[IT3-2][CVPR2014] Image Reconstruction from Bag-of-Visual-Words
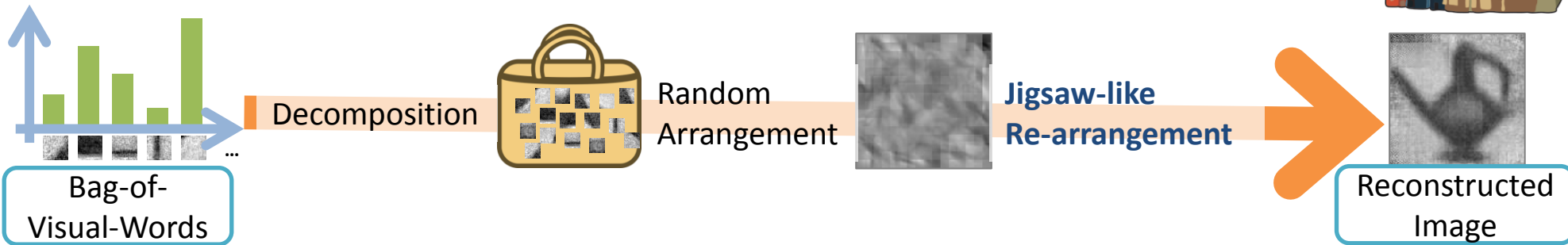
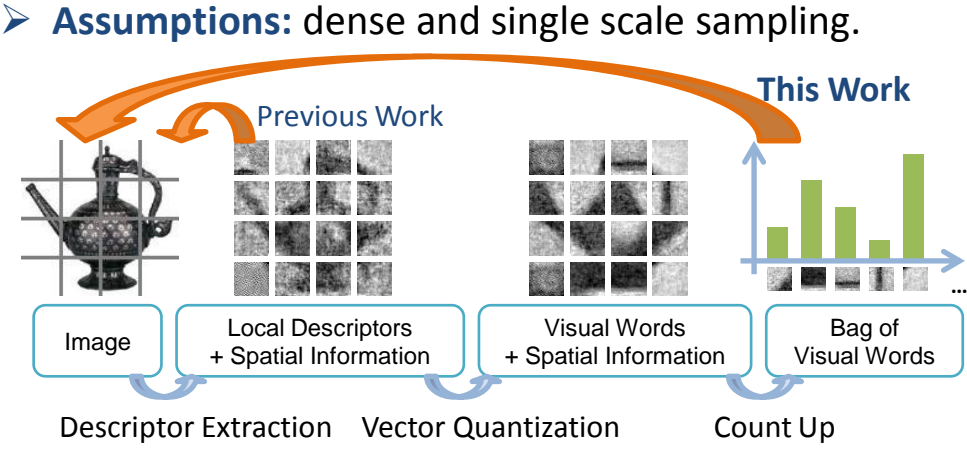Hiroharu Kato   Tatsuya Harada   - The University of Tokyo

Bag-of-Visual-Words → Decomposition → Random Arrangement → Jigsaw-like Re-arrangement → Reconstructed Image

Overview

➤ **Background**: To reconstruct an image from its image feature is useful for understanding the feature intuitively. It is the first work to reconstruct images from BoVW.

➤ **Main problem**: BoVW lacks spatial information of visual words.

➤ **Solution**: Spatial arrangement of them is estimated like solving a jigsaw puzzle, using statistics of local co-occurrences and absolute positions of visual words in an image database.
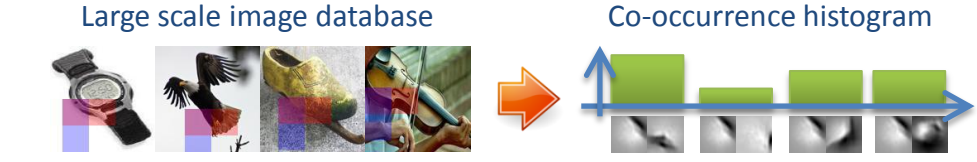
Reconstruction Method

Our proposed method consists of two steps.
1. Estimation of spatial arrangement of visual words.
2. Generation of an image patch from each visual word.
For the latter, we use HOGgles[*1]. For the former, we maximize the naturalness of adjacencies and global locations of them. We use an external image database to learn the naturalness.

Extraction of Bag-of-Visual-Words

➤ **Assumptions**: dense and single scale sampling.

Previous Work | This Work

Image → Local Descriptors + Spatial Information → Visual Words + Spatial Information → Bag of Visual Words

Descriptor Extraction   Vector Quantization   Count Up

➤ **Objective function**

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

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ik} = 1, \quad \sum_{k=1}^{n} x_{ik} = 1, \quad x_{i,k} \in \{0,1\}$$

➤ **Adjacency Cost C^a**

Large scale image database → Co-occurrence histogram

➤ **Global Location Cost C^l**

Similar images by BoVW → Spatial histogram of

Experimental Results

Good Results

		Original Images					Obtained from BoVW

Original Images							
Obtained from BoVW	**Our method**						
	HOGgles[*1]						
	Image Retrieval						

We used 101 object images. Five results of them are shown here.

[Settings]
- Image size: 128x128 px
- Descriptor size: 32x32 px
- Local descriptor: SIFT
- Vocabulary size of BoVW: 5000
- Descriptor extraction step: 8px

*1 Vondrick et al., ICCV, 2013.

Discussion

➤ **Reconstructability**: images of single objects are reconstructed more stably than that of complicated textures.

➤ **Computational Cost**: About a minute for one image. The bottleneck is optimization.

➤ **Limitations**: there are several assumptions which are desirable to be relaxed, such as single scale sampling and hard assignment of local descriptors.

THE UNIVERSITY OF TOKYO

All experimental results are available at
http://www.mi.t.u-tokyo.ac.jp/kato/cvpr2014.html