

# Using $n$ -grams models for semantic place recognition



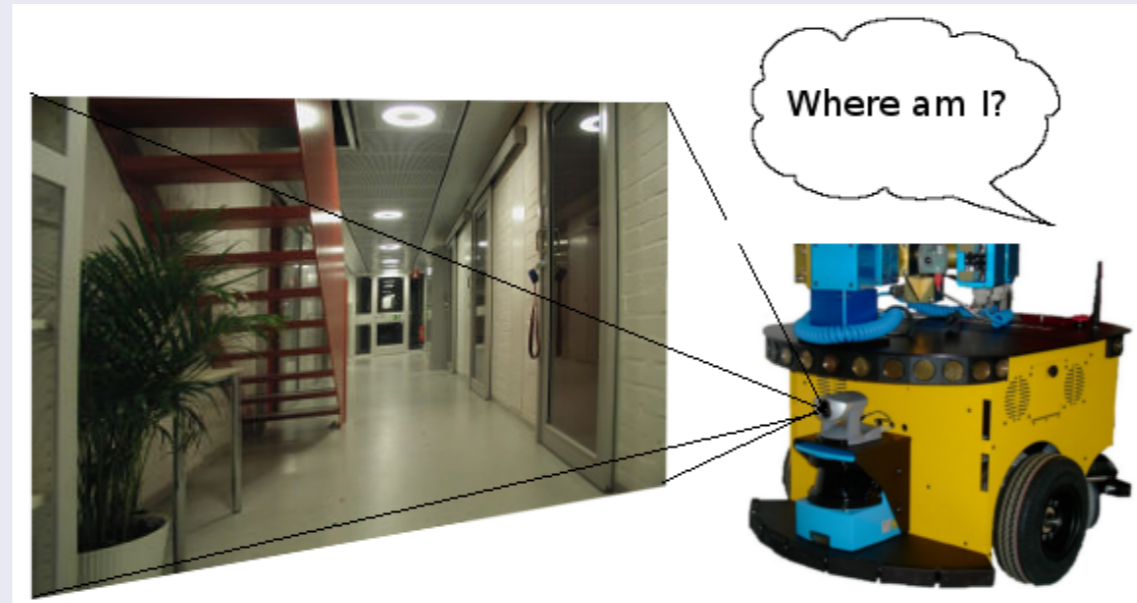
Mathieu Dubois, Emmanuelle Frenoux, Philippe Tarroux

LIMSI-CNRS & Université Paris-Sud  
first.last@limsi.fr

## Goal

Recognize current place (semantic context) from its appearance:

- without map
- new environment



Robotics constraints: incremental & on-line learning, real-time.

## Problem

A place has different appearances ( $\neq$  scene categorization):



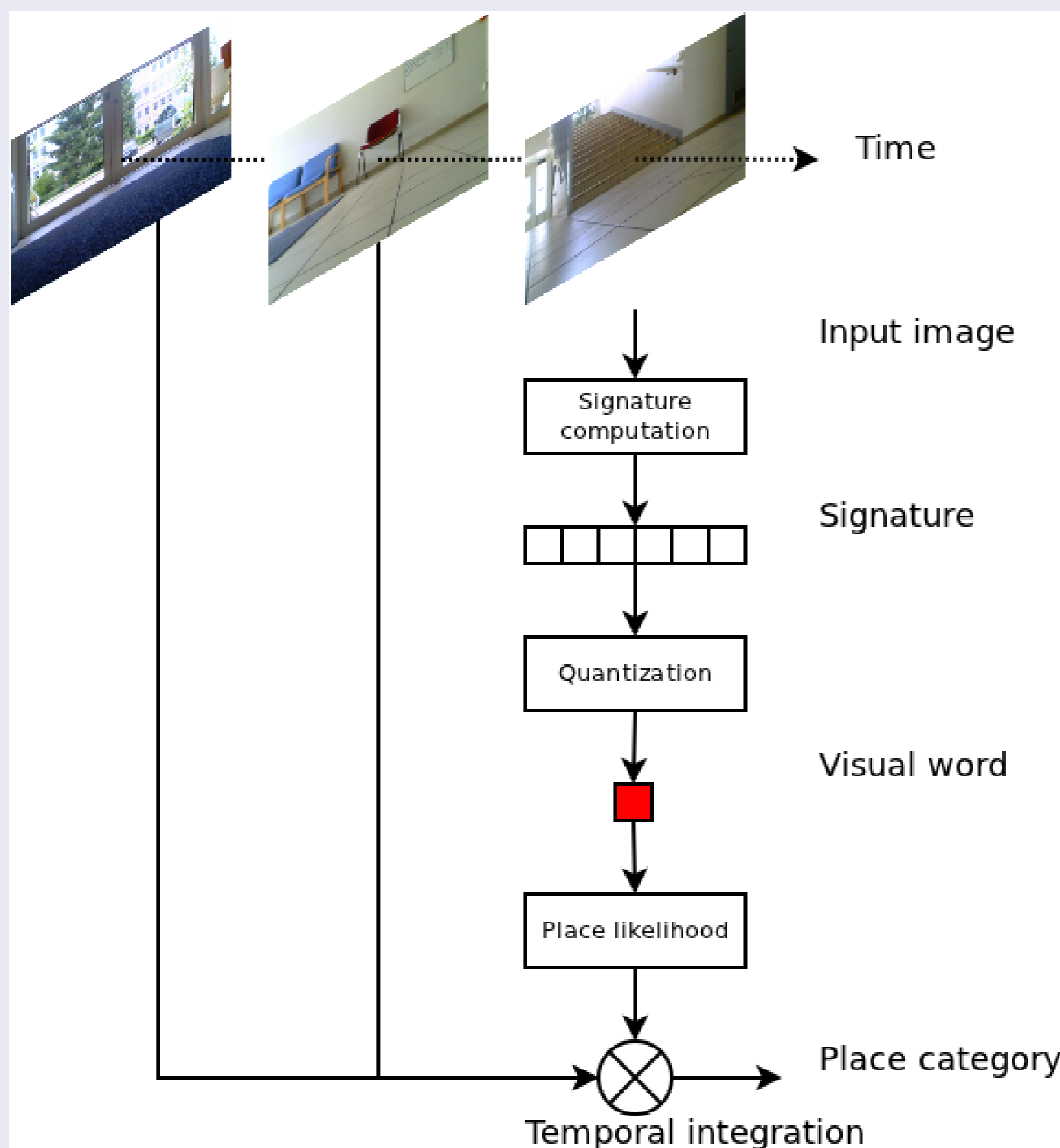
Outside      Conference room      Stairs  
**Different scenes but same place**

## A system for place recognition based on visual words

- Supervised learning
- Don't recognize objects (supported by psychological studies)
- Visual context**: use global appearance
- Temporal integration**: use several images

Analogy with visual words:

- Global image characterization: GIST ([1])
- Dictionary of global signatures  $\Rightarrow$  global visual words
- Analogy with Natural Language Processing (NLP):
- Image sequence  $\Rightarrow$  sentence of visual words**
- A place  $\Leftrightarrow$  a language**

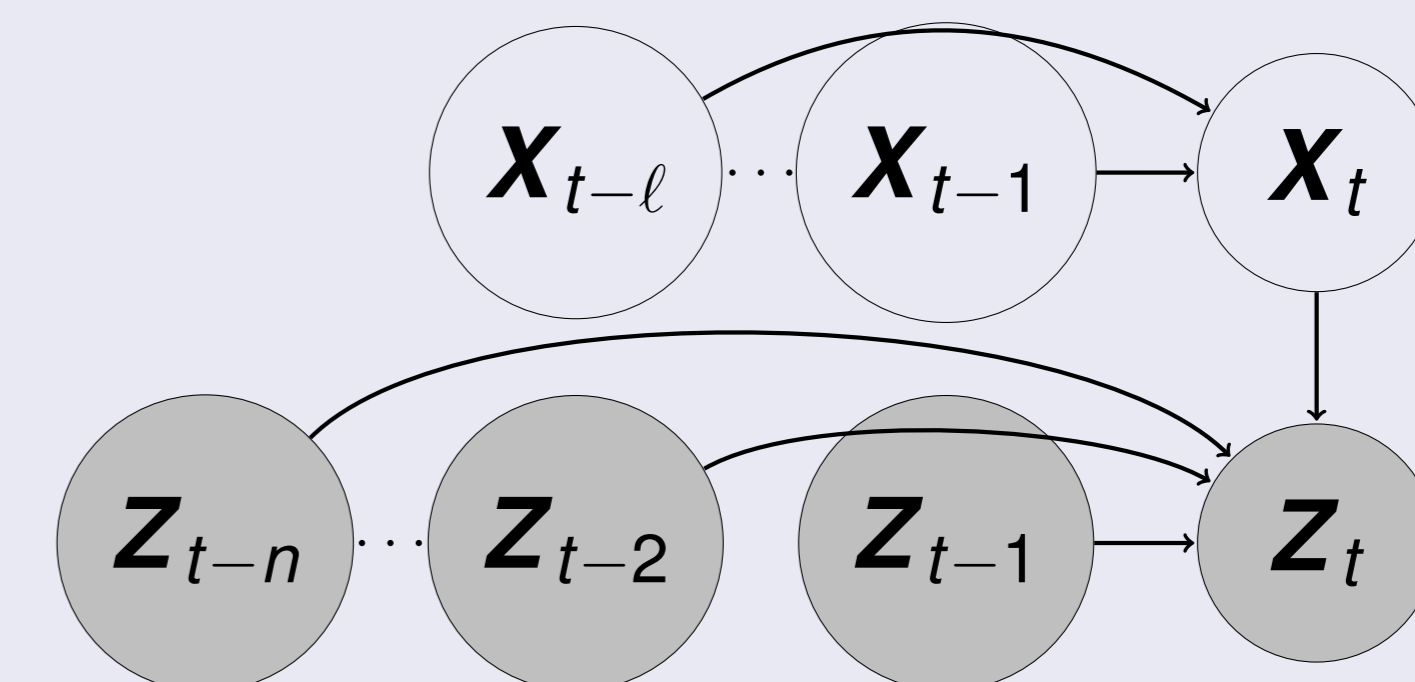


## Temporal integration

High-Order Double Chain Markov Model ([2]):

**High-order model**: State  $\mathbf{X}_t$  depends on  $\ell$  past states

**Incomplete hidden state** (doesn't provide all the information): observation  $\mathbf{Z}_t$  depends on  $\mathbf{X}_t$  and the last  $n$  observations



Previous work:

[3]: standard HMM ( $\ell = 1$  and  $n = 1$ )

[4]: use previous observation ( $\ell = 1$  and  $n = 2$ )

**Assuming  $\ell = 1$ :**

$$\text{bel}(\mathbf{x}_t) = \underbrace{P(\mathbf{z}_t | \mathbf{z}_{t-n:t-1}, \mathbf{x}_t)}_{\text{Place model}} \sum_{c_i \in \mathcal{C}} \underbrace{P(\mathbf{x}_t | \mathbf{x}_{t-1})}_{\text{Topology}} \text{bel}(\mathbf{x}_{t-1})$$

**Hypothesis: larger  $n$  improve performance.**

## Questions

- How to assign probability to unobserved sequences when learning the place model?
- Successive images are frequently mapped to the same word: how to simplify sentences?

## Method

1 In NLP  $P(\mathbf{z}_t | \mathbf{z}_{t-n:t-1}, \mathbf{x}_t = c_i)$  is called a  $n + 1$ -grams  $\Rightarrow$  efficient estimation methods:

- Laplace-Lidstone (LL)
  - Witten-Bell (WB)
- 2 Filter words (attention):
- Select 1 image over  $s$  (subsample)
  - Compress sequences of  $m$  identical words (compress)
  - Use only unique words (unique)

Experiments:

- Instance recognition on the COLD database
- Vary observation order ( $n$ ), estimation method, size of vocabulary ( $K$ ) and filtering method

## Conclusions & future works

Conclusions:

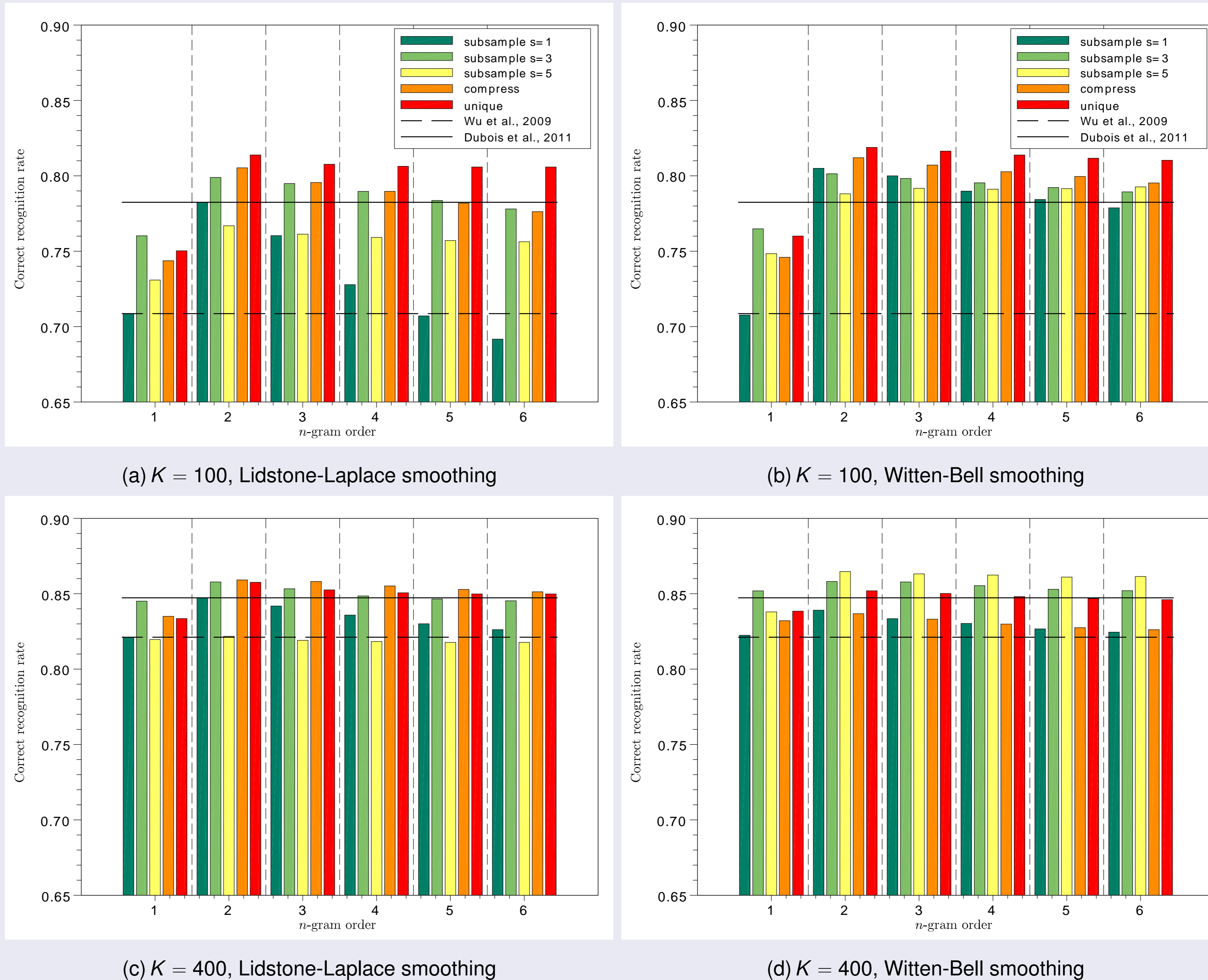
- using 2 words is better (coherent with [4]) but increasing  $n$  doesn't lead to best performance
- Witten-Bell smoothing increase performance
- the "unique" filtering is efficient for small vocabularies; the "subsample" filtering is efficient for larger vocabularies

## Technics from NLP help place recognition

Future works:

- Better visual words & descriptors
- Better word selection
- High-order model ( $\ell > 1$ )

## Results



**Figure**: Results on the instance recognition task. The vertical axis is the correct recognition rate (in %). The horizontal axis is the value of  $n$ . Upper-row: results for 100 words. Lower row: results for 400 words. Left column: results for Lidstone-Laplace smoothing. Right column: results for Witten-Bell smoothing. Dashed line: [3]. Plain line: [4]

## References

- A. Torralba et al. "Context-based vision system for place and object recognition". In: *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003)*, 14-17 October 2003, Nice, France. Vol. 1. IEEE Computer Society, Oct. 2003, pp. 273-280. ISBN: 0-7695-1950-4.
- A. Berchtold. *High-order extensions of the double chain Markov model*. Tech. rep. 356. University of Washington, 2002.
- J. Wu, H. Christensen, and J. Rehg. "Visual place categorization: Problem, dataset, and algorithm". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009 (IROS 2009)*. IEEE. St. Louis, USA, 2009, pp. 4763-4770.
- M. Dubois et al. "Visual place recognition using Bayesian filtering with Markov chains". In: *Proceedings of the European Symposium on Artificial Neural Networks (ESANN 2011)*. Ed. by M. Verleysen. 2011.