Sparse-Coded Features for Image Retrieval

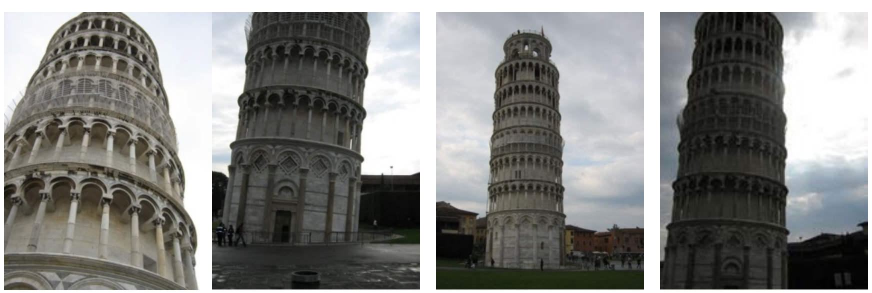
Tiezheng Ge	Qifa Ke	Jian Sun
University of Science and Technology of China	Microsoft Bing	Microsoft Research Asia

Problem Statement

- Retrieve images representing the same object/scene

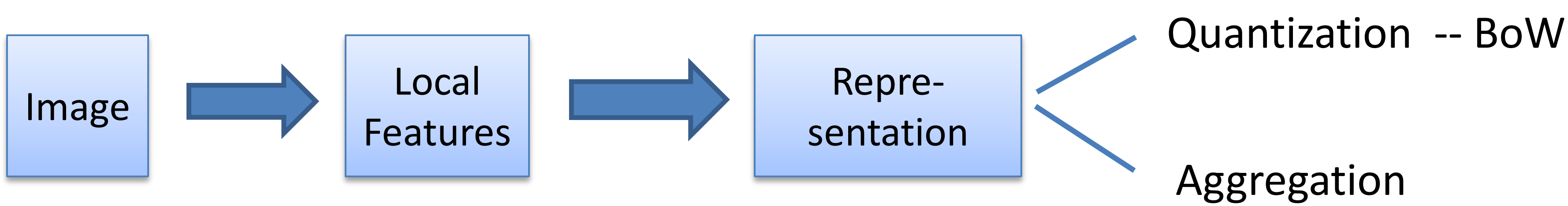
Query

Return



Previous Work

- Applying local feature

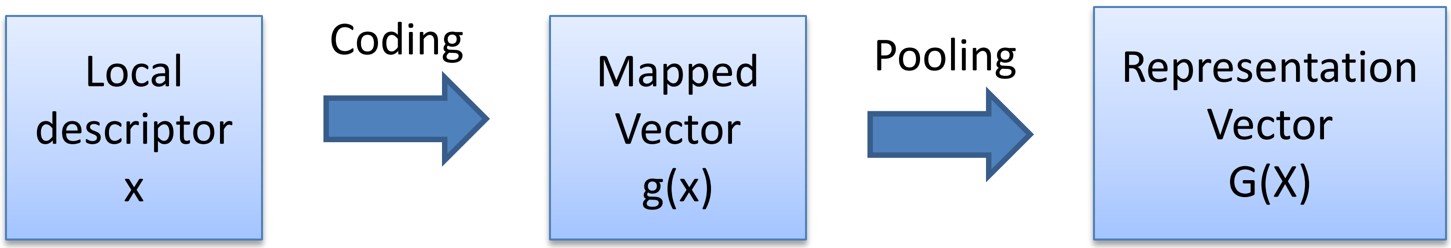


- Bag of Visual Word(BoW) [Sivic & Zisserman 03]
  - Large (hierarchical) vocabularies [Nister & Stewenius 03]
  - Hamming embed [Jegou et al 08]
  - Geometry preserving [Zhang et al 11]
  - Query expansion [Chum et al 07, Arandjelovic & Zisserman 12]

- Aggregation based method
  - VLAD [Jegou et al 10, Arandjelovic & Zisserman 13]
  - Fisher Kernel [Perronnin et al 07, 08, 10, Douze et al 11]

Previous Work

- Aggregation formulation: Coding & Pooling



- Try more coding method!

Sparse coding for image search

- For image classification:
  ScSPM [Yang et al 09]   LLC [Wang et al 10]

- Formulation:

codebook   sparse code

$$\min_{u} ||\boldsymbol{x} - \boldsymbol{u}V||_2^2 + \lambda ||\boldsymbol{u}||_1$$
$$\text{s.t. } \mathbf{u} \succcurlyeq 0$$

➢ Encoding: $g(\boldsymbol{x}) = \boldsymbol{u}$

➢ Pooling: max pooling

Sparse coding for image search

- Slightly differs from classification:

➢ Interest/key points, not dense sampled ones.

➢ No SPM (shift and rotation)

Sparse coding for image search

- Active

Sparse coding for image search

- Active

$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad 0.5]$

$x_2 \;\; g(x_2) = [0.3 \quad 0.6 \quad 0.2]$

$x_3 \;\; g(x_3) = [0 \quad 0.4 \quad 0\ ]$

x_1

x_2

x_3

Sparse coding for image search

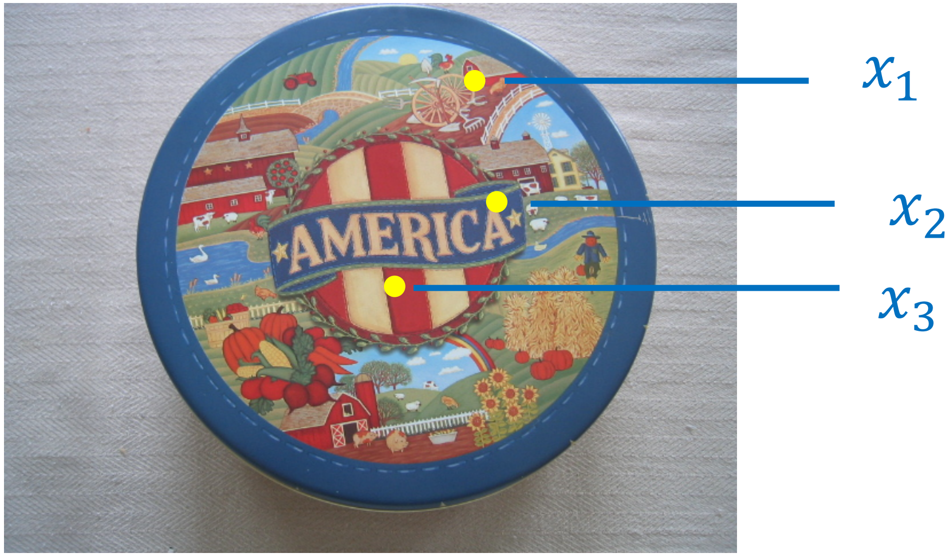
- Active

$x_1 \quad g(x_1) = [0.1 \quad 0 \quad 0.5]$

$x_2 \quad g(x_2) = [0.3 \quad 0.6 \quad 0.2]$

$x_3 \quad g(x_3) = [0 \quad 0.4 \quad 0 \;]$

Pooling $G(X) = [0.3 \quad 0.6 \quad 0.5]$



Sparse coding for image search

- Active

$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad 0.5]$

$x_2 \;\; g(x_2) = [0.3 \quad 0.6 \quad 0.2]$

$x_3 \;\; g(x_3) = [0 \quad 0.4 \quad 0\;]$

Pooling $G(X) = [0.3 \quad 0.6 \quad 0.5]$

x_1, x_2 are active

G(x) depends solely on x_1, x_2

Sparse coding for image search

- Active

$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad \color{red}{0.5}]$

$x_2 \;\; g(x_2) = [\color{red}{0.3} \quad \color{red}{0.6} \quad 0.2]$

$x_3 \;\; g(x_3) = [0 \quad 0.4 \quad 0\;]$

Pooling $G(X) = [\color{red}{0.3} \quad \color{red}{0.6} \quad \color{red}{0.5}]$

x_1, x_2 are active

G(x) depends solely on x_1, x_2

x_1

x_2

Active points

Sparse coding for image search

- Co-active

$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad 0.5]$

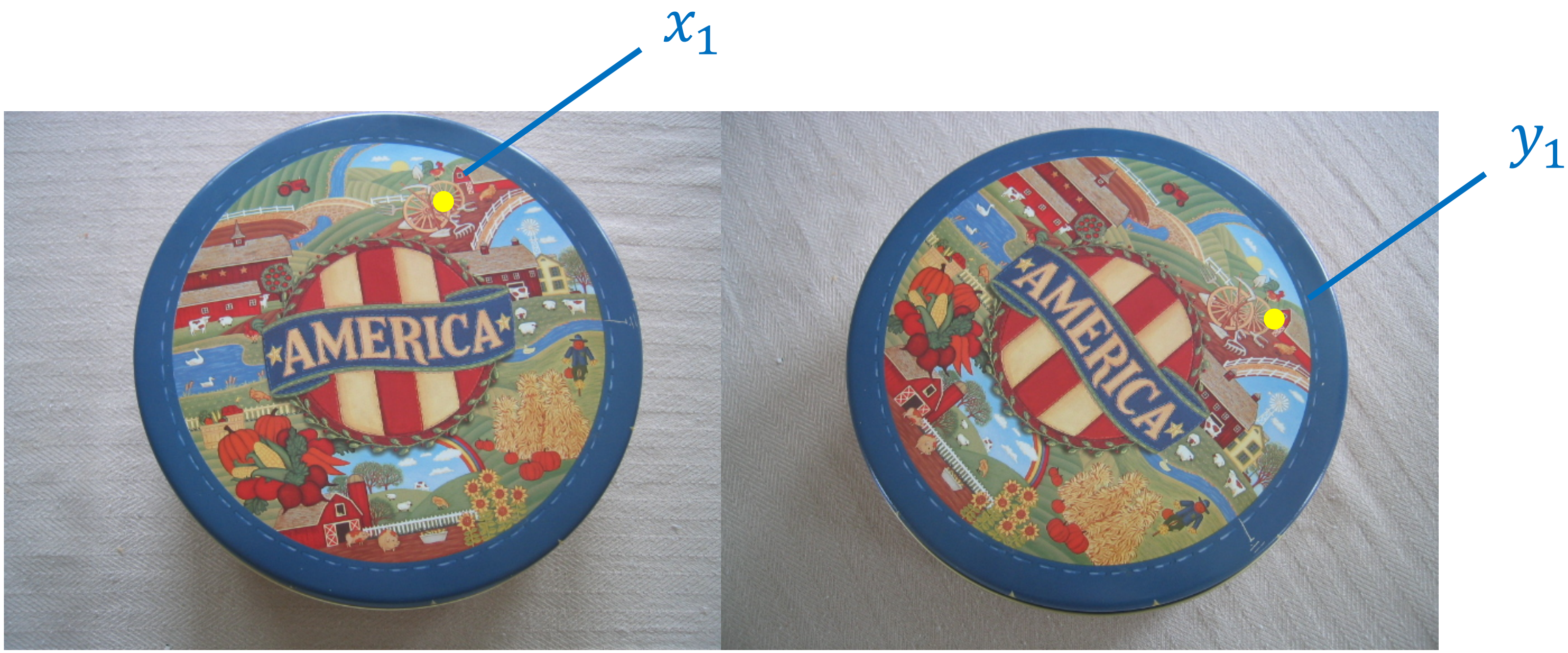
x_1

Sparse coding for image search

- Co-active

$$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad 0.5]$$

$$y_1 \;\; g(y_1) = [0.2 \quad 0.1 \quad 0.6]$$



Sparse coding for image search

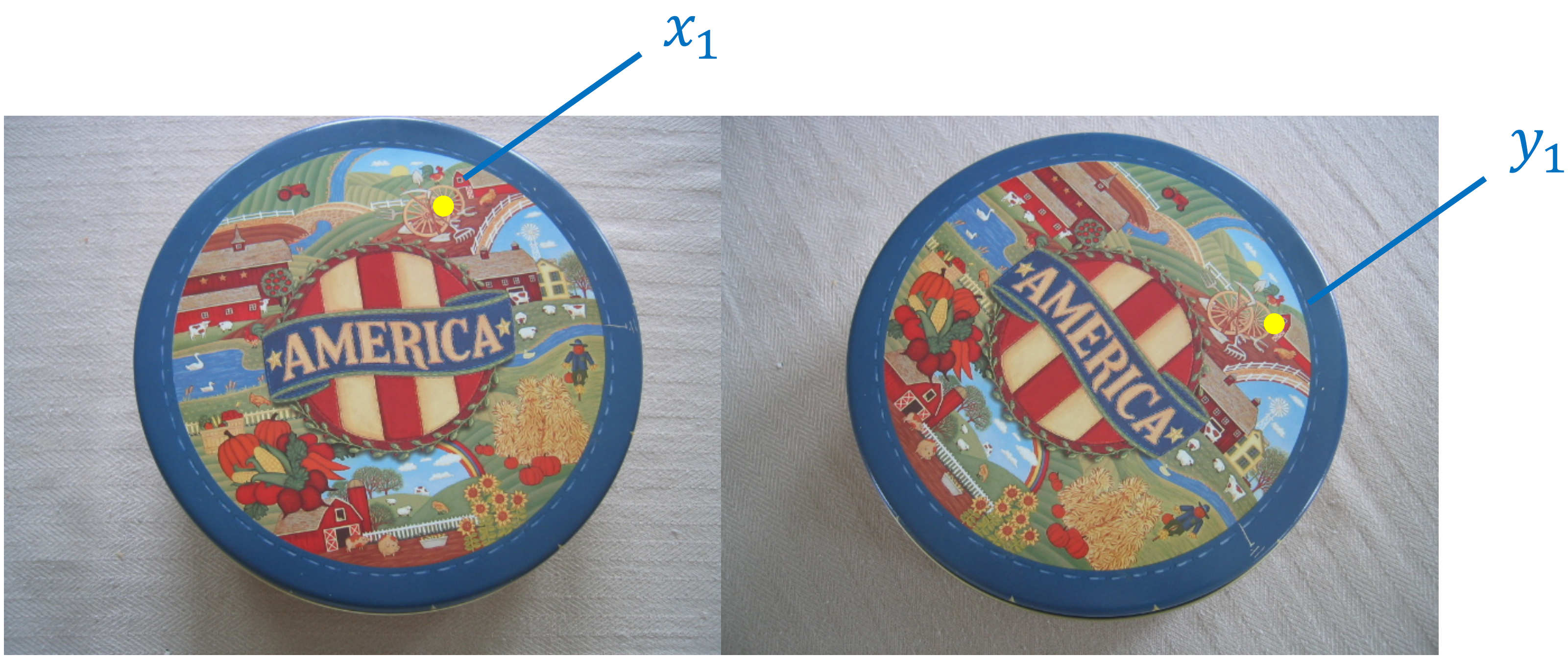
- Co-active

$x_1 \;\; g(x_1) = [0.1 \quad 0 \quad 0.5]$

$y_1 \;\; g(y_1) = [0.2 \quad 0.1 \quad 0.6]$

x_1, y_1 are co-active

Should be true active pair



Sparse coding for image search

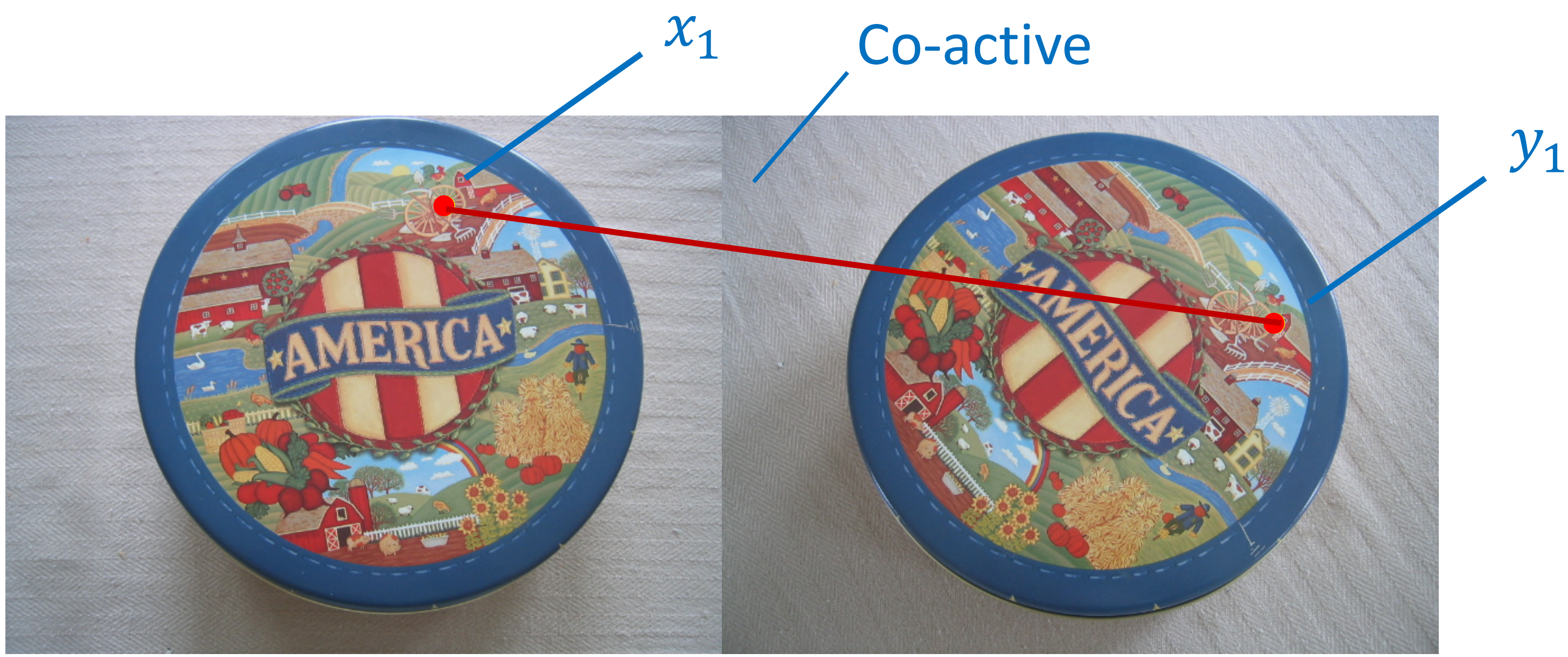
- Co-active

$$x_1 \quad g(x_1) = [0.1 \quad 0 \quad 0.5]$$

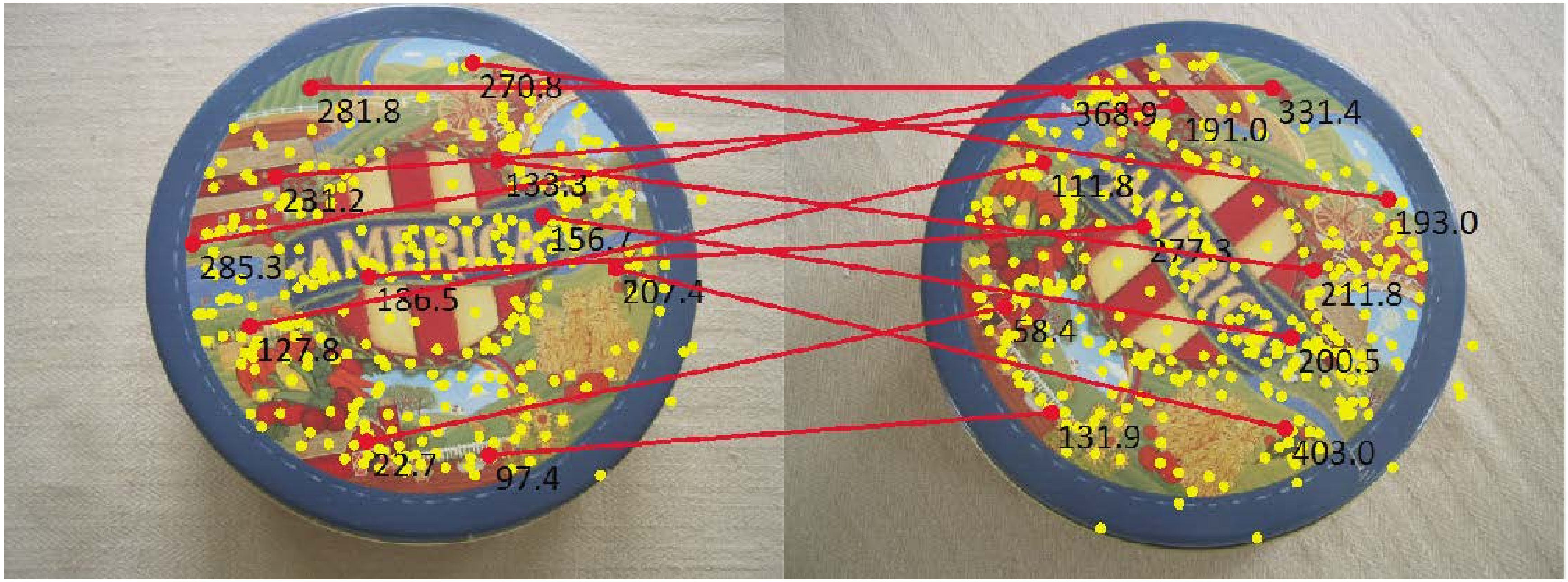
$$y_1 \quad g(y_1) = [0.2 \quad 0.1 \quad 0.6]$$

x_1, y_1 are co-active

Should be true active pair

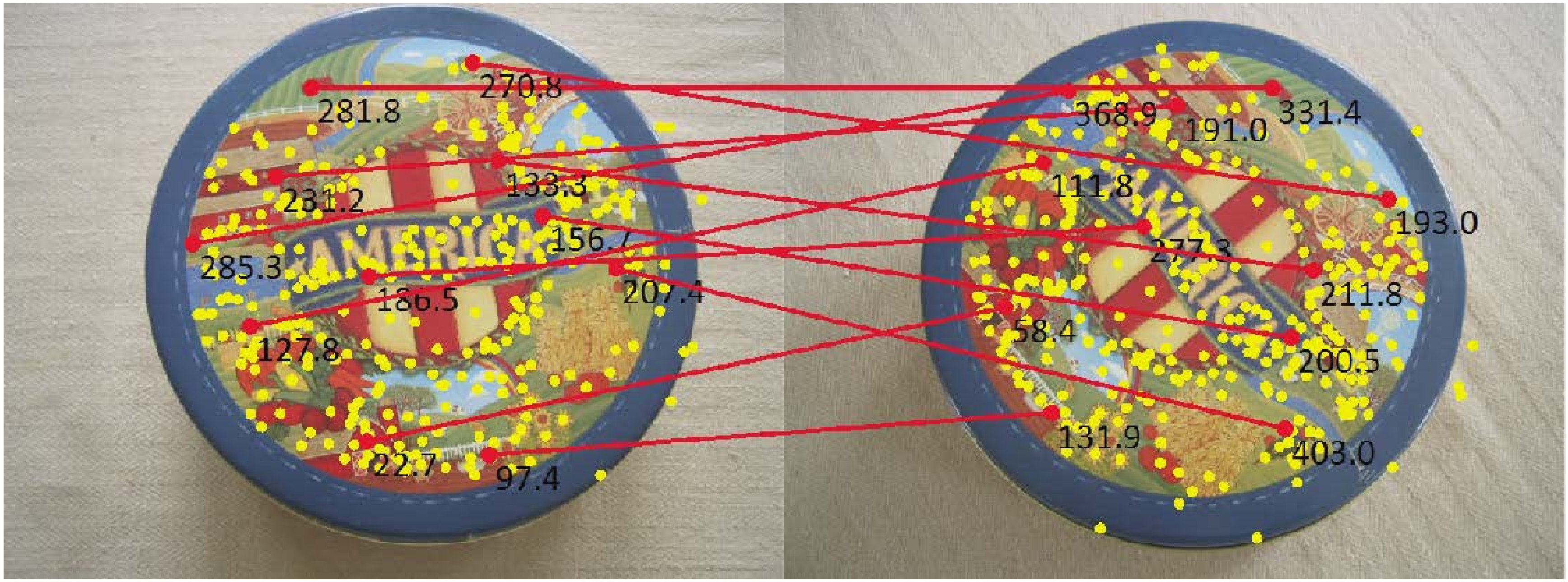


Sparse coding for image search



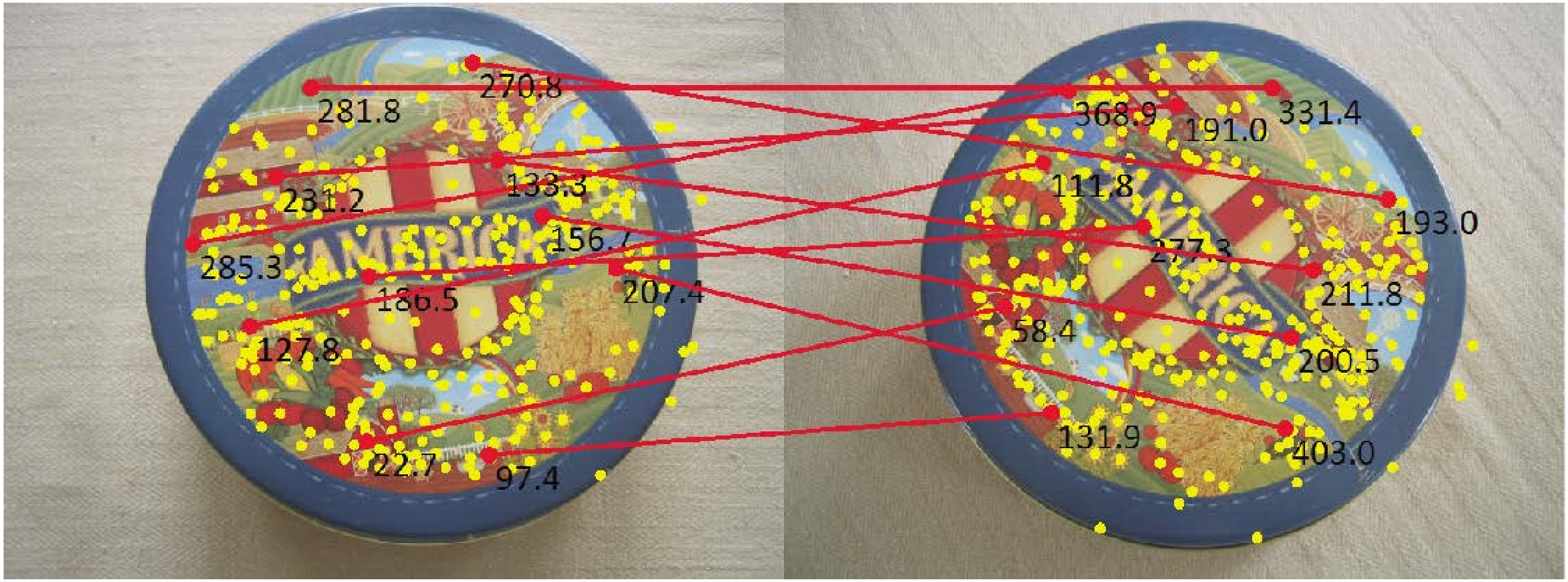
Sparse coding for image search

- Most descriptors are active



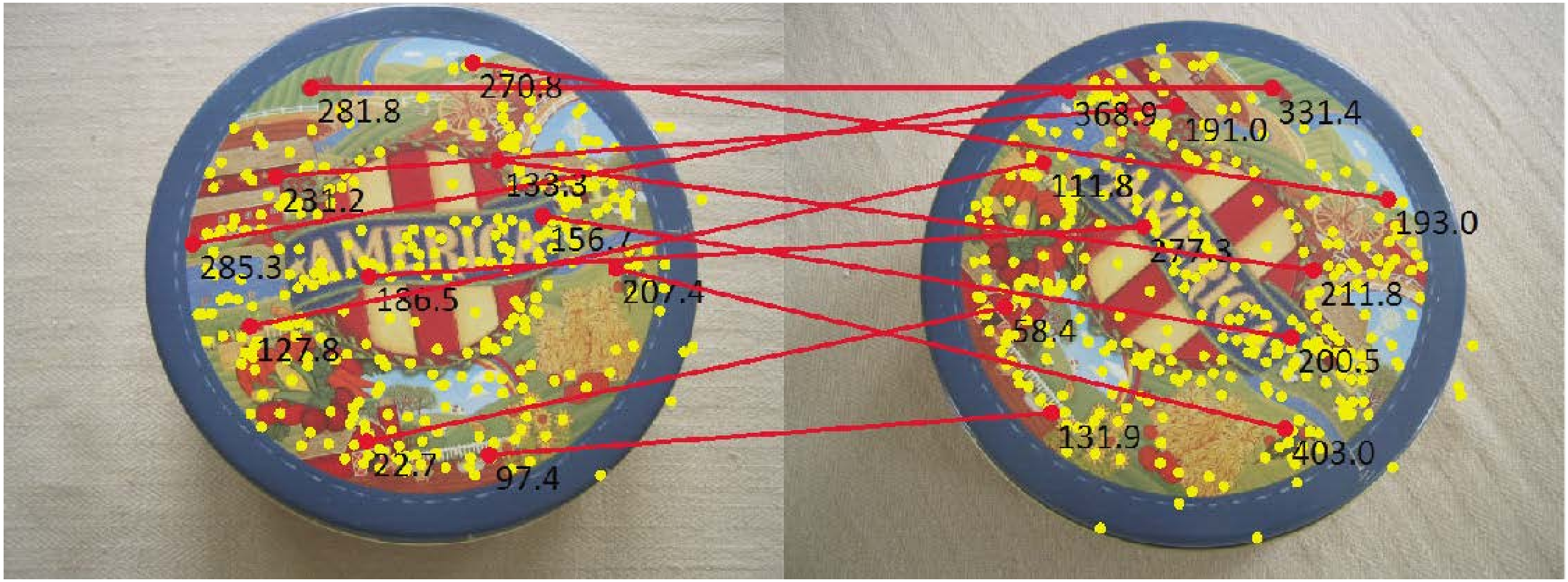
Sparse coding for image search

- Most descriptors are active
- Many correct corresponding pairs!



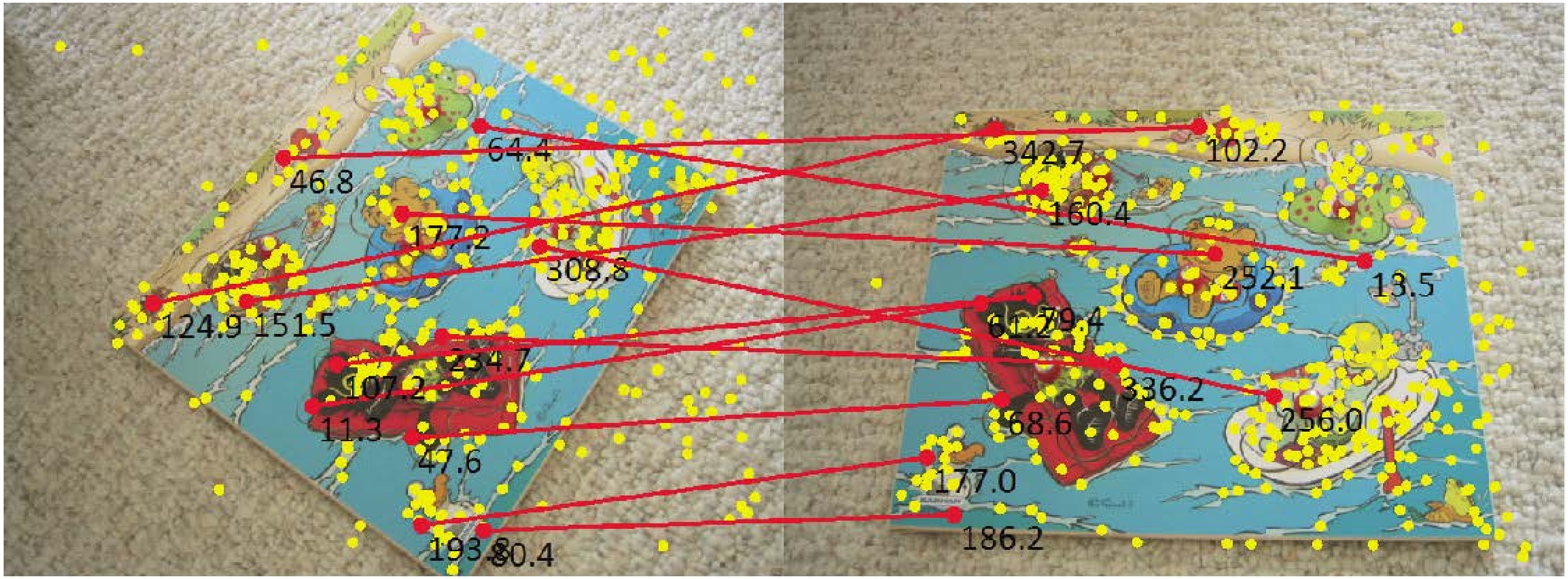
Sparse coding for image search

- Most descriptors are active
- Many correct corresponding pairs!
- Sparse coding is a feature matcher



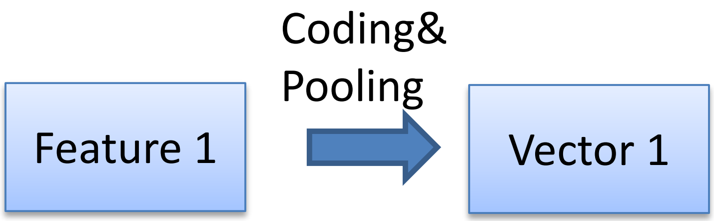
Sparse coding for image search

- Another example:

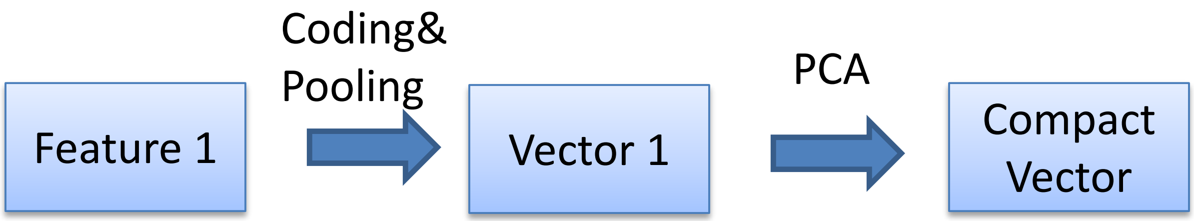


Multiple feature

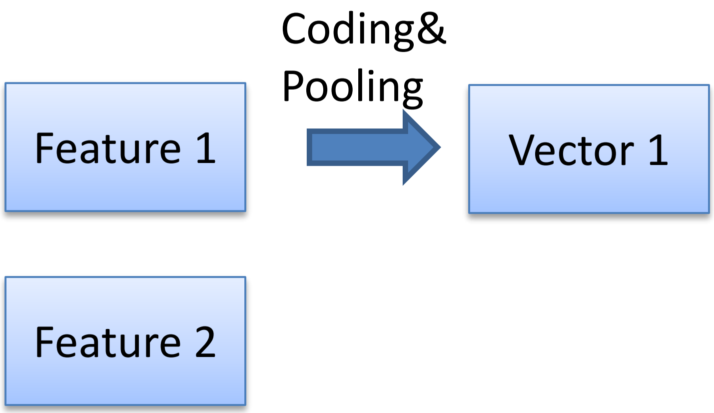
Multiple feature



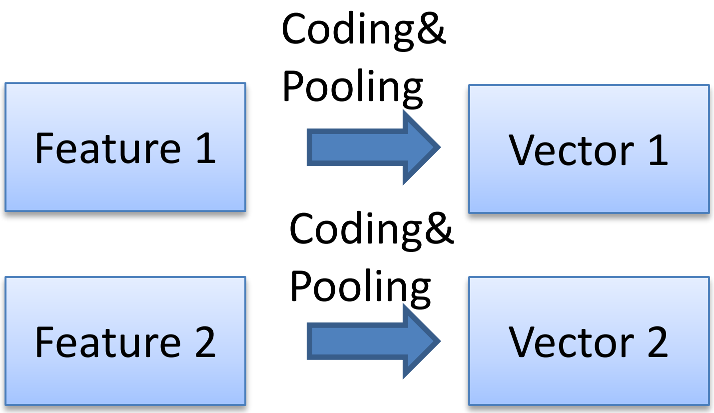
Multiple feature



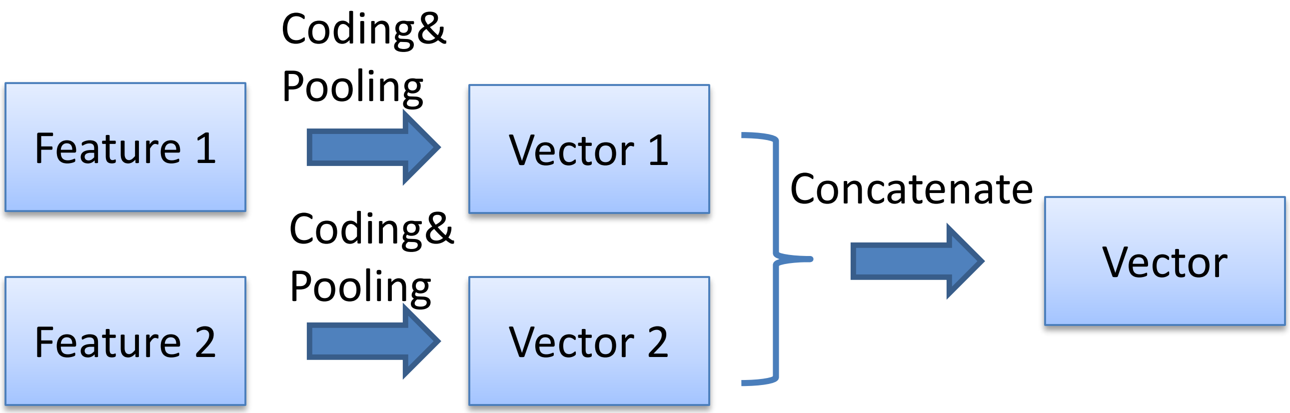
Multiple feature



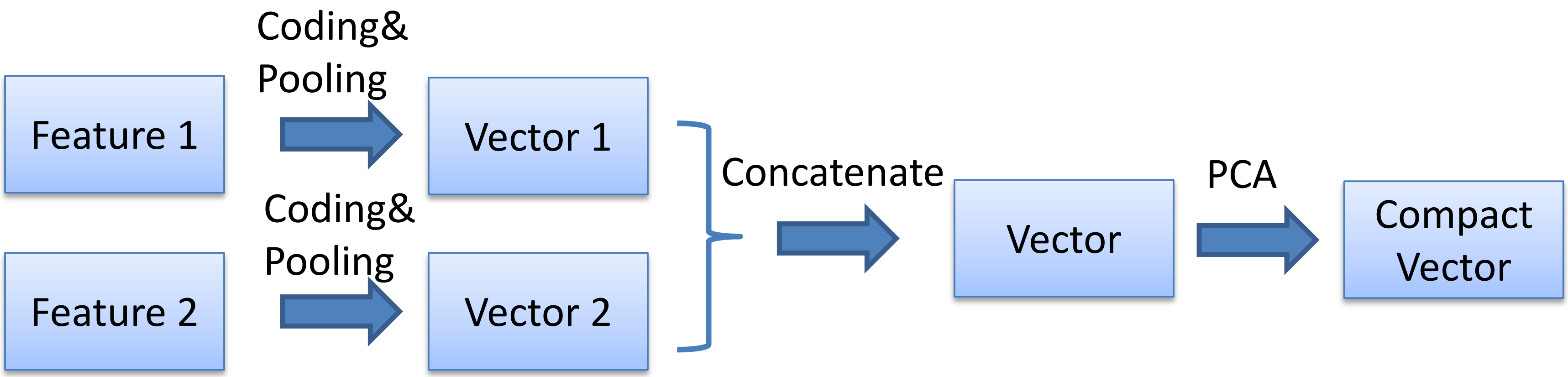
Multiple feature



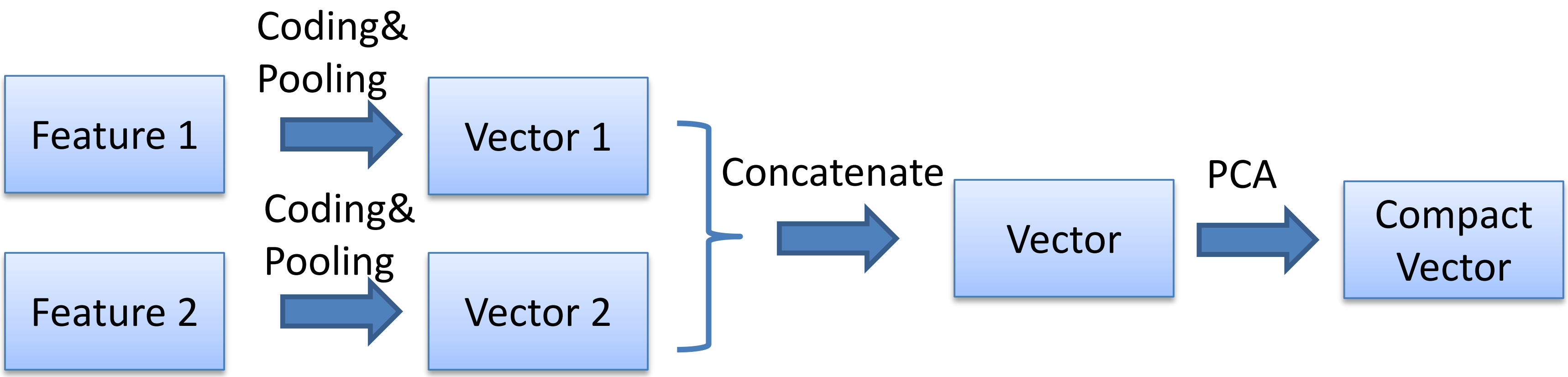
Multiple feature



Multiple feature

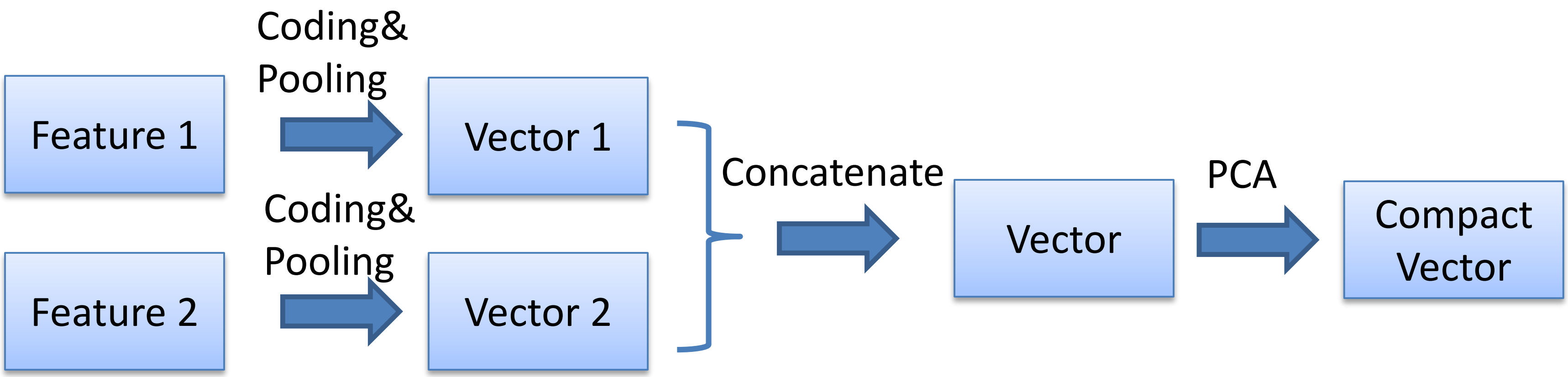


Multiple feature



- Do not hurt memory & computation efficiency

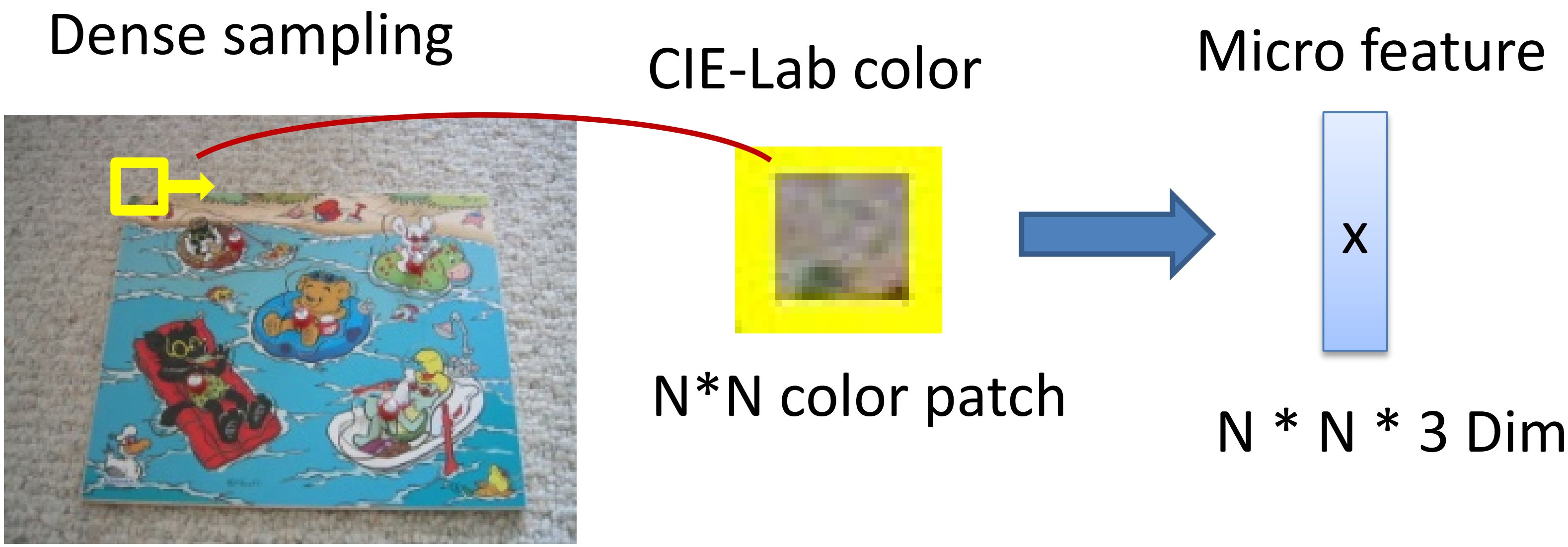
Multiple feature



- Exploit feature combination:
- ➢ Detector: Harris corner, LOG (Laplacian of Gaussian)
- ➢ Descriptor: SIFT, DAISY
- ➢ Best configuration: Harris-DAISY(HD) + LOG-SIFT(LS)

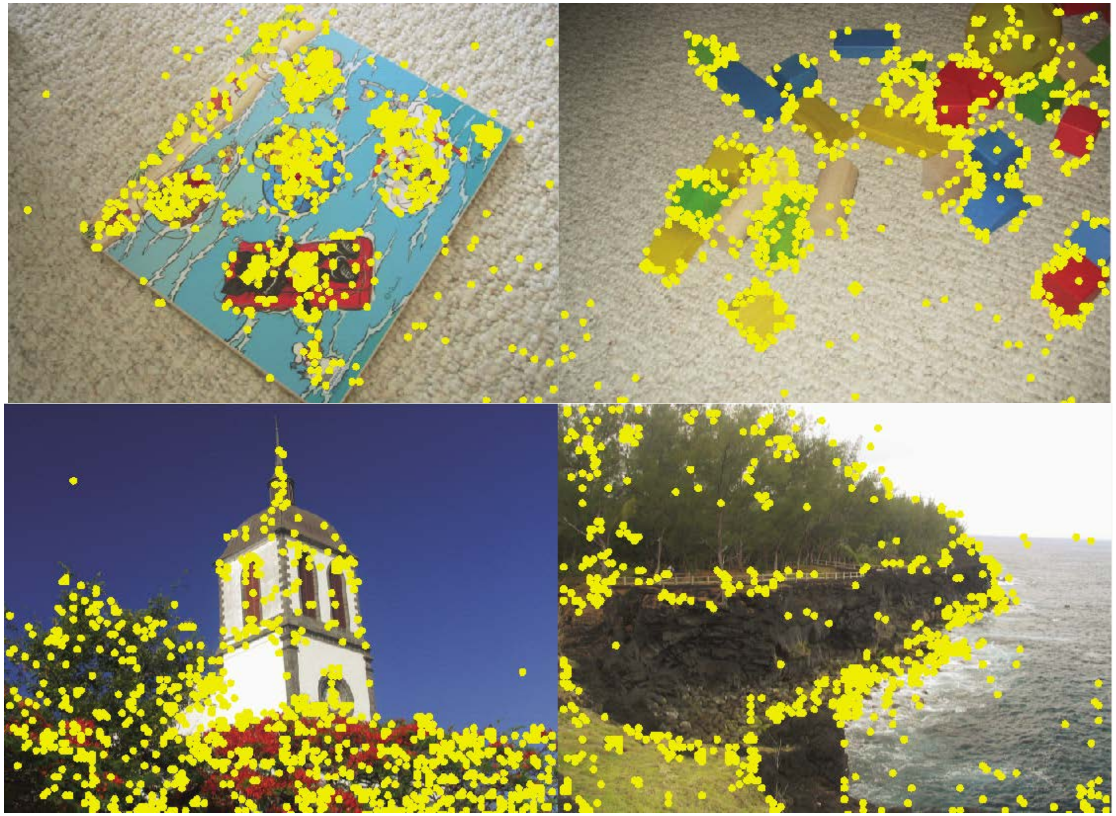
Sparse-coded micro feature

- Color features
- Inspired by bag-of-colors(BOC) [Wengert et al 11]



Sparse-coded micro feature

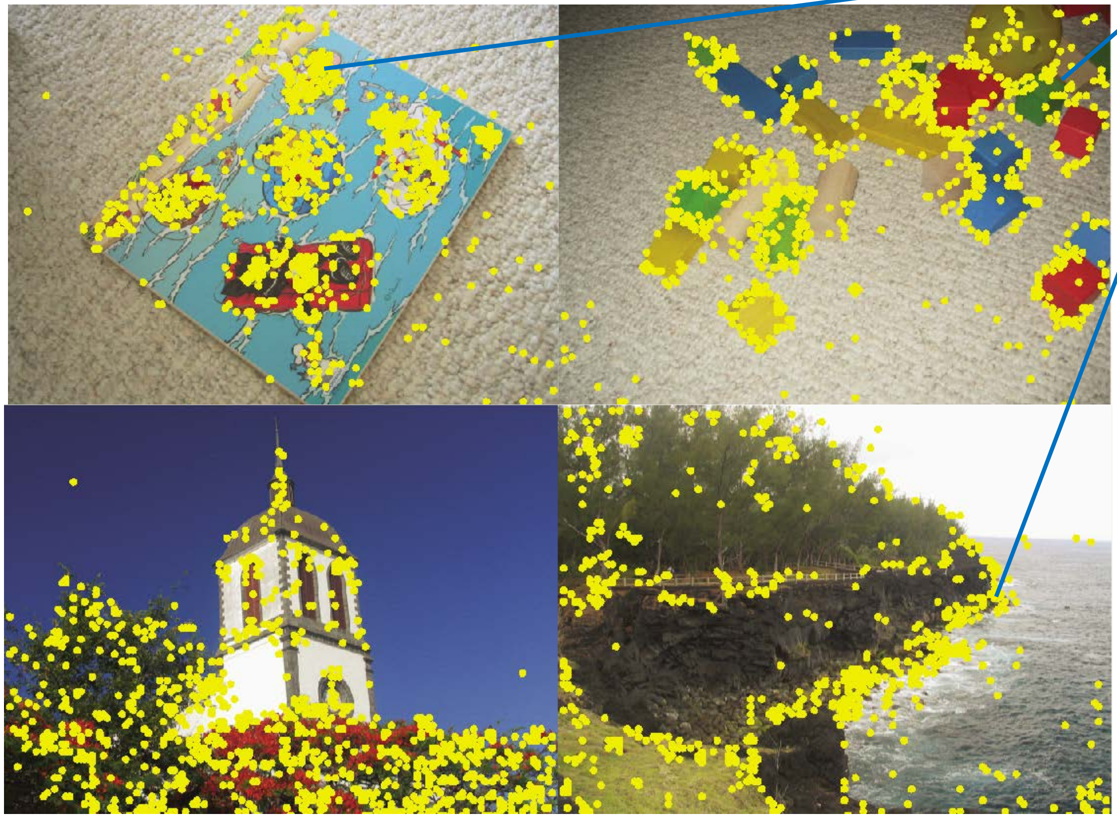
- Active points



Sparse-coded micro feature

- Active points

Focus on
distinctive points

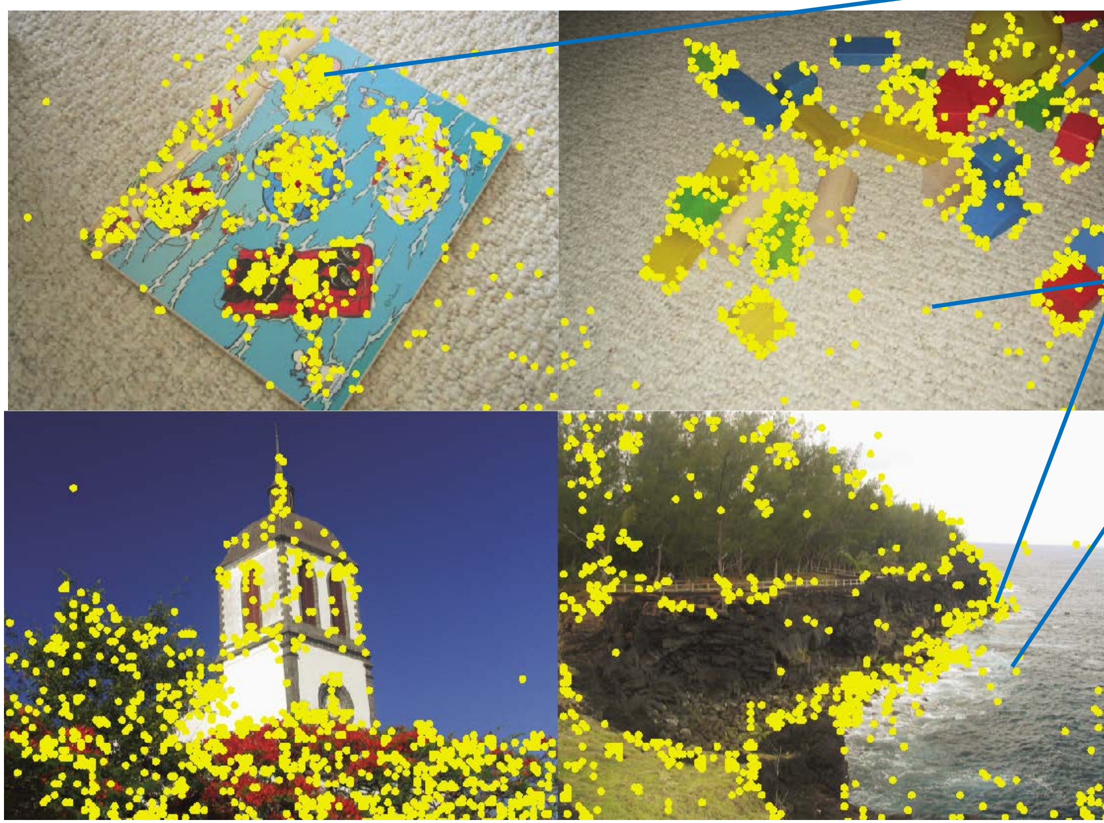


Sparse-coded micro feature

- Active points

Focus on distinctive points

patches in smooth region



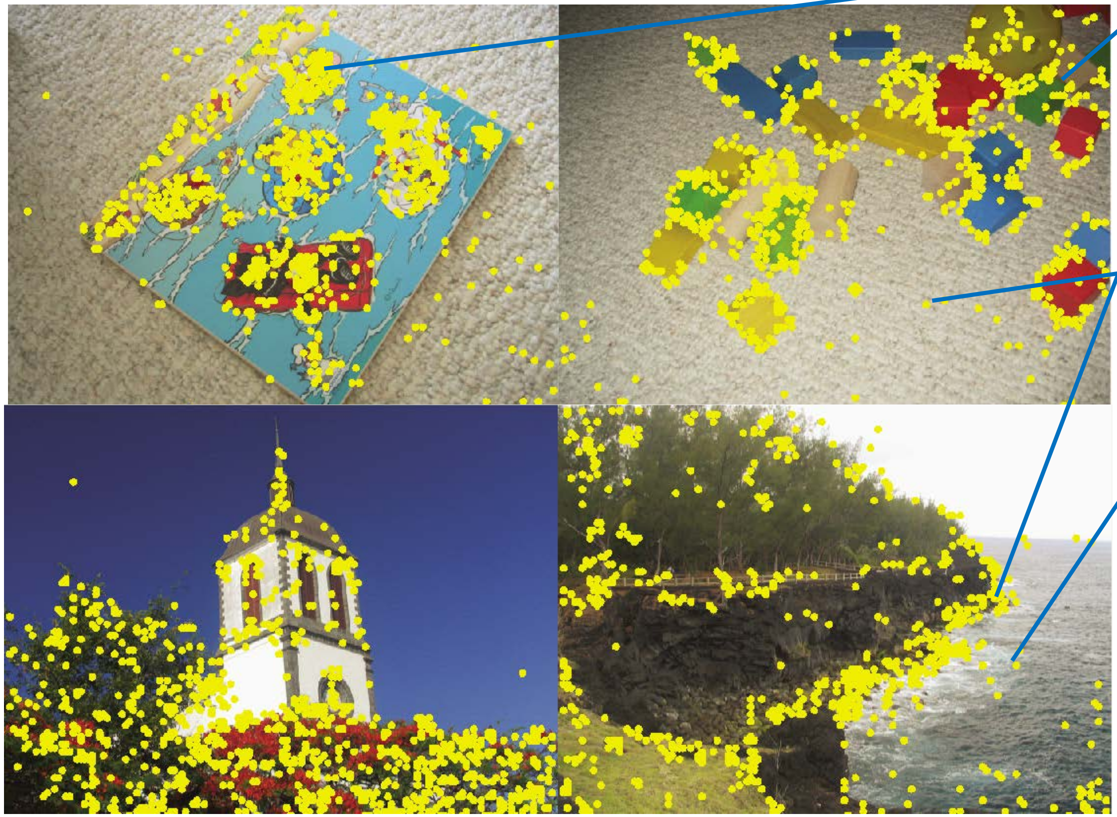
Sparse-coded micro feature

- Active points

Focus on distinctive points

patches in smooth region

SC is a filter!



Single feature comparison

- Compare SC framework with VLAD and Fisher Kernel using the same local feature --- Harris-DAISY(HD, 104Dim)
- Datasets: INREA Holidays(mAP) & UKB(score/4)

Approaches	**Dim**	**Holidays(mAP)**			**UKB(score/4)**		
		D	->128	->64	D	->128	->64
Fisher(HD)	6656	0.566	0.530	0.499	3.15	3.09	3.02
VLAD(HD)	6656	0.559	0.527	0.496	3.14	3.05	3.00
SC(HD)	5000	0.599	0.525	0.505	3.40	3.29	3.21

Multiple features

- Adding more features
- HD – Harris-Daisy   LS – LOG-SIFT   Micro – Micro feature

Approaches	**Dim**	**Holidays(mAP)**		**UKB(score/4)**	
		D	->128	D	->128
SC(HD)	5000	0.599	0.525	3.40	3.29
SC(HD+LS)	10000	0.664	0.599	3.50	3.45
SC(HD+LS+Micro)	11024	0.767	0.727	3.76	3.67

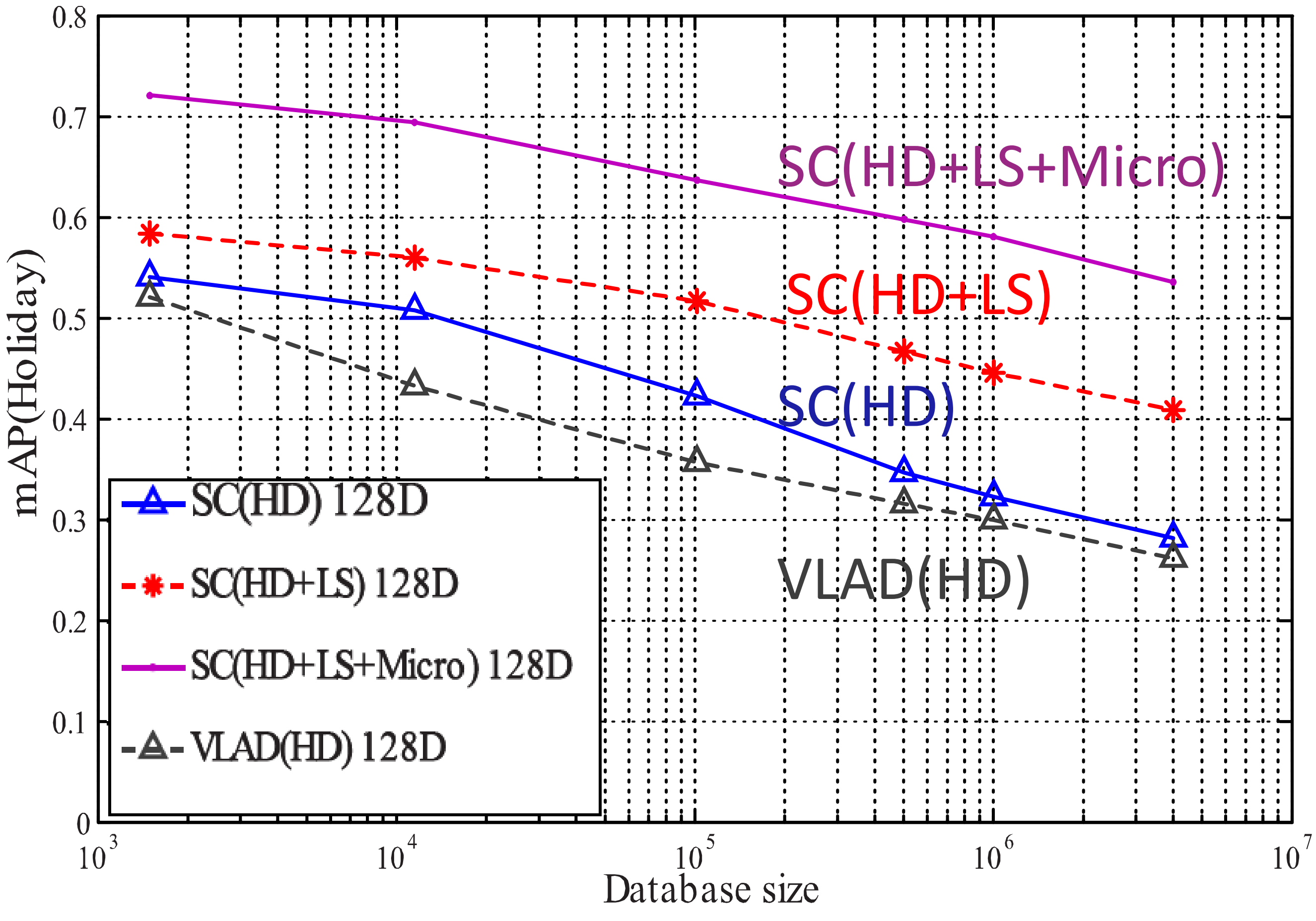
Multiple features

- Adding more features
- HD – Harris-Daisy   LS – LOG-SIFT   Micro – Micro feature

Approaches	Dim	Holidays(mAP)		UKB(score/4)	
		D	->128	D	->128
SC(HD)	5000	0.599	0.525	3.40	3.29
SC(HD+LS)	10000	0.664	0.599	3.50	3.45
SC(HD+LS+Micro)	11024	0.767	0.727	3.76	3.67

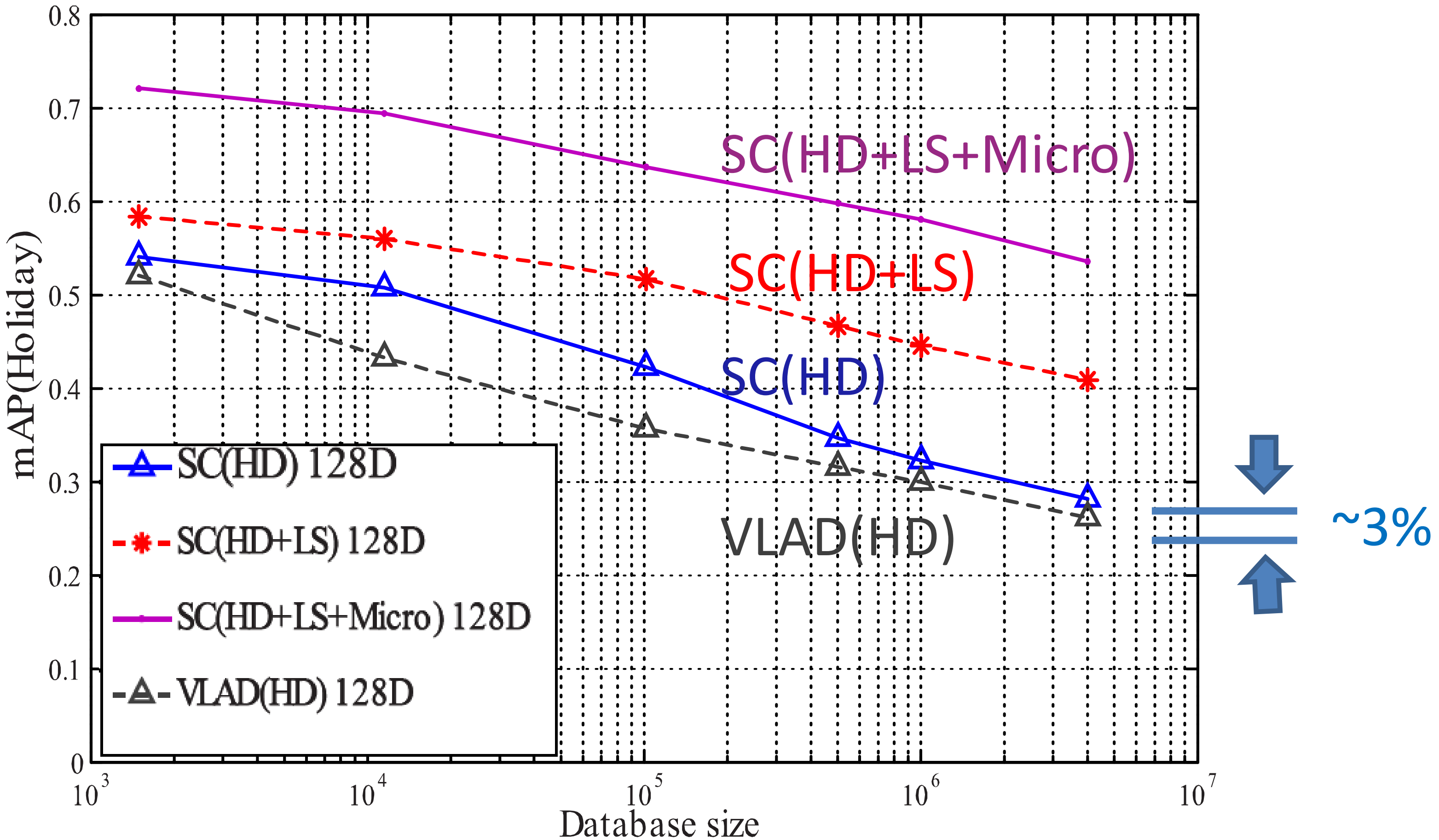
Scalability study

- Holidays + 4M images from *Flickr*.



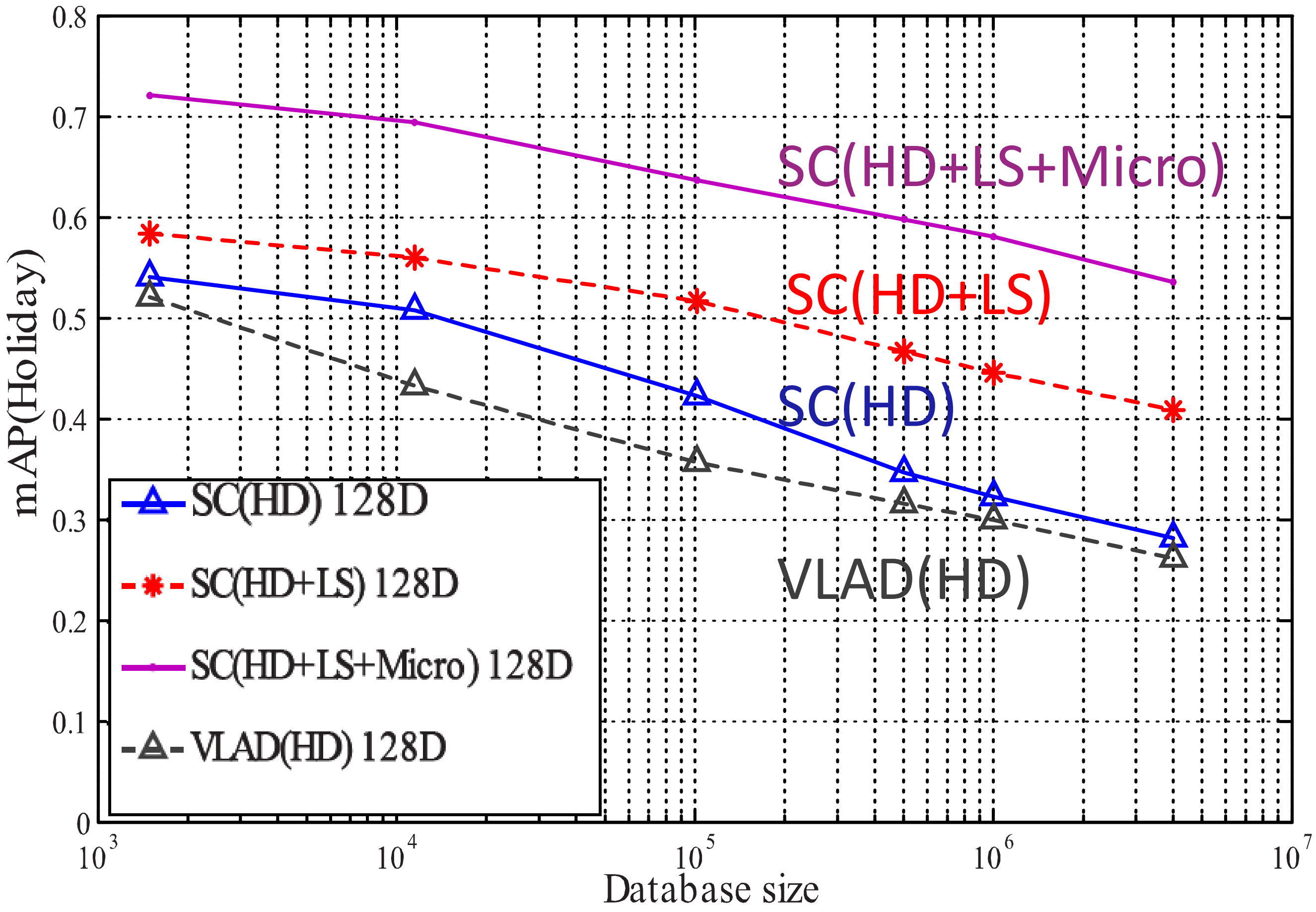
Scalability study

- Holidays + 4M images from *Flickr*.



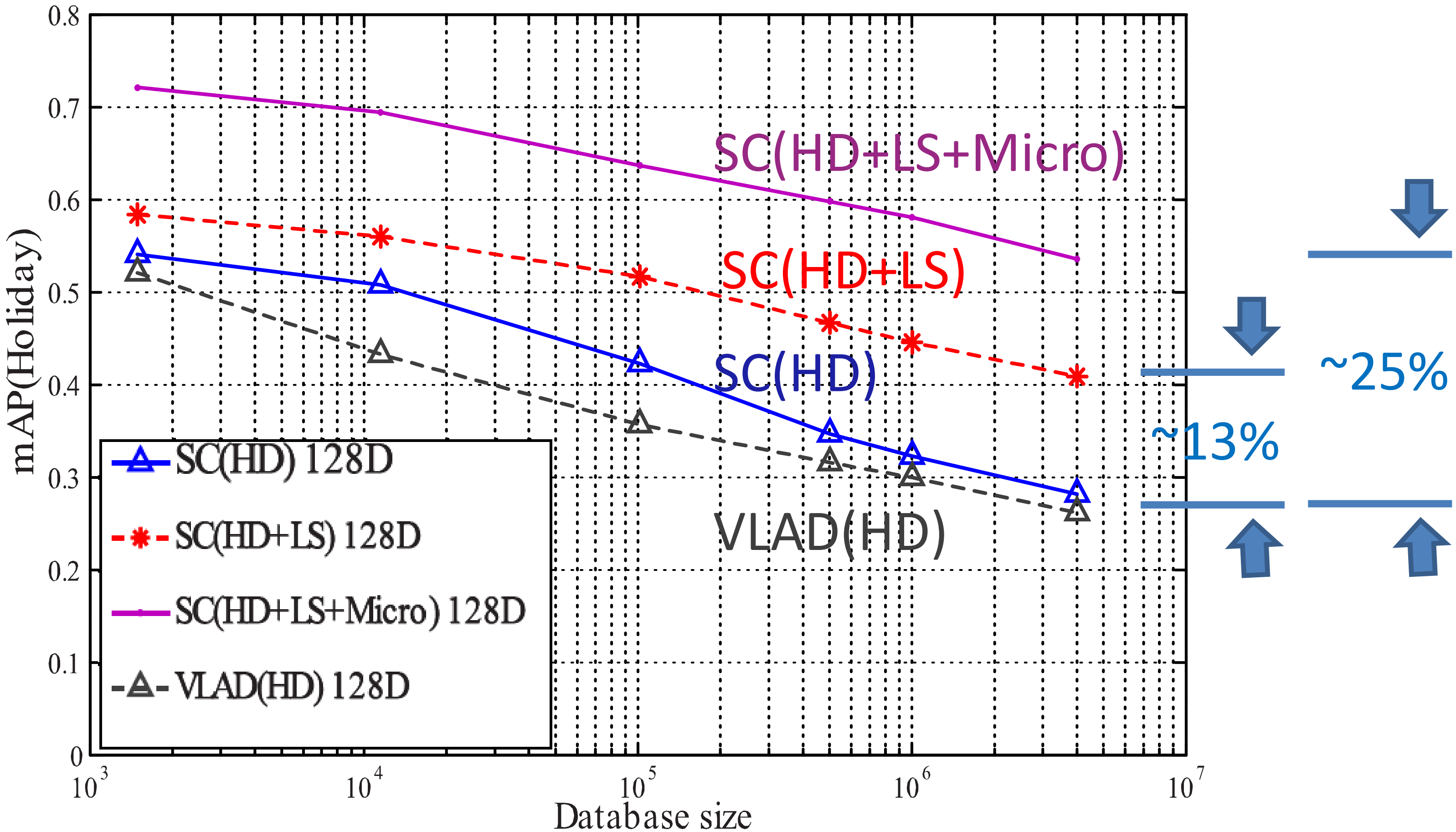
Scalability study

- Holidays + 4M images from *Flickr*.



Scalability study

- Holidays + 4M images from *Flickr*.



Conclusion

- Our work:

➢ Is based on local feature aggregation
➢ Applies sparse coding
➢ Utilize multiple features
➢ Designs novel “Micro feature”

Thank you!