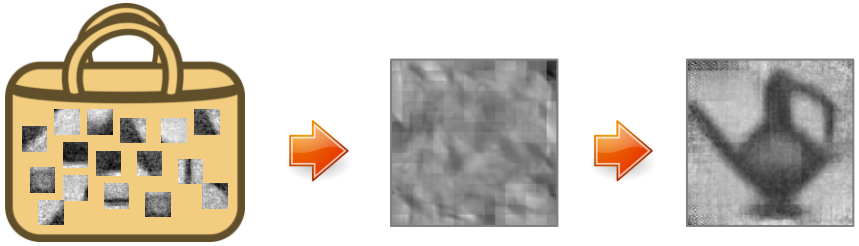
MIRU 2014 Invited Talk

[CVPR 2014]

Image Reconstruction from Bag-of-Visual-Words



Hiroharu Kato

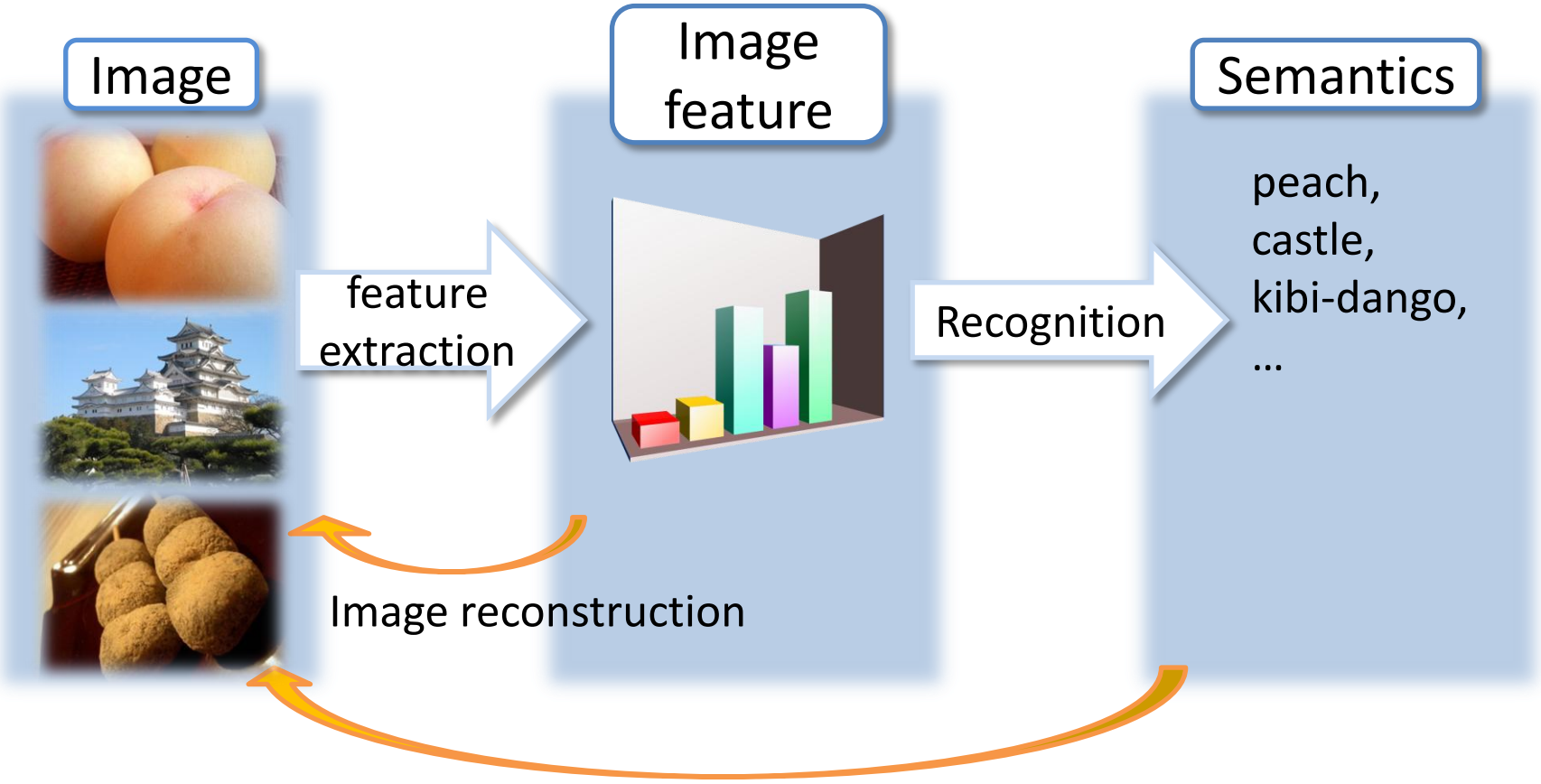
Univ. of Tokyo (-2014/03)

Tatsuya Harada

Univ. of Tokyo

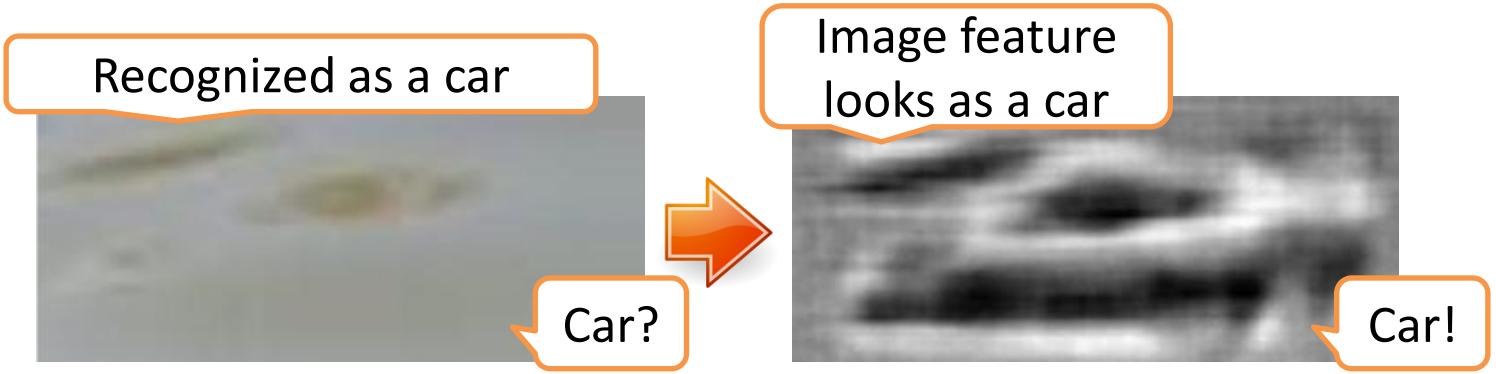
Introduction: Image Features

- Image features
  - Connection between images and semantics



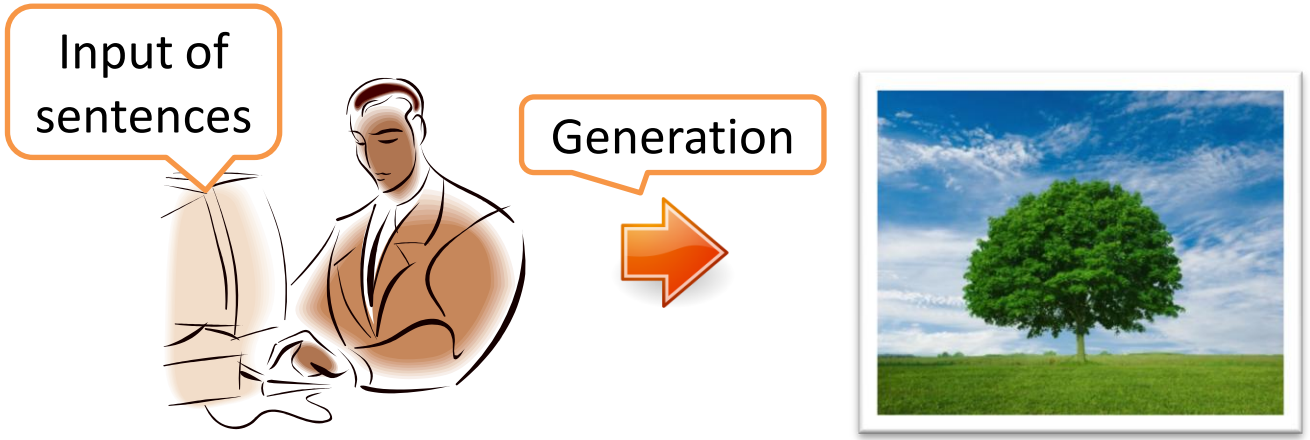
Introduction : Image Reconstruction

- Image reconstruction from features
  - Enables intuitive understanding of image features



[Vondrick *et al.*, 2013]

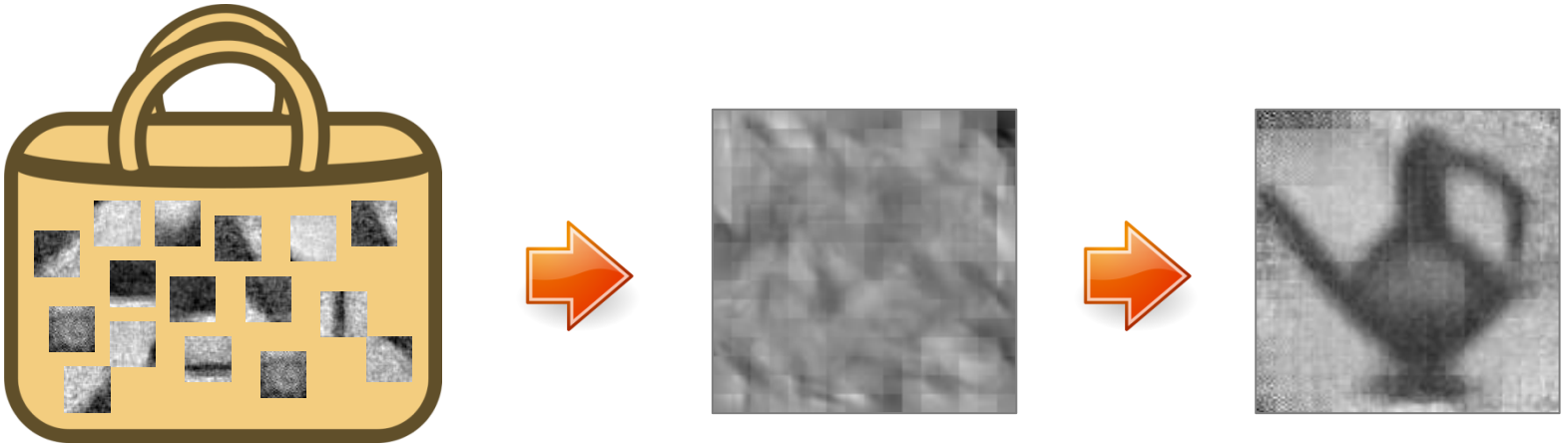
- Image generation from natural sentences



Introduction : Objective

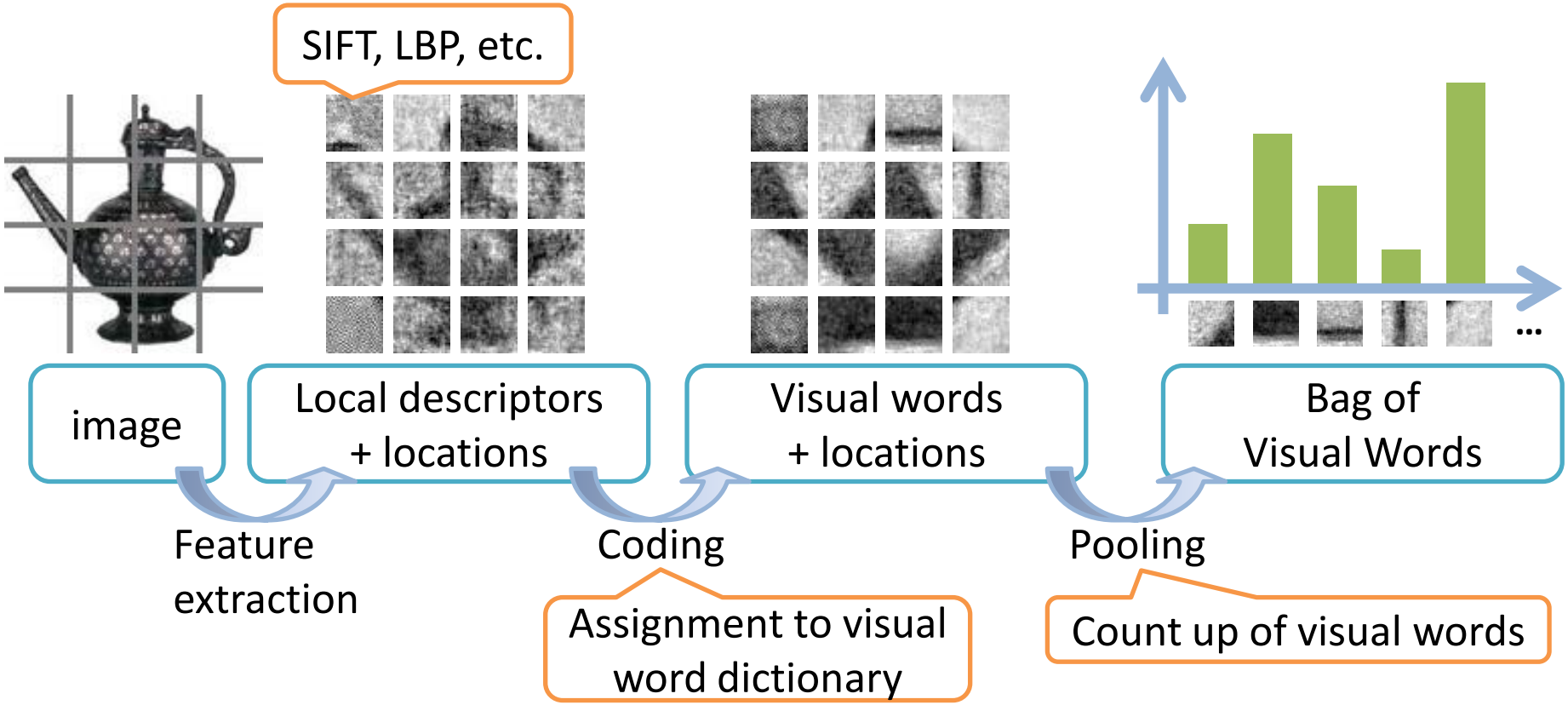
- Bag-of-Visual-Words (**BoVW**)
  - De facto standard for recognition and retrieval
  - extended to many modern features
  - Not reconstructed yet

- Objective of this work
  - To reconstruct the original image from BoVW



Bag-of-Visual-Words

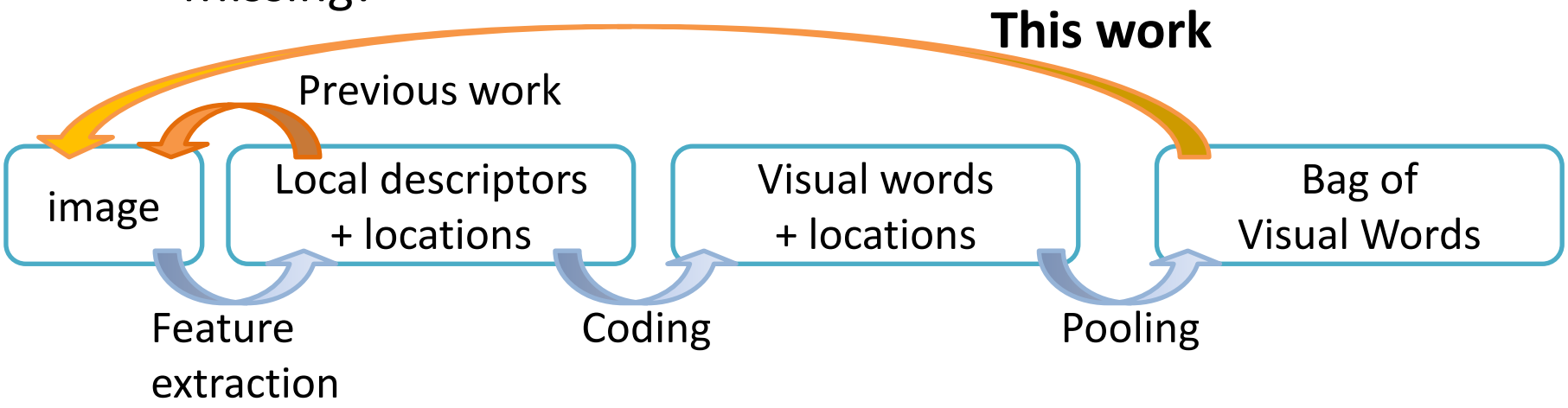
- De facto standard feature
  - For retrieval [Sivic *et al.*, 2003], for recognition [Csurka *et al.*, 2004]



Related Work: Image Reconstruction

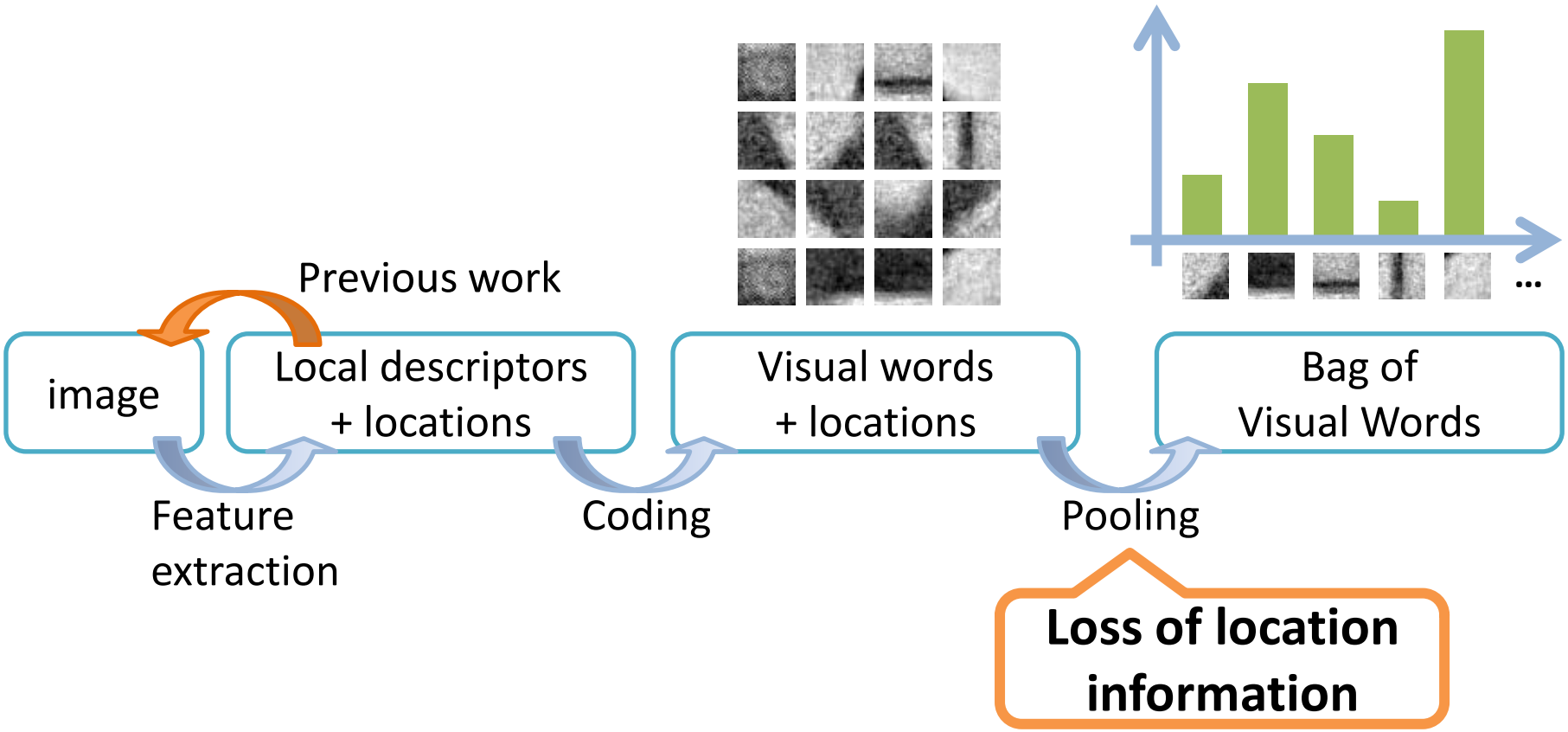
- Image patch generation from local descriptors
  - SIFT [Weinzaepfel *et al.*, 2011]
  - BRIEF/FREAK [d'Angelo *et al.*, 2012]
  - HOG [Vondrick *et al.*, 2013]

- Image reconstruction from set of descriptors
  - Missing!



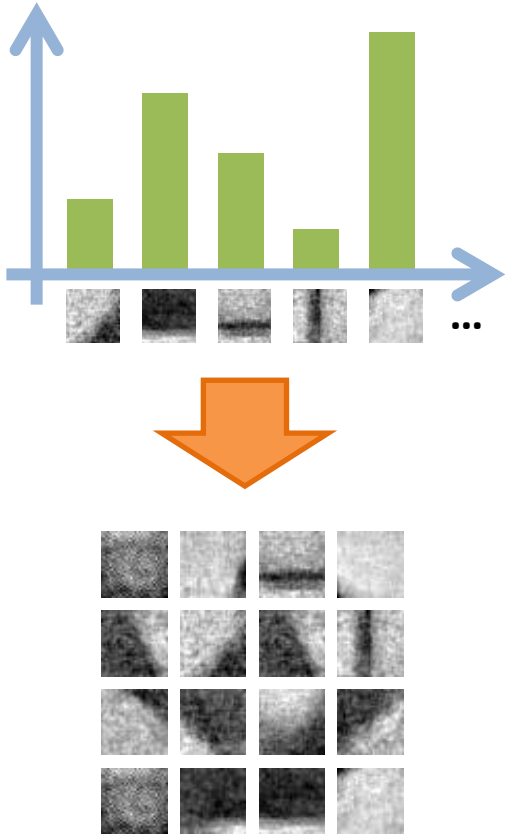
Method: Main Problem

- Additional information loss
  - Location information of visual words



Method: Estimation of Locations

- Re-arrangement of visual words in BoVW
  - To assign *n* visual words in *n* grid points



- Two possible strategies
  - Naturalness of adjacencies

ex. 

  - Naturalness of absolute locations

ex.  does not lie in the center

Method: Adjacency Cost

- Naturalness of adjacencies of visual words
  - Learnt from an image database

    ex. frequencies in the database are expected as

  - C^a_{ijkl} : a cost to assign visual word *i, j* to location *k, l*

$$H^a_{ijkl} = \text{Frequency of co-occurrences of visual words } i, j \text{ at the relative location } k, l$$

$$C^a_{ijkl} = -\log(H^a_{ijkl} + 1)$$

Method: Global Location Cost

- Naturalness of absolute locations of visual words
  - Learnt from similar images

    ex. in the similar images

     tends to lie in the edge of the image

  - C_{ik}^{p} : a cost to assign visual word *i* at location *k*

$$H_{ik}^{p} = \text{Frequency of visual word } i \text{ at location } k$$

$$C_{ik}^{p} = -\log(H_{ik}^{p} + 1)$$

Method: Optimization

- Estimation of the best arrangement
  - $x_{ik} = 1$: if visual word *i* lies location *k*

    $x_{ik} = 0$: otherwise

$$\min \quad (1-\lambda) \sum_{i.j.k.l=1}^{n} C_{ijkl}^{a} x_{ik} x_{jl} + \lambda \sum_{i,k=1}^{n} C_{ik}^{p} x_{ik}$$

Naturalness of global location

Naturalness of adjacency

$$s.t. \quad \sum_{i=1}^{N} x_{ik} = 1 \quad (1 \le k \le N)$$

One visual word at one location

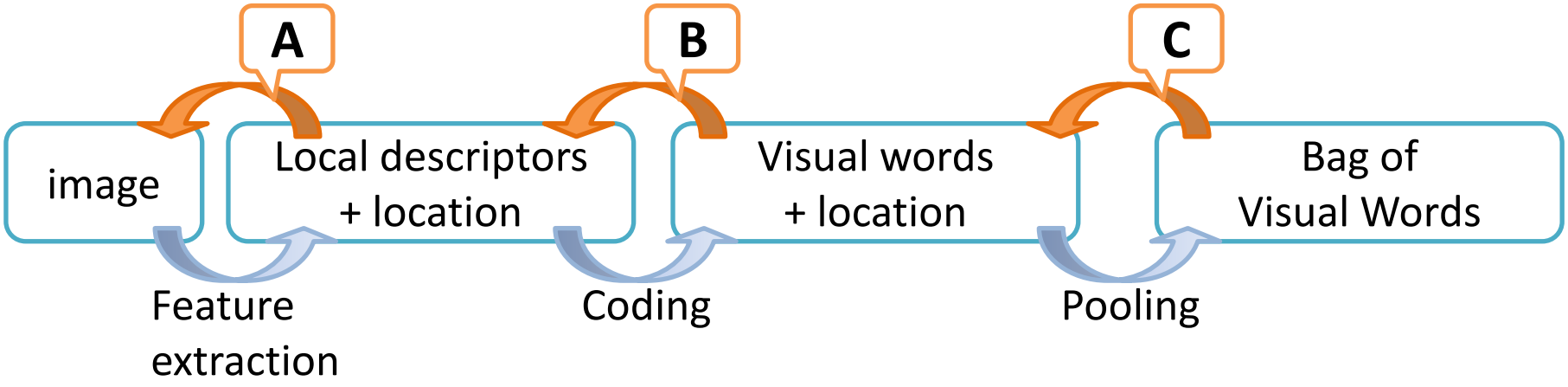
$$\sum_{k=1}^{N} x_{ik} = 1 \quad (1 \le i \le N)$$

$$x_{ik} \in \{0,1\}$$

- This result in the Quadratic Assignment Problem
  - Solved by Genetic Algorithm + Hill Climbing

Method: Summary

A) Image reconstruction from “descriptors + locations”
   – By previous work [Vondrick *et al.*, 2013]
B) To assume “local descriptors = visual words”
C) To estimate spatial layout of visual words
   – Maximizing naturalness of adjacency and global location



Experiment: Settings

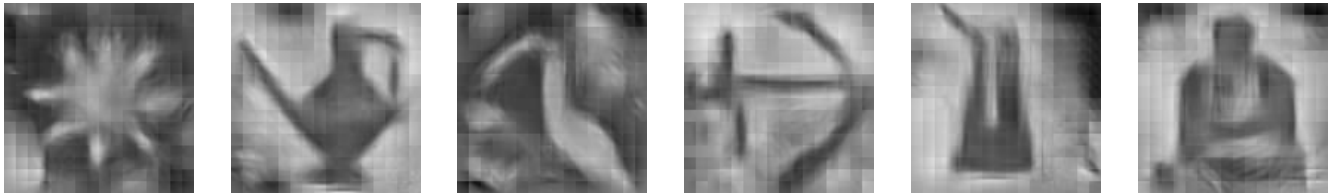
- Comparison of three methods
  - Our method, HOGgles [Vondrick *et al.*, 2013], image retrieval
- Images
  - Reconstruction: 101 objet images (from Caltech 101)
  - Image database: 1M object images (from ILSVRC 2012)
- Other settings
  - Local descriptor is SIFT, the number of descriptors is 13*13=169, the size of visual word dictionary is 8192, weight parameter λ is 0.8

Example of images used in experiments

Experiment: Comparison of Methods

Original image

Our method



HOGgles
[Vondrick *et al*., 2013]

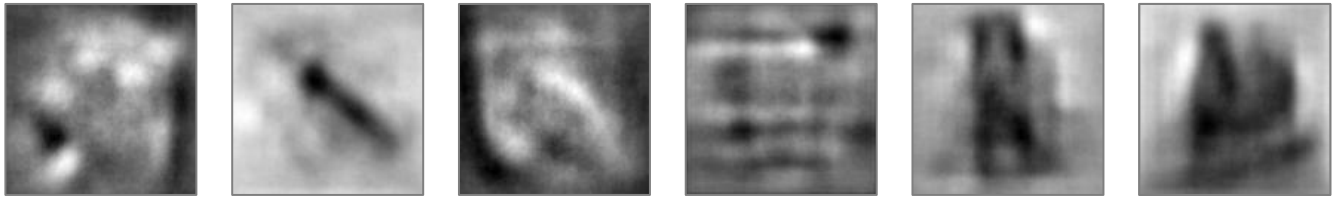
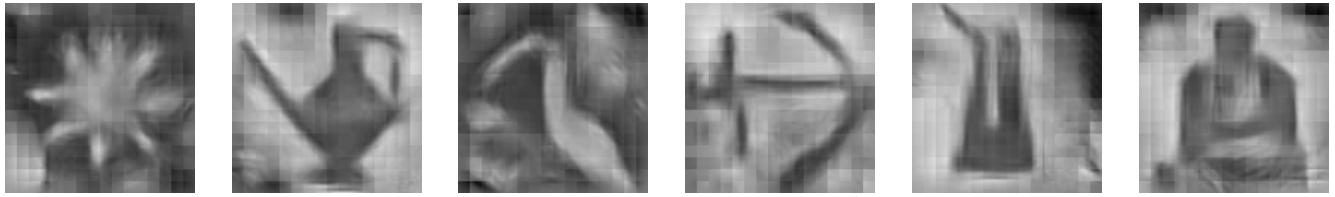


Image retrieval

Experiment: Comparison of Methods

Original image

Our method



HOGgles
[Vondrick *et al*., 2013]

Closely similar to the original images

Able to be recognized

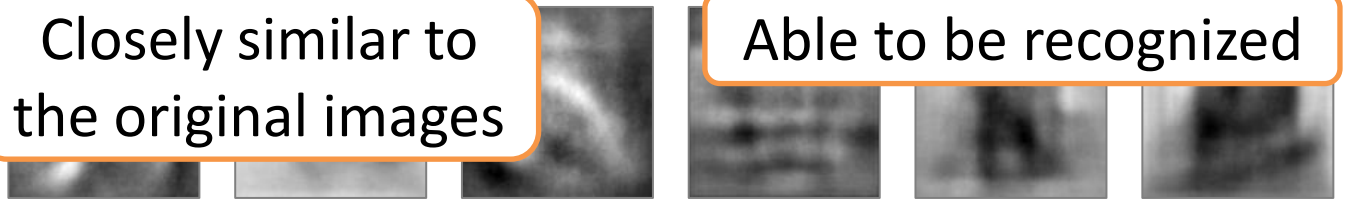


Image retrieval

Experiment: Comparison of Methods

Original image






Our method

general method for arbitrary features




HOGgles
[Vondrick *et al*., 2013]

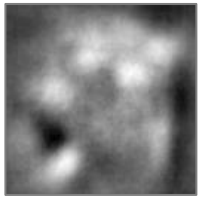
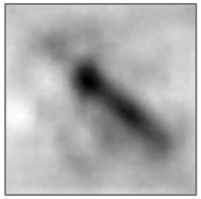
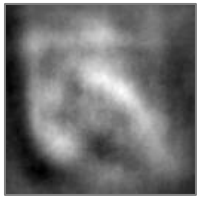





Image retrieval

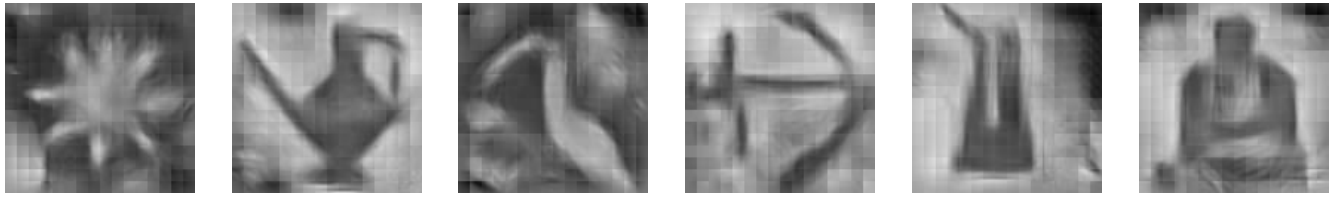


Heavily blurred and difficult to understand

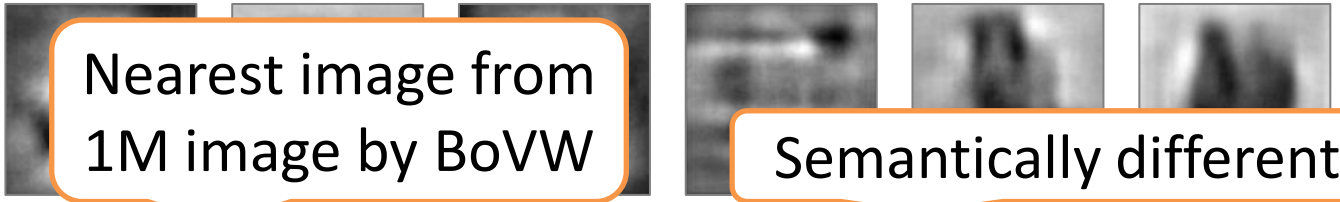
Experiment: Comparison of Methods

Original image

Our method



HOGgles
[Vondrick *et al*., 2013]

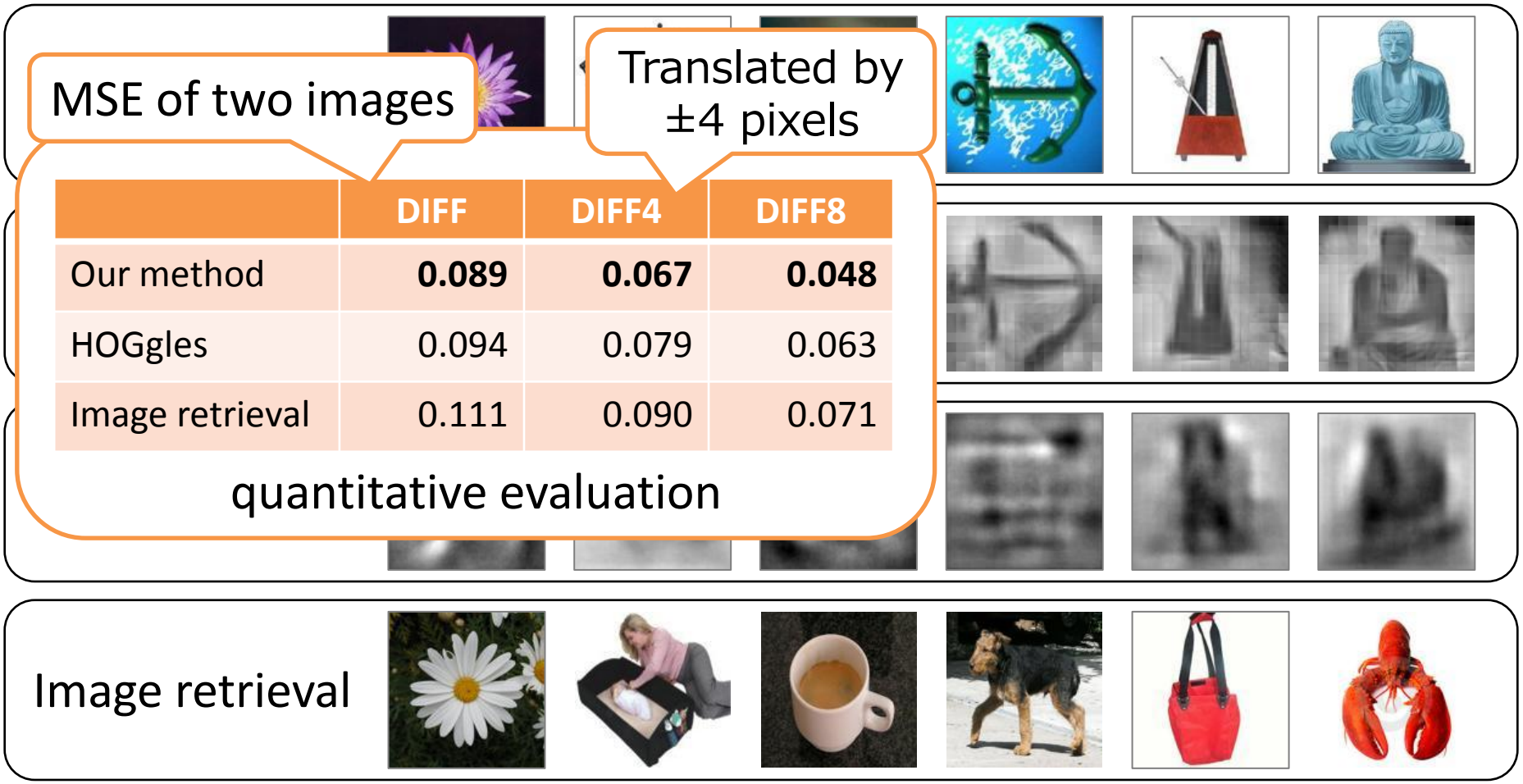


Nearest image from 1M image by BoVW

Semantically different

Image retrieval

Experiment: Comparison of Methods



MSE of two images

Translated by ±4 pixels

	DIFF	DIFF4	DIFF8
Our method	**0.089**	**0.067**	**0.048**
HOGgles	0.094	0.079	0.063
Image retrieval	0.111	0.090	0.071

quantitative evaluation

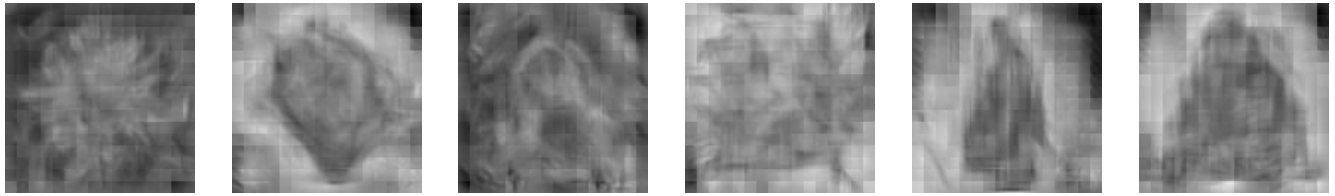
Image retrieval

Experiment: Weight Parameter

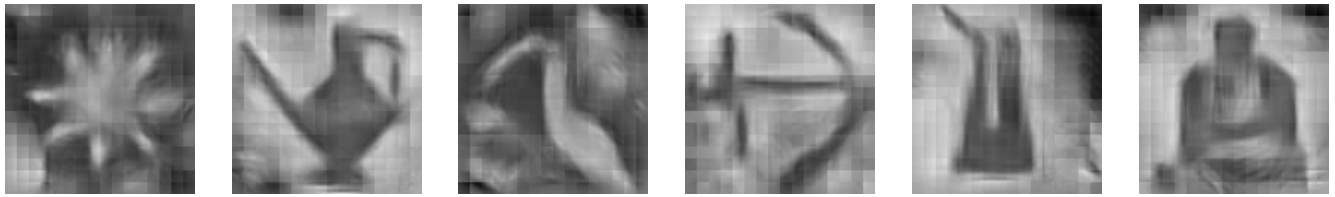
Cost function = (1-λ)*(Naturalness of absolute locations)
+ (λ)*(Naturalness of adjacencies)

Original image

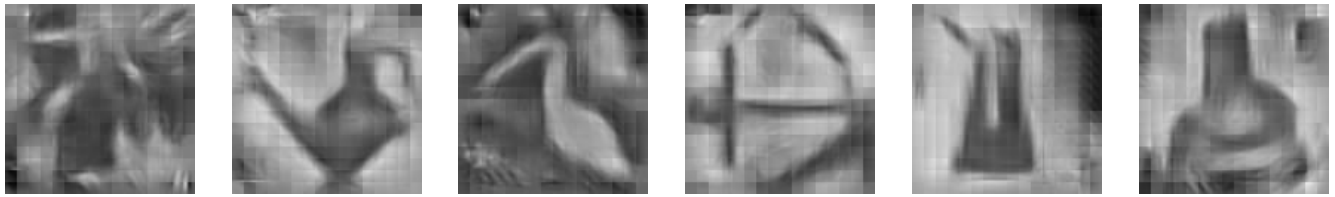
λ = 0.0



λ = 0.7



λ = 1.0

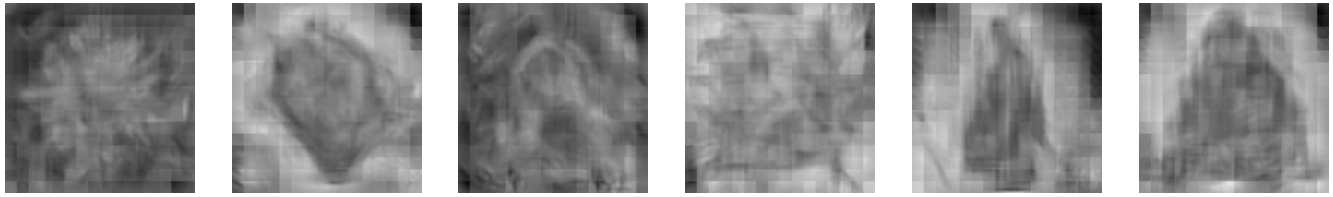


Experiment: Weight Parameter

Cost function = (1-λ)*(Naturalness of absolute locations)
+ (λ)*(Naturalness of adjacencies)

Original image

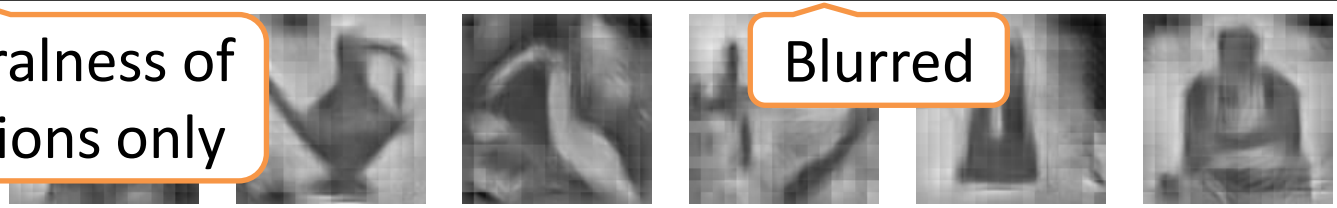
λ = 0.0



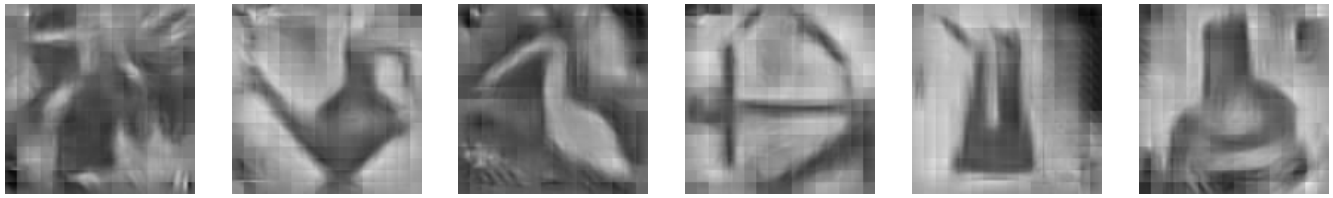
λ = 0.7

Naturalness of locations only

Blurred



λ = 1.0

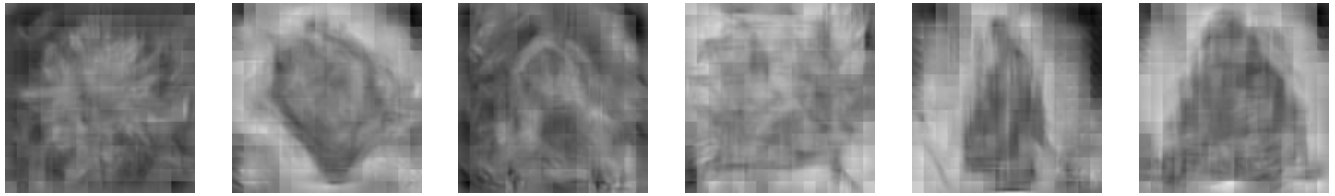


Experiment: Weight Parameter

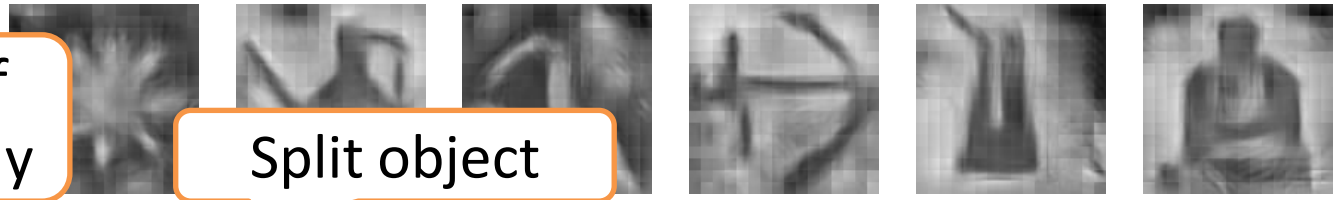
Cost function = (1-λ)*(Naturalness of absolute locations)
+ (λ)*(Naturalness of adjacencies)

Original image

λ = 0.0

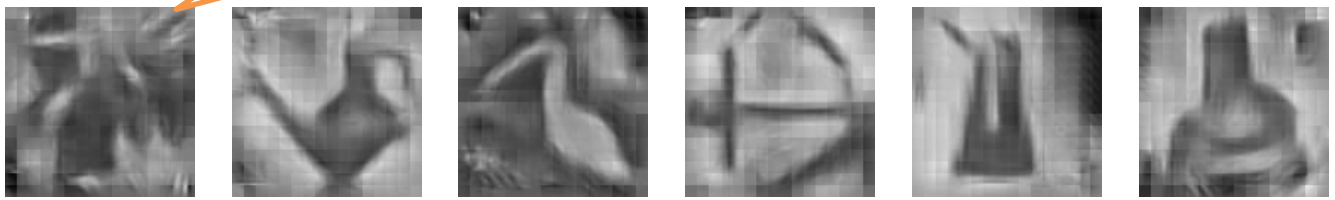


Naturalness of adjacencies only



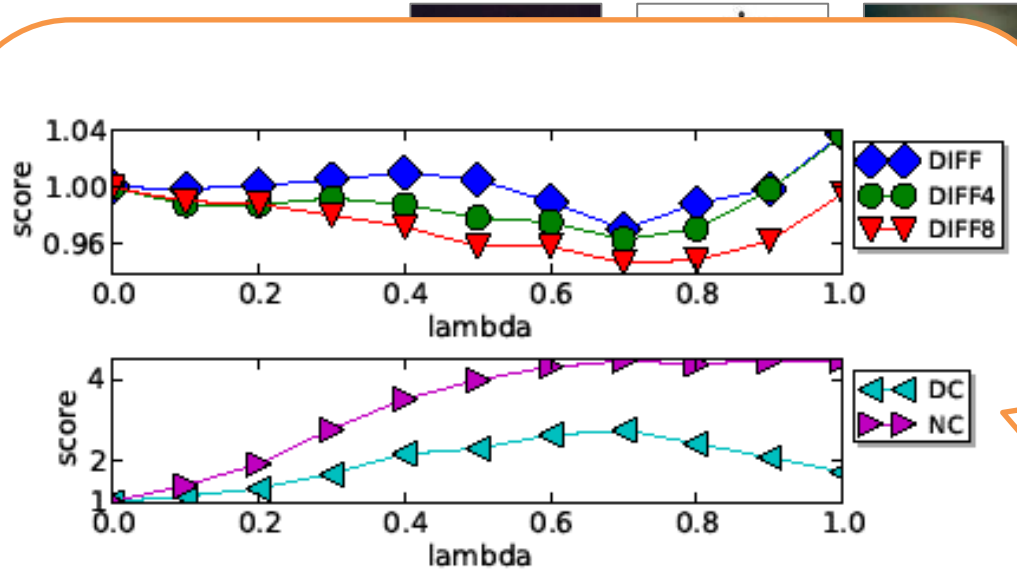
Split object

λ = 1.0



Experiment: Weight Parameter

Cost function = (1-λ)*(Naturalness of absolute locations)
+ (λ)*(Naturalness of adjacencies)



Score is maximized at λ=0.7

Difference from original images

Correctness of spatial layout

λ = 1.0

Conclusion and Feature Work

- Conclusion
  - Novel method for image reconstruction from BoVW
  - We demonstrated that
    - The spatial layout of visual words can be recovered
    - Modeling naturalness of 1) adjacency and 2) global position of them are both effective

- Future work
  - Extend for more sophisticated coding methods
  - Image generation via image features