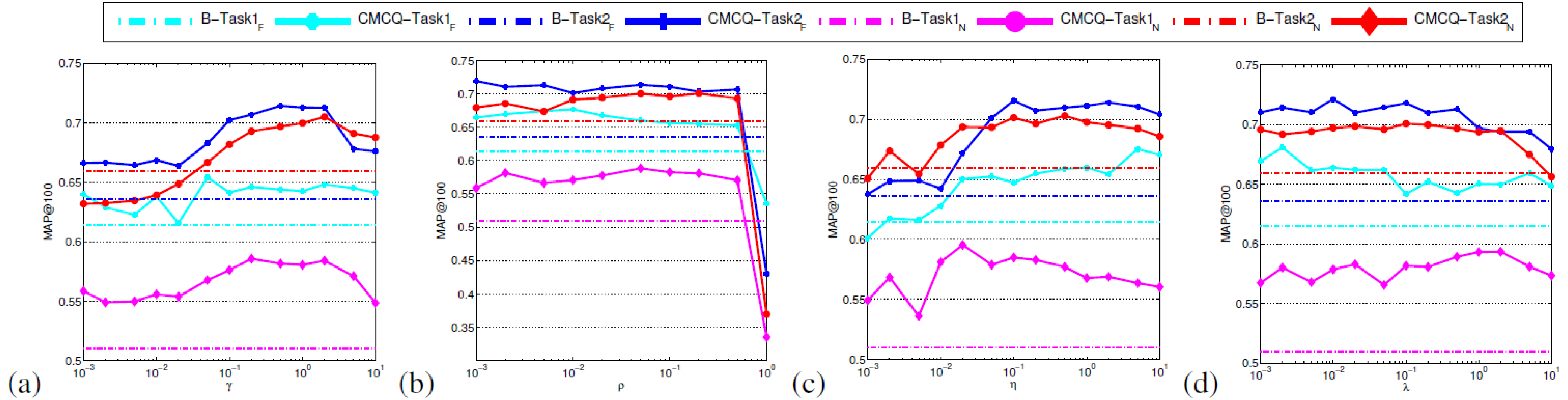


Figure 4. Parameter sensitive analysis of our algorithm with respect to (a) γ, (b) ρ, (c) η, and (d) λ over image to text (task1) and text to image (task2) on two datasets: FLICKR25K (F) and NUS-WIDE (N) with 32 bits. The dashdot line shows the best results obtained by other baseline methods and is denoted as B, e.g., B-Task1_F denotes the best baseline results over the image to text task on FLICKR25K.

Take-home message

- A quantization-based compact coding approach for cross-modal similarity search

- Learns the quantizers for both modalities by exploring the intra-document correlation

Outline

- Overview
- Neighborhood graph Search
- Quantization
  - Composite quantization
  - Supervised quantization
  - Multi-modality quantization
- Application

Application

- Bing image search
- Bing web search
- Bing cluster
- Bing Ads
- XiaoIce
- …

XiaoIce 2.0: Image-Based Chat

微信
ms-xiaoice

微博
小冰

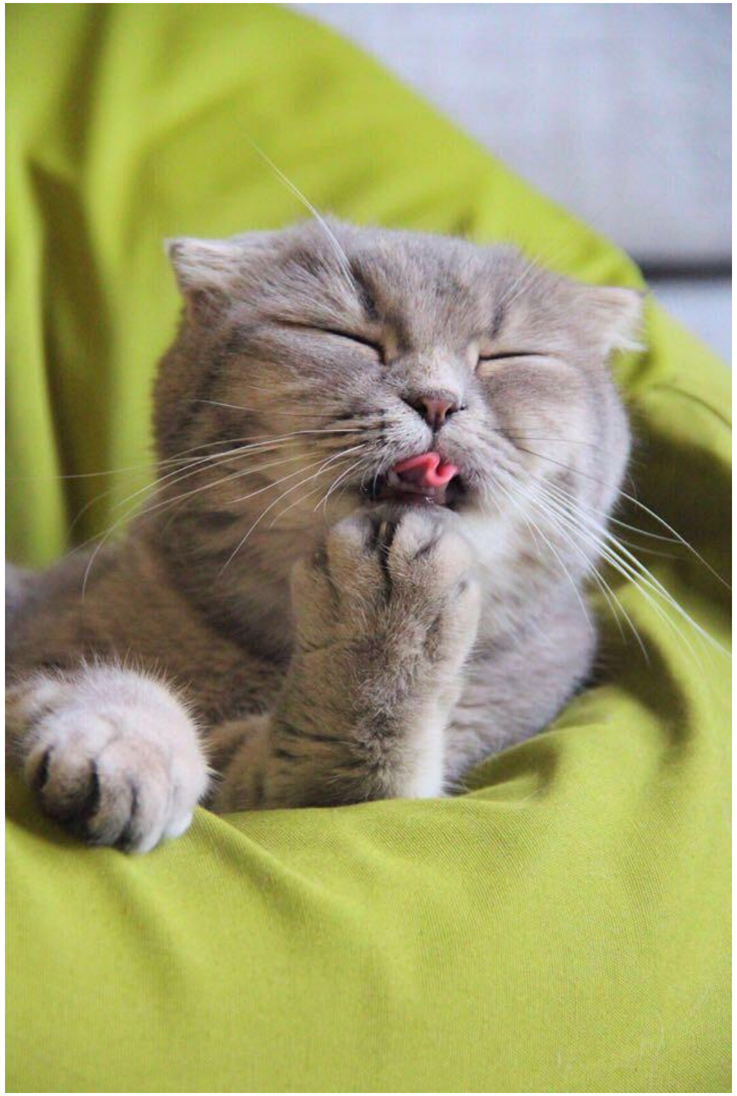
米聊
小冰


低 标准 偏高 高
肥胖类型
身高
体重
年龄
性别
9.1 %
BMI 23.1
1612
OMRON 身体脂肪测量器 HBF-306
电源
设定
测量


[SimilarImage]突破自身的秘诀是 “不要命”...,0.6254609
SIGMA
0.0
MAX. SPEED
55.84


瞅你咋地？
你瞅啥？


[SimilarImage]小眼神儿太TMD犀利了 [大笑],0.128493264




[SimilarImage] 这小舌头。。,0.6241474

“这么多 …… 很撑吧?”

“口水三千丈，肥肉日日长。”

“恭喜啊，啥时候请我们吃喜糖啊，呵呵~~”

“在顶层办公会是一种什么样的体验。”

“不是空气污染就知足吧。。。”

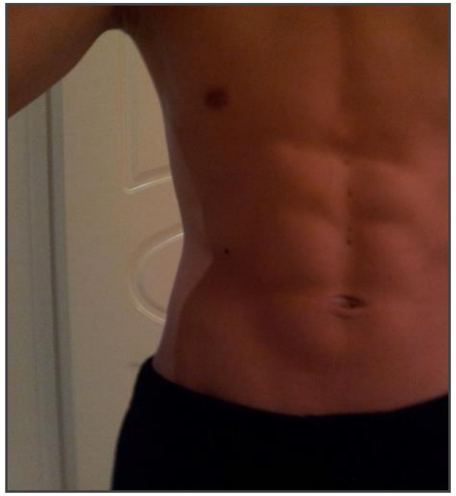
“你！又！粗！去！玩！还！不！带！着！我！”

微软小冰可以评价她看到的任何图像，她具有自己的观察角度，而不仅仅是解释图像内容

Beyond Image Recognition

“瞧这小舌头。。。”

“这朵黄色得都有些透明了，真美”



“大叔真努力！”

“不是重度污染就知足吧”

Collaborators

- Jing Wang, You Jia, Naiyan Wang
- Ting Zhang, Xiaojuan Wang
- Guo-jun Qi, Jinhui Tang, Shipeng Li
- …

Thanks!
Q&A

Neighborhood graph search:
Coming soon

CQ Code:
https://github.com/hellozting/CompositeQuantization

Homepage:
https://jingdongwang2017.github.io/