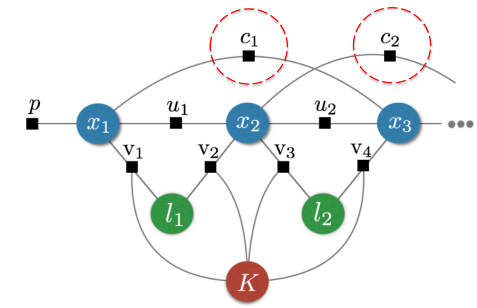
Recall SLAM factor graph



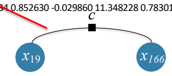
E.g., in “input_INTEL_g2o.g2o”

Odometric constraints: u_i

EDGE_SE2 1224 1225 0.625700 0.003903 0.013764 11.124947 1.838454 0.000000 255.403355 0.000000 2432.575223
EDGE_SE2 1225 1226 0.106504 -0.000109 -0.016731 13.283390 -138.281257 0.000000 8813.714781 0.000000 2418.398573
EDGE_SE2 1226 1227 0.000000 0.000000 -0.000662 11.111282 -0.257444 0.000000 399.999830 0.000000 2496.693284

Loop closure constraints: c_i

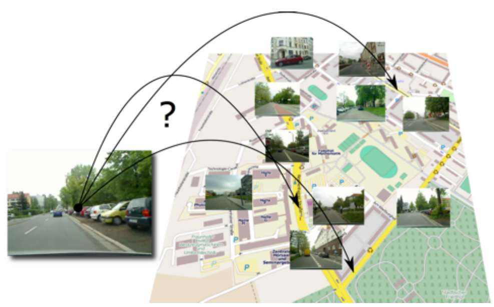
EDGE_SE2 19 166 -2.459689 0.241111 0.252800 11.731352 1.696454 0.000000 15.751178 0.000000 1592.856013
EDGE_SE2 19 172 1.067903 0.915786 0.149470 22.211594 -17.729393 0.000000 39.428022 0.000000 1892.102792
EDGE_SE2 25 172 -2.539634 0.852630 -0.029860 11.348228 0.783018 0.000000 13.696833 0.000000 2357.130503



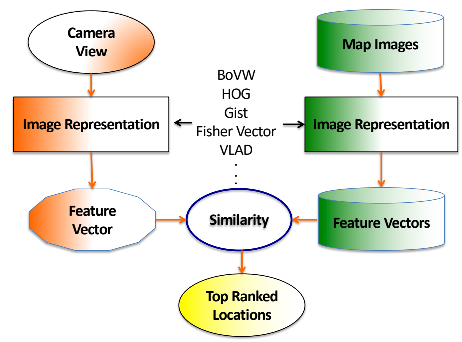
What are h_k, $\mathcal{X}_k$ and z_k?

$$\min_{\boldsymbol{R}_i, \boldsymbol{t}_i} \sum_{(i,j)\in\mathcal{E}} \left\|\mathrm{Log}\left(\bar{\boldsymbol{R}}_{ij}^{\mathsf{T}} \boldsymbol{R}_i^{\mathsf{T}} \boldsymbol{R}_j\right)\right\|_{\Omega_{ij}^r}^2 + \left\|\bar{\boldsymbol{t}}_{ij} - \boldsymbol{R}_i^{\mathsf{T}}(\boldsymbol{t}_j - \boldsymbol{t}_i)\right\|_{\Omega_{ij}^t}^2$$

Visual loop closure detection (a.k.a. visual place recognition)



General LCD/VPR framework



Visual LCD/VPR

Typical LCD processing pipeline:

1. Selection of top ranked map candidates
2. Verification of the candidates (loop closure verification)

Major challenges:

1. Scalability: LCD in a map of numerous poses (images)
2. Illumination Invariance: LCD in spite of significant illumination changes between visits
3. Viewpoint invariance: LCD in spite of significant viewpoint changes between visits

Do these images come from the same place?

Mapillary HK

Scalability: BoW-based solution

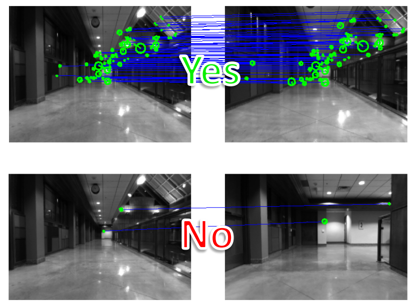
11. Sivic, J. and Zisserman, A. "Video Google: A Text Retrieval Approach to Object Matching in Videos", Proceedings of the International Conference on Computer Vision (ICCV), 2003.

Galvez-Lopez, Dorian, and Juan D. Tardos. "Real-time loop detection with bags of binary words." *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011.

7

Image Matching

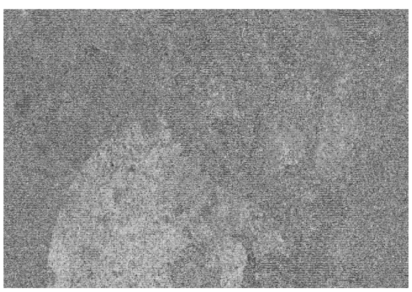
Current View A Map Location



8

Image Matching?

53,000 Map Locations Current View



= ?

9

Brute-Force Feature Matching

$$O(n^2dN)$$

n: number features in an image
d: dimension of a feature descriptor
N: number of map locations

E.g., $300^2 \times 128 \times 10^5 \approx 10^{11}$

10

Bag-of-Words (BoW)

$$O(ndN/k) = O(ndN')$$

Cost of NN search for vector quantization

n: number features in an image
d: dimension of a feature descriptor
N: number of map locations
k: a constant scaler

E.g., $300 \times 128 \times 10^5/10^2 \approx 10^7$

11

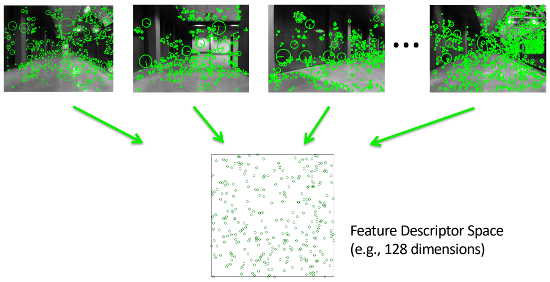
Bag of Words?

- How does Google search the internet so quickly?
- Answer: indexing
- However, we are dealing with visual data and must express images as "text files" before indexing can be applied.

12

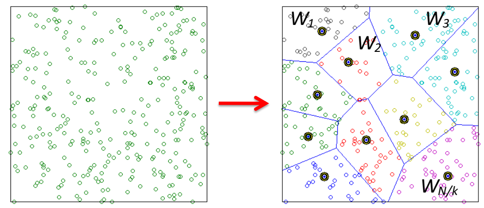
BoW Image Descriptor

1. Extract visual features from training images



BoW Image Descriptor

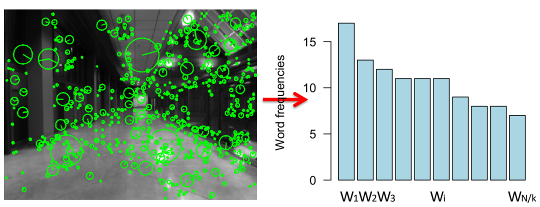
2. Cluster the extracted features



Training features

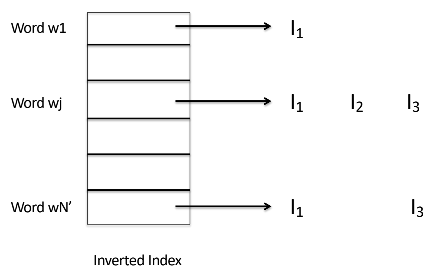
or dictionary: $\{W_i\}$

BoW Image Descriptor



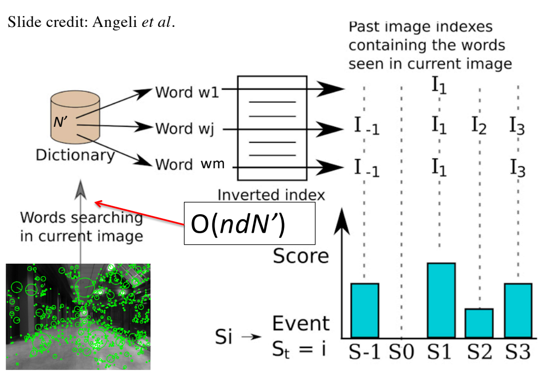
Each image is described in terms of a histogram of visual word frequencies

BoW Inverted Index

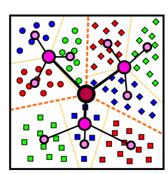


N' = N/k

Slide credit: Angeli *et al*.



Bag-of-Words w/ Vocabulary Tree

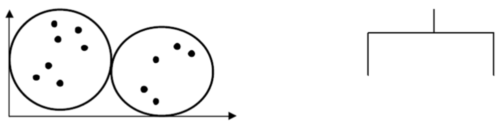


$O(nd\log[N'])$

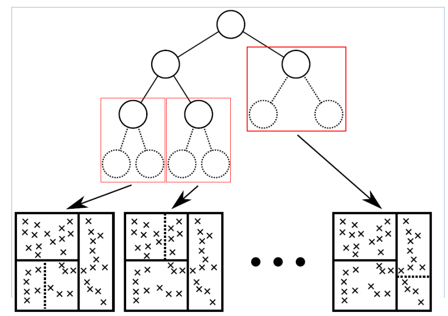
Cost of ANN search for vector quantization

n: number features in an image
d: dimension of a feature descriptor
N': *N/k*
N: number of map locations
k: a constant scaler

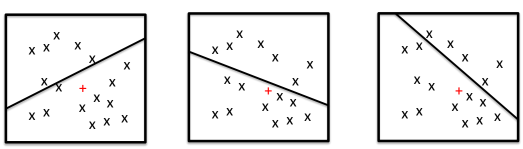
Bag-of-Words w/ Vocabulary Tree



Bag-of-Words w/ K-D Tree

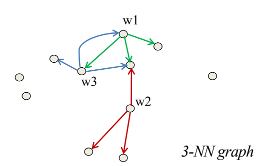


Bag-of-Words w/ LSH



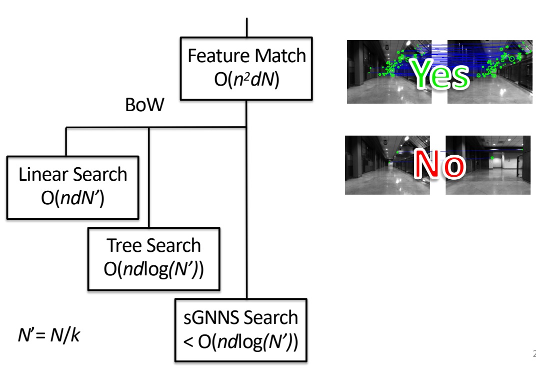
- Choose random directions
- Project visual words to these directions
- Pick a random threshold to bipartite the words

Bag-of-Words w/ GNNS



1. Construct a k-nearest neighbor graph (offline)
2. Randomly initialize the search (online)
3. Local hill-climbing until local minimum or a max step count is reached

Location/Place Recognition

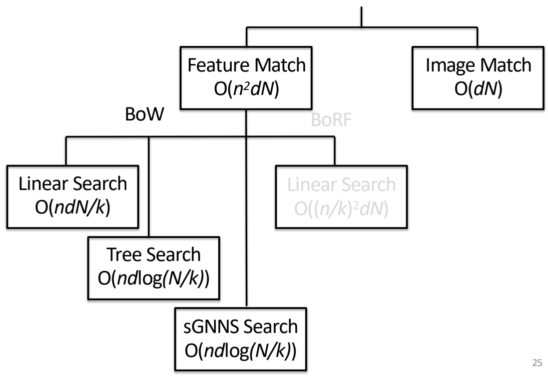


Whole Image Descriptor

$$O(1*dN)=O(dN)$$

d: dimension of the descriptor
N: number of map locations

Location/Place Recognition

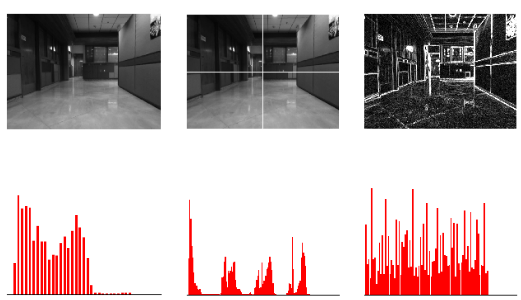


Whole Image Descriptor

O(dN)

Examples: Color Histogram, HOG, Gist, SIFT, SURF, Down Sampled (DS) Image, *Binarized DS Image*

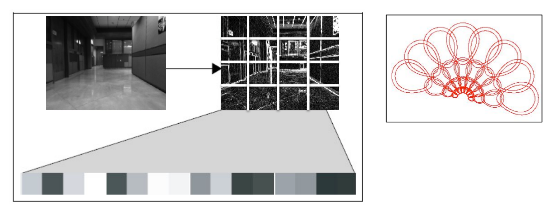
Whole Image Descriptor



Histogram Local histogram Gradient based

Whole Image Descriptor

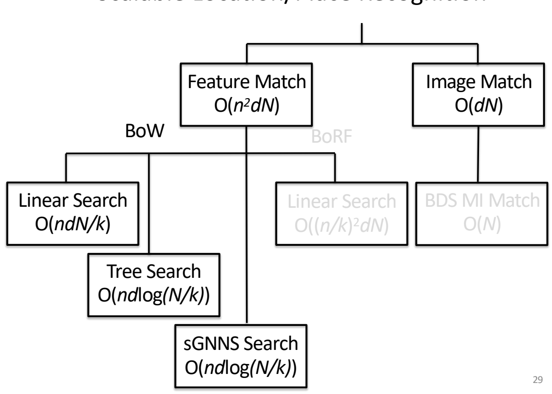
- Gabor-Gist



d = 16 x 20 = 320

Yang Liu and Hong Zhang, **Visual Loop Closure Detection with a Compact Image Descriptor**, 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Algarve, Portugal, October 2012

Scalable Location/Place Recognition



Summary

- Solutions to scalability
  - BoW based: sGNNS
  - Whole image based: BDS+MI
- Open issue: robustness to dynamic changes:
  - illumination
  - seasonable
  - moving objects
- Ongoing work
  - Semantics SLAM