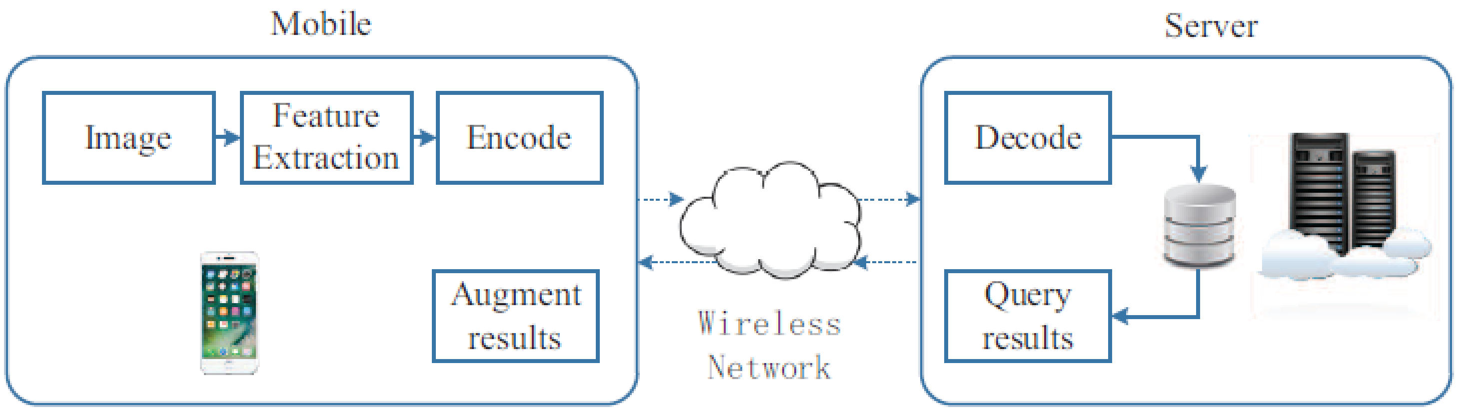
Visual Query Compression with Embedded Transforms on Grassmann Manifold

Zhaobin Zhang, Li Li and Zhu Li
Dept of Computer Science & Electrical Engineering
University of Missouri, Kansas City
lizhu@umkc.edu

Outline

- Background
- Related work
    - Compact feature descriptors, e.g., SURF, SIFT, etc.
    - MPEG: Compact Descriptor Visual Search
- Proposed algorithm
    - Hierarchical partition tree
    - Local transform optimization via Grassmann manifold
- Experiments
- Conclusions

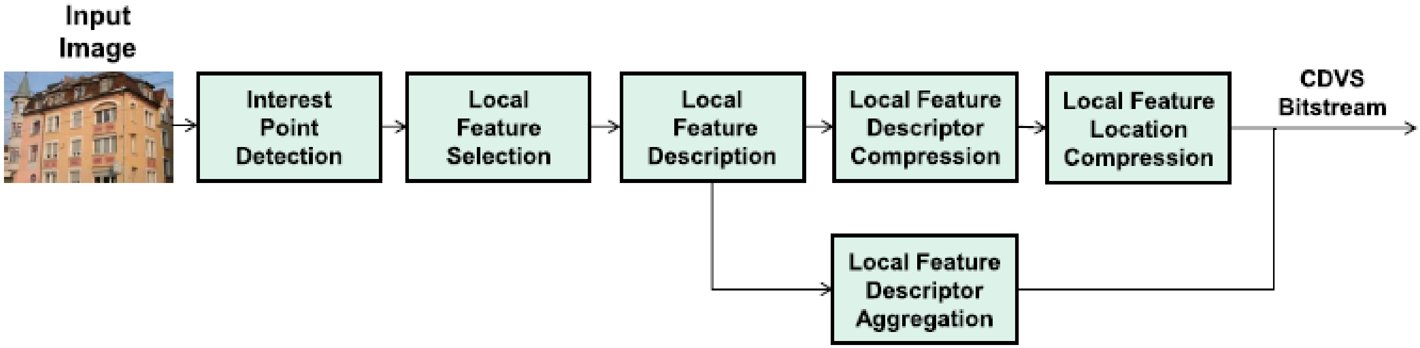
Mobile Visual Search



- Mobiles and tablet have become an indispensable part of our lives.
- Wireless has a limited bandwidth.
- Challenge: How to minimize transmission bits to reduce network latency?

Related work

- Compact feature descriptors
  - SURF (Speeded-Up Robust Features) *[Herbert, 2008]*
  - **SIFT** (Scale-Invariant Feature Transform) *[David G. Lowe, 2004]*
  - MPEG: CDVS *[2016]*



Overview of the MPEG-CDVS Standard, Ling-Yu, 2016

- High-dimension image descriptors are not efficient, e.g., 128 for SIFT.
- A single transform is not sufficient to capture all the identification information.

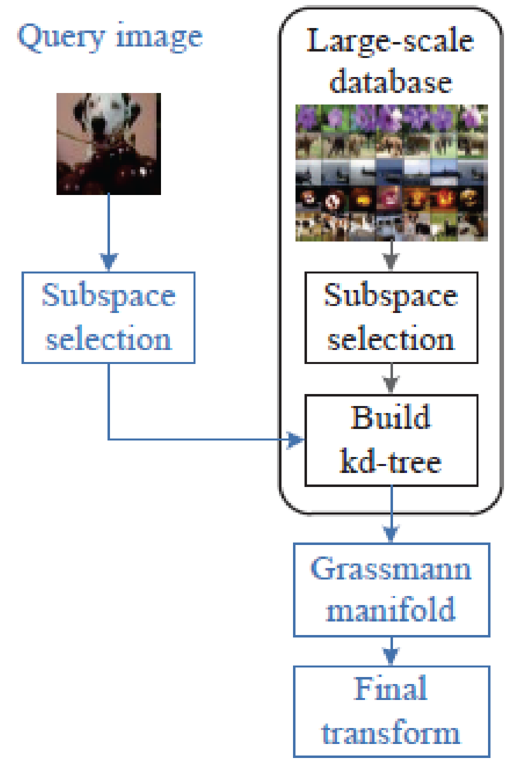
Proposed algorithm

- Hierarchical partition tree
  - Apply PCA for the current node
  - Find the median value for the variance for all dimensions
  - Split into left and right child node
- Piece-wise linear projection
  - Apply PCA at local subspaces
- Grassmann optimization
  - Merge the nodes which have the smallest Binet-Cauchy Grassmann distance
  - Stops when predefined condition is satisfied, e.g., number of transforms

Hierarchical partition tree

Given n features in d dimensions, and the height of kd-tree ht, the procedure of constructing the data partition tree are:

1. Calculate the variance of the first dimension $\{v_1, v_2, \ldots, v_n\}$.
2. Find the median of the variance m_1.
3. Split the whole space into two parts whose border is m_1.
4. For $i = 1$ (level one), calculate the variance of each dimension, find the median of the dimension which has the largest variance and split at the median.
5. Repeat step 4 until finish the partition.



Subspace distance on Grassmann manifold

- Given two subspace models A_1 and A_2, find the rotations that can max align two:
  - Rotating A_1 and A_2 in $G(p, d)$ such that they are maximally aligned

$$\max_{R_1, R_2} trace(R_1^T A_1^T A_2 R_2), \qquad \text{s.t.}\, R_1, R_2 \in O_p$$

  - Solving by SVD:

$$[S, V, D] = svd(A_1^T A_2)$$

  - The diagonal of $S = [s_1, s_2, \dots, s_p]$ are the principle angles.

$$\theta_k = cos^{-1}(s_k)$$

- Binet-Cauchy distance
  - Def: $d_{bc}(A_1, A_2) = (1 - \prod_i cos^2\theta_i)^{1/2}$

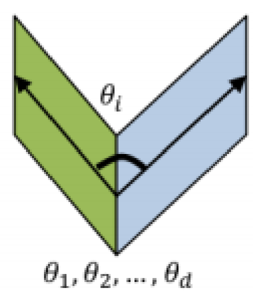
Principal angles

- The principal angles between two subspaces:
  - For A_1 and A_2 in $G(p, d)$, their principal angles are defined as

$$\cos\theta_k = \max_{u_k \in span(A_1), v_k \in span(A_2)} u_k^T v_k$$

$$s.t. \begin{cases} u_k^T u_k = 1, v_k^T v_k = 1 \\ u_k^T u_k = 0, v_k^T v_k = 0 \end{cases}$$

where $\{u_k\}$ and $\{v_k\}$ are called principal dimensions for $span(A_1)$ and $span(A_2)$



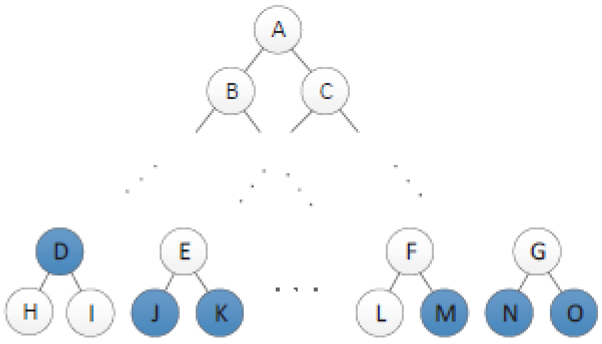
Local transform optimization

- Merge occurs when two nodes have the shortest Grassmann distance.

1. Merge on Grassmann Manifold

   - The set of existing k node
   $$N = \{n_1, n_2, \dots, n_k\}$$
   - The number of samples in each node
   $$W = \{w_1, w_2, \dots, w_k\}$$
   - The distance between two nodes
   $$D_{ij} = w_i \times d_{ic} + w_j \times d_{jc}$$
   where $d_{ic} = d_{BC}(i, c)$, c is the Lowest Common Ancestor of i and j
   - Nodes having the shortest distance will be merged.
   - Finally we get m optimal transforms.

2. Only $m \times \log k$ bits are needed to signal the transform due to the disorder characteristic of SIFT feature.



H and I are merged to D

Experiment dataset

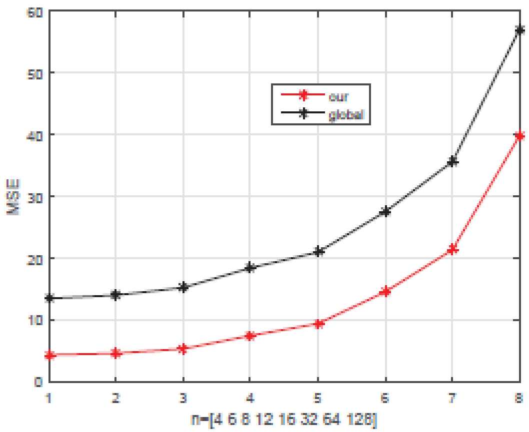
- Evaluate in CDVS dataset
- Input: SIFT features
- Output: repeatability and bitrate
- Select $\mathrm{k} = 8$ transforms from $\mathrm{n} = 127$ (kd-tree height $ht = 6$)

Table 2. A brief view of CDVS dataset

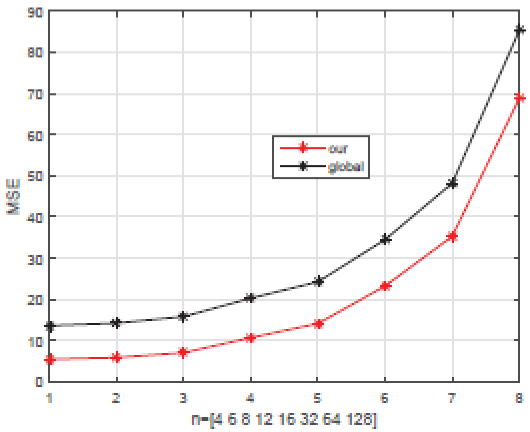
Dataset	MP	NMP
1. CDs, DVDs, books, business cards (Mixed text + graphics)	3000	29,903
2. Museum paintings	363	3639
3. Video frames	399	3999
4. Landmarks and buildings	1789	17,949
5. Common object or scenes	2549	21,307

Reconstruction Error

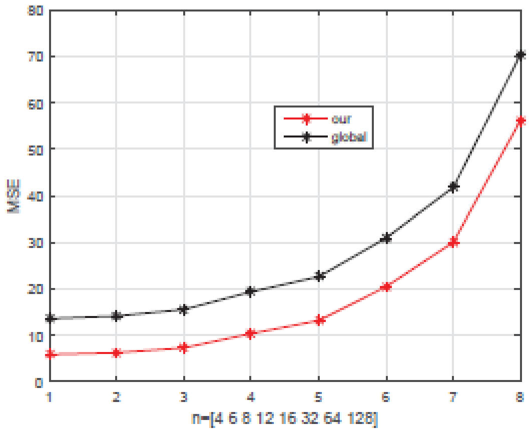
- MSE at different number of quantization bins



Kd=4



kd=6

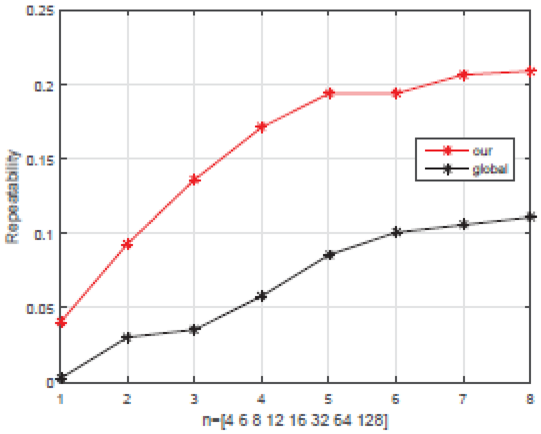


kd=8

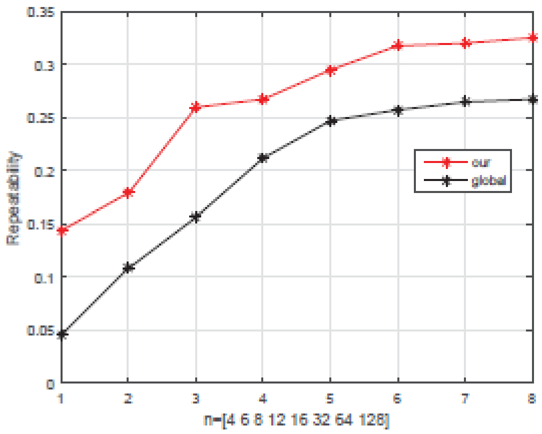
- Average reduction: 47.12%, 35.89% and 34.31% respectively at kd=4, 6 and 8.

Query Accuracy

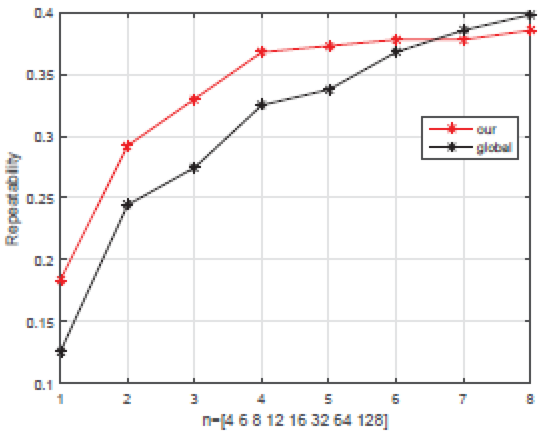
- Repeatability at different number of quantization bins



Kd=4



kd=6



kd=8

- Average improvement: 134.48%, 35.27% and 5.26% respectively at kd=4, 6 and 8.

Conclusions

- In large-scale dataset retrieval tasks, data partition tree is efficient for exploiting the local matching characteristics.
- Multiple transforms can help preserve more identification information thus improve retrieval accuracy.
- Grassmann distance which is used to measure the similarity between two orthogonal transforms can be used in feature space to optimized the projections.

Thanks!