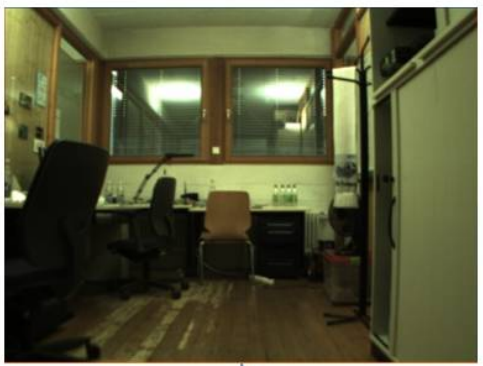
Robust Semantic Place Recognition with Vocabulary Tree and Landmark Detection

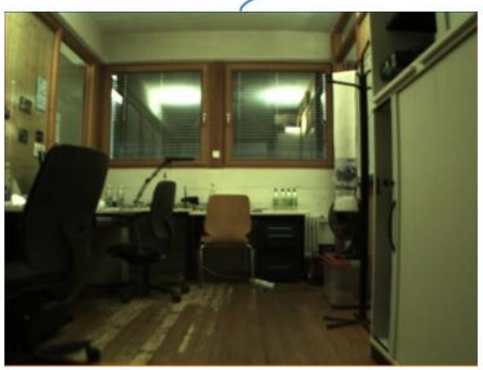
Lin Yuan, Kaichi Chan, C.S.George Lee

School of Electrical and Computer Engineering
Purdue University

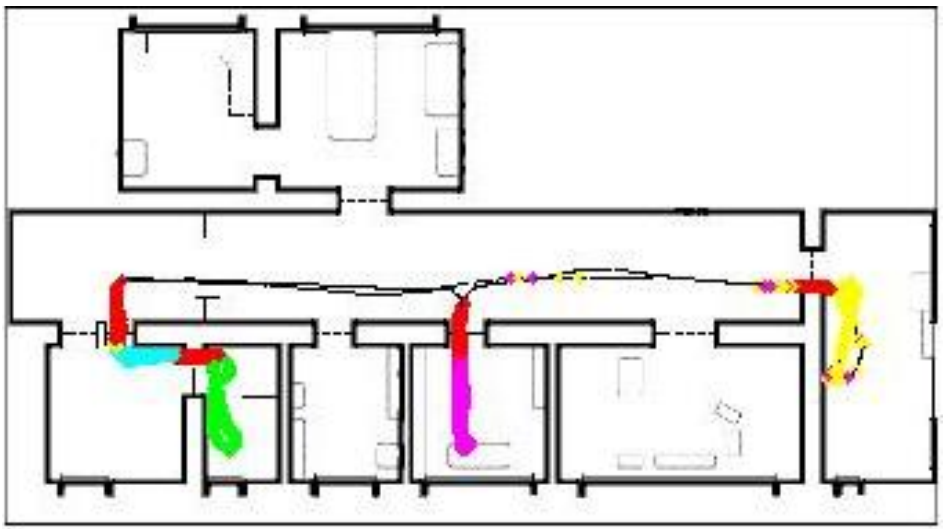
September 26, 2011

Place Recognition: Office A v.s. Office B

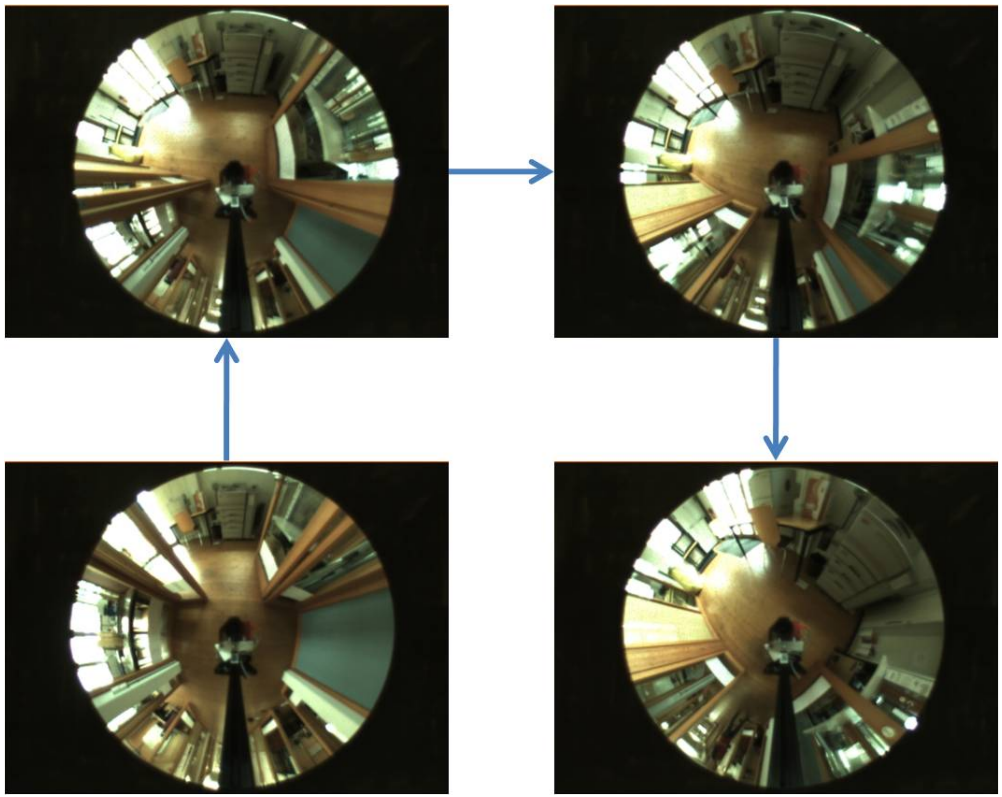




- Semantic Place Classification is the problem of classifying different type of environments.
- Semantic Place Recognition differs in the sense that it also needs to distinguish between “office A” and “office B”.

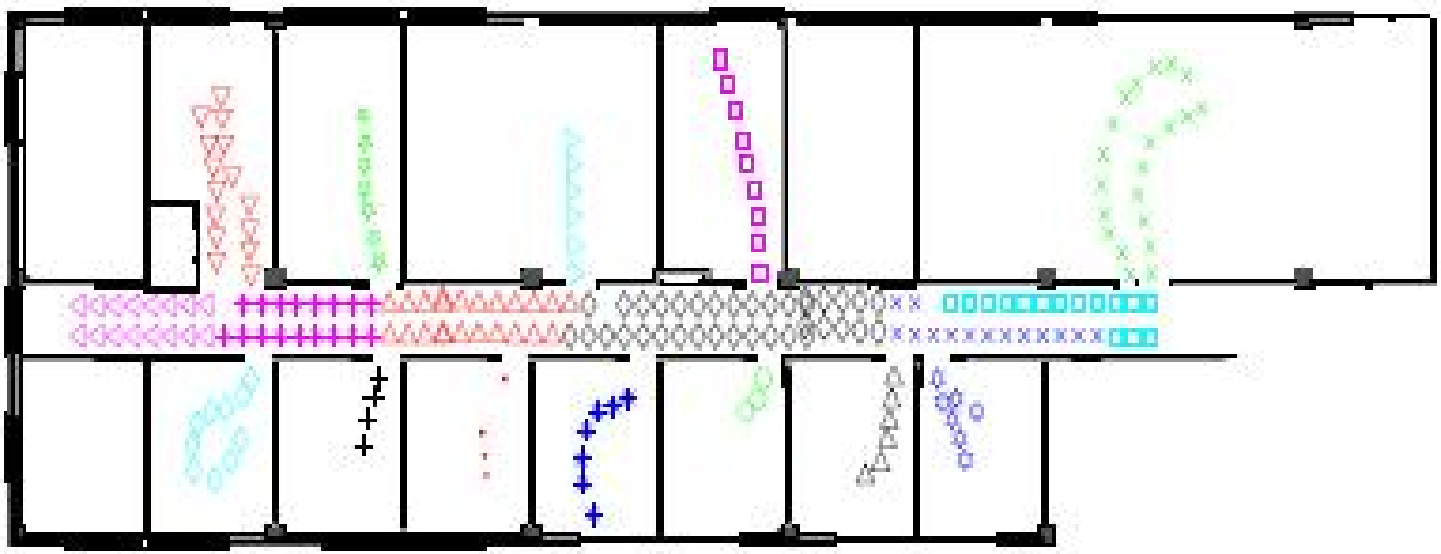


- Our method is concerned with the Semantic Place Recognition problem.
- But we look at it from a different point of view: instead of giving labels of different semantic places, how about we give labels for transitions between places?



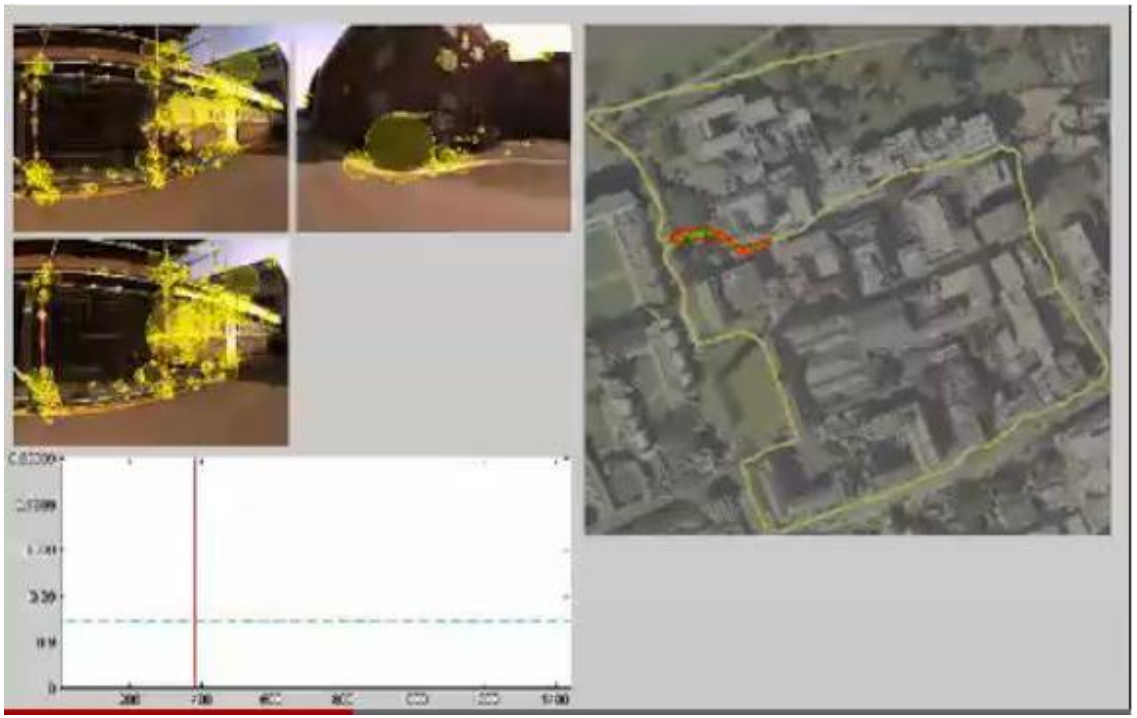
Images from KTH COLD dataset

Vision-based localization methods (Zivkovic *et al*. [6, 2]) tend to observe misclassifications at boundaries between places.



Courtesy of Zoran Zivkovic

Topological mapping (Cummins and Newman [1]) community use Bag-of-Words model and graphical models (Chow-Liu Tree) for loop-closure detection and scene classification.

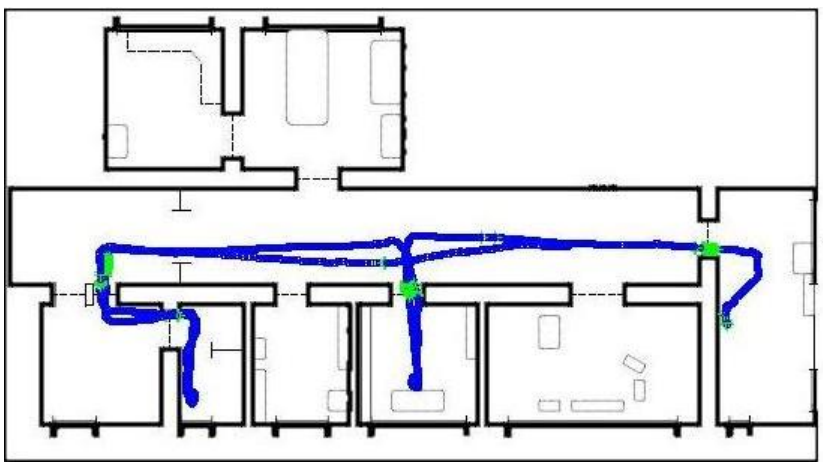


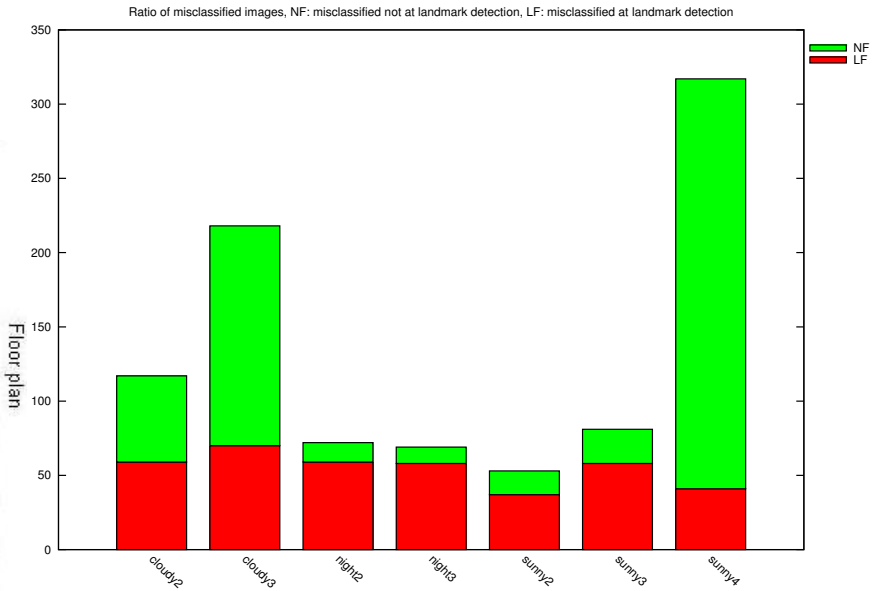
Courtesy of Mark Cummins

Related work (Ranganathan *et al*. [4]) refer to the boundaries as "landmarks", and we will adopt this naming convention.
Existing approaches (Wang *et al*. [5],Pronobis *et al*. [3]) for Semantic Recognition doesn't specifically consider "landmark" positions.

Hypothesis

- Do most of the misclassified samples come from the “landmark positions”?



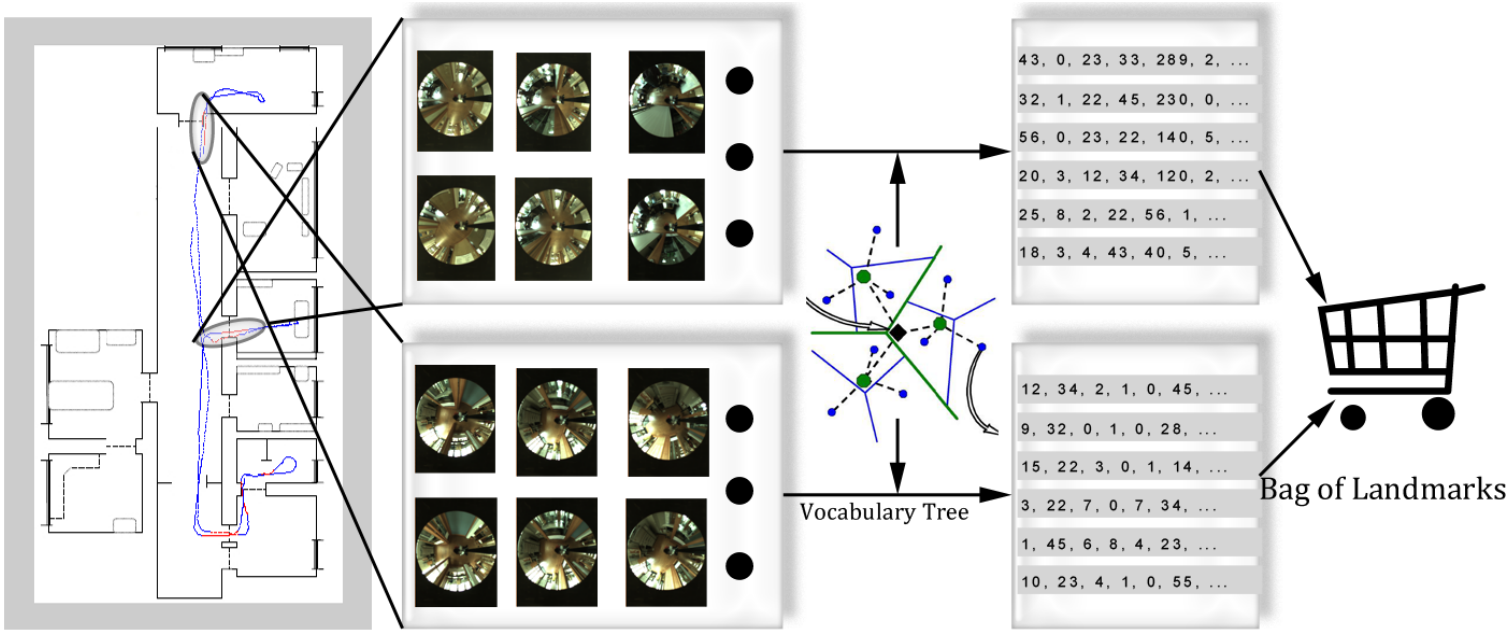


Detect the "Landmark"

- "Landmarks" re-defined (doorways, junctions, etc.)
- How should we deal with sample images taken from "landmark positions"?
- Can we detect the "Landmark" positions?
- – Unsupervised method (PLISS by Ananth Raganathan in RSS2010)
- – Supervised method (Our method, BoLTS)

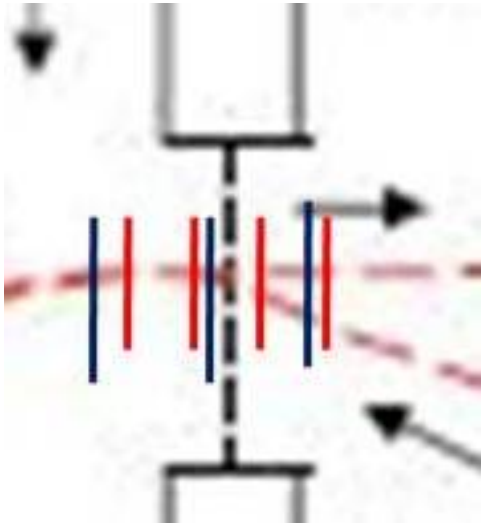
BoLTS (Bag-of-Landmarks using Time Series)

- Feature generation: Vocabulary Tree method.
- Feature integration: concatenate into a Time Series data format.
- Similarity measure: Dynamic Time Warping and Histogram Intersection Kernel.

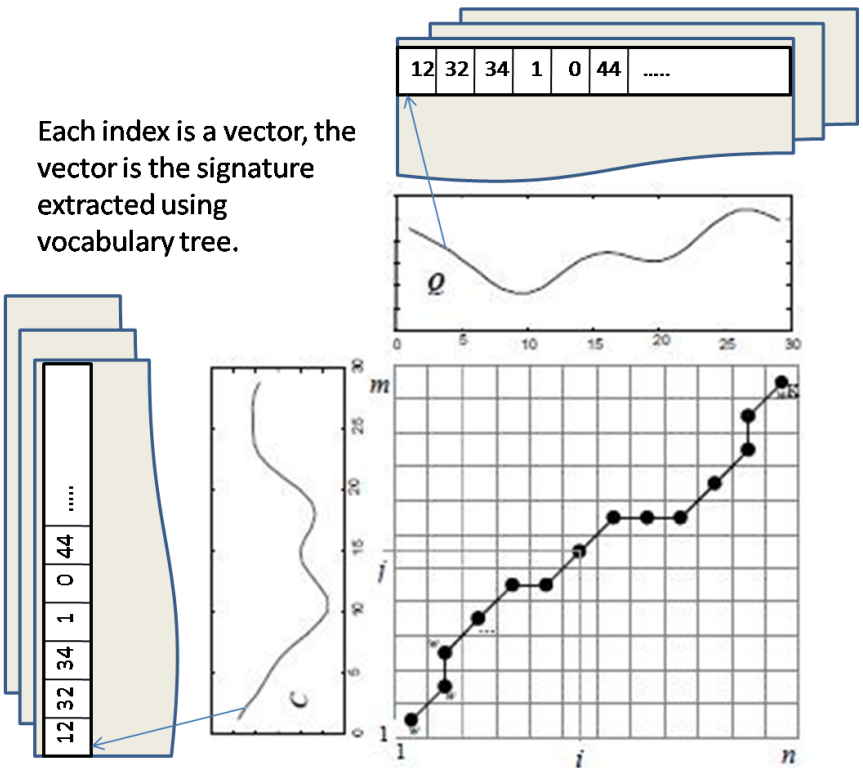


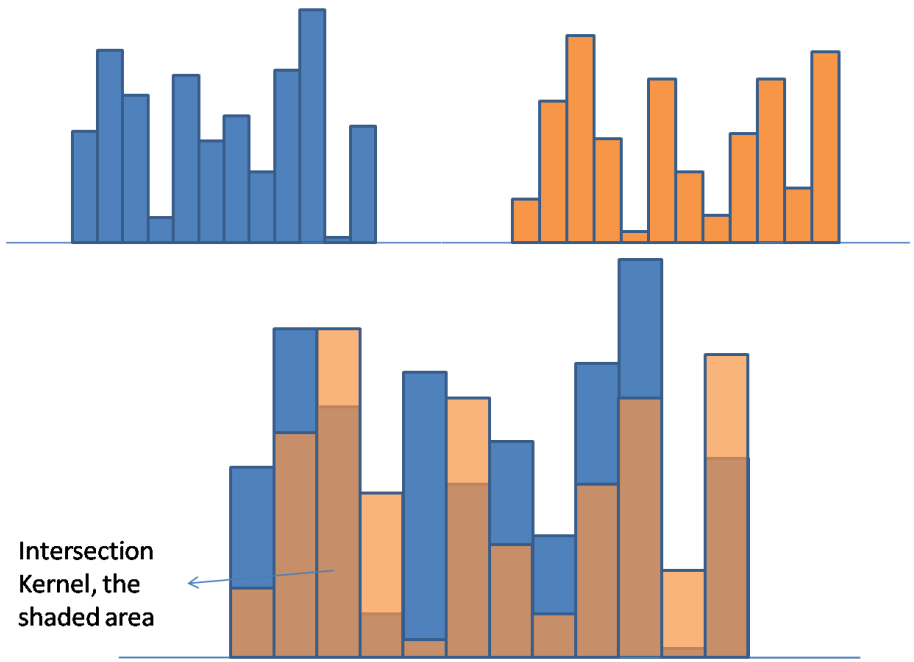
Dynamic Time Warping(DTW) and Iterative Vector-DTW I

- Why time-series?
- – Doorways are usually short of rich visual features.
- – Temporal constraint makes detection “Landmark” positions more robust.
- Dynamic Time Warping is in nature complied with the movement of robot.

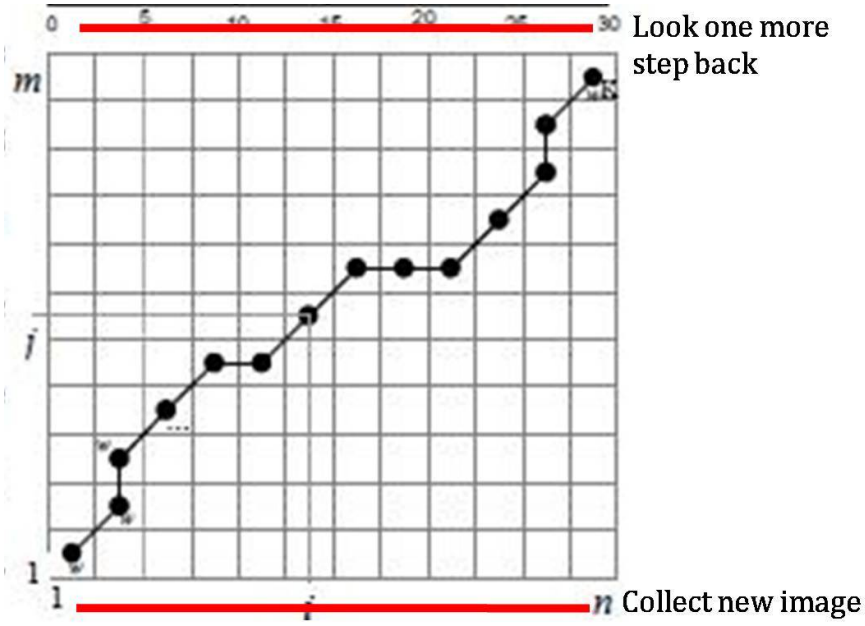


Dynamic Time Warping(DTW) and Iterative Vector-DTW II



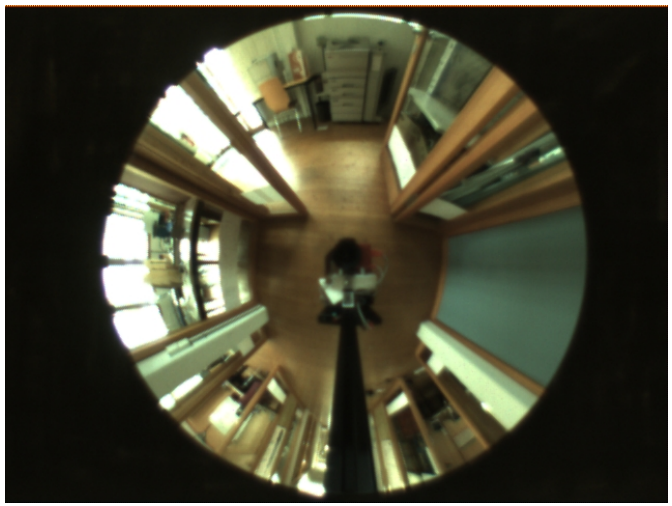


- Performance issues.
- – Each time index itself is a vector.
- – Images are collected incrementally, detection decisions need to be made in online mode instead of in batch mode.
- Perform DTW algorithm iteratively, memoizing calculated similarity measures d and the matching matrix D for dynamic programming.

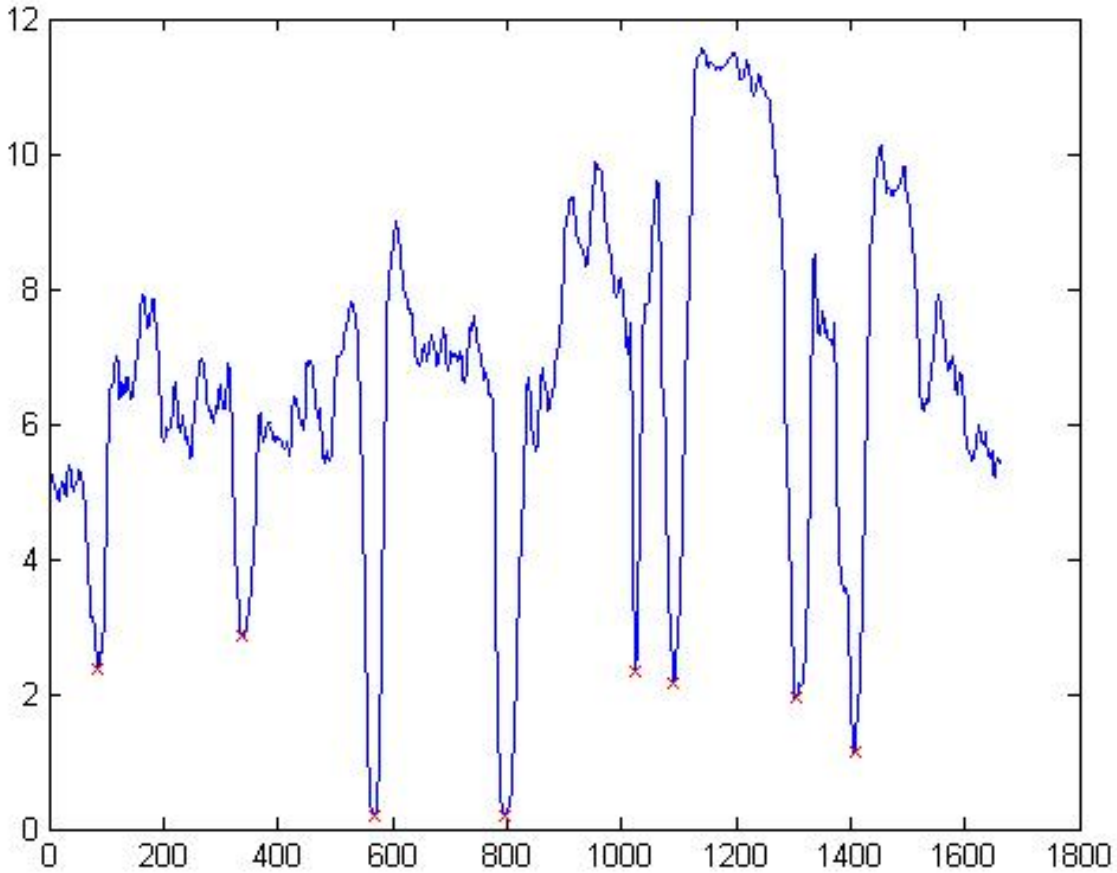


Experiment Setup

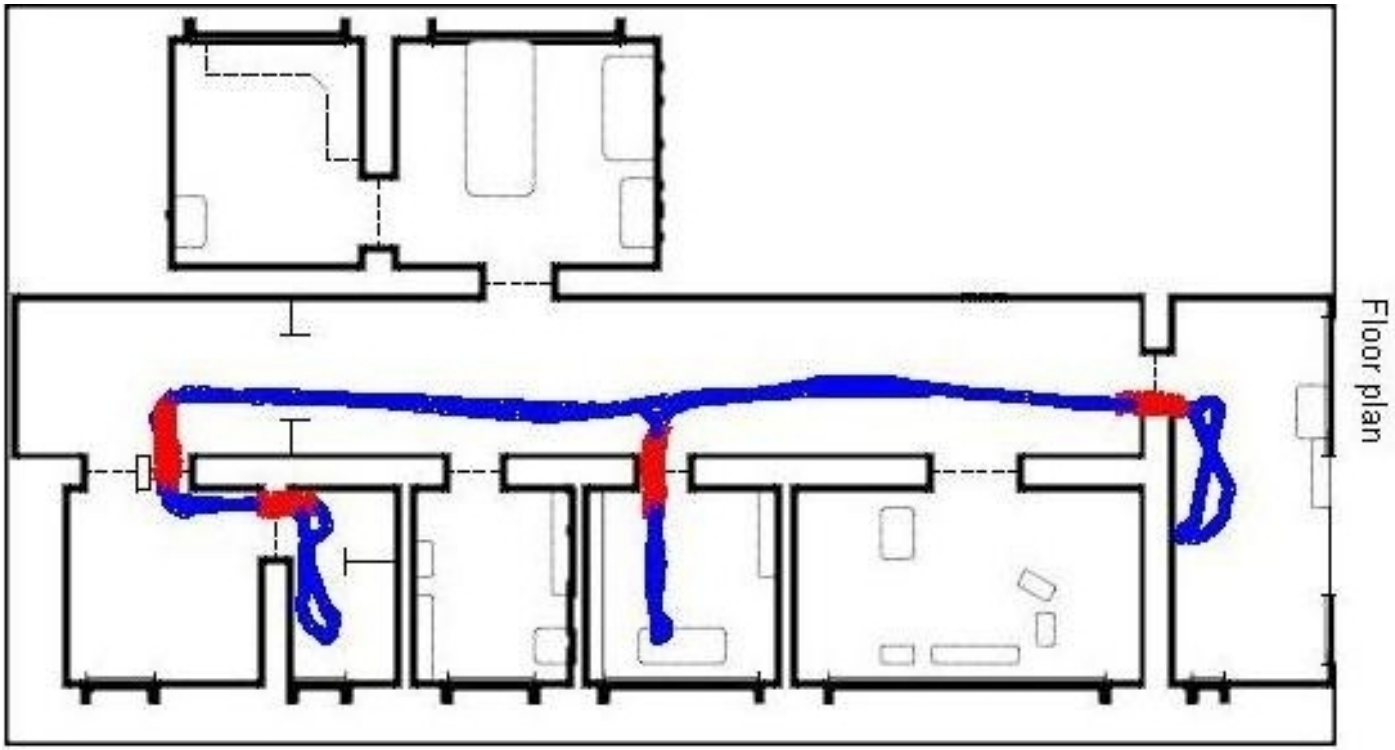
- Using the COLD dataset.
- Performed experiments on the same data as in (Wang *et al*. [5])
- Only panoramic images are used, no multi-model classification.



Landmark Detection

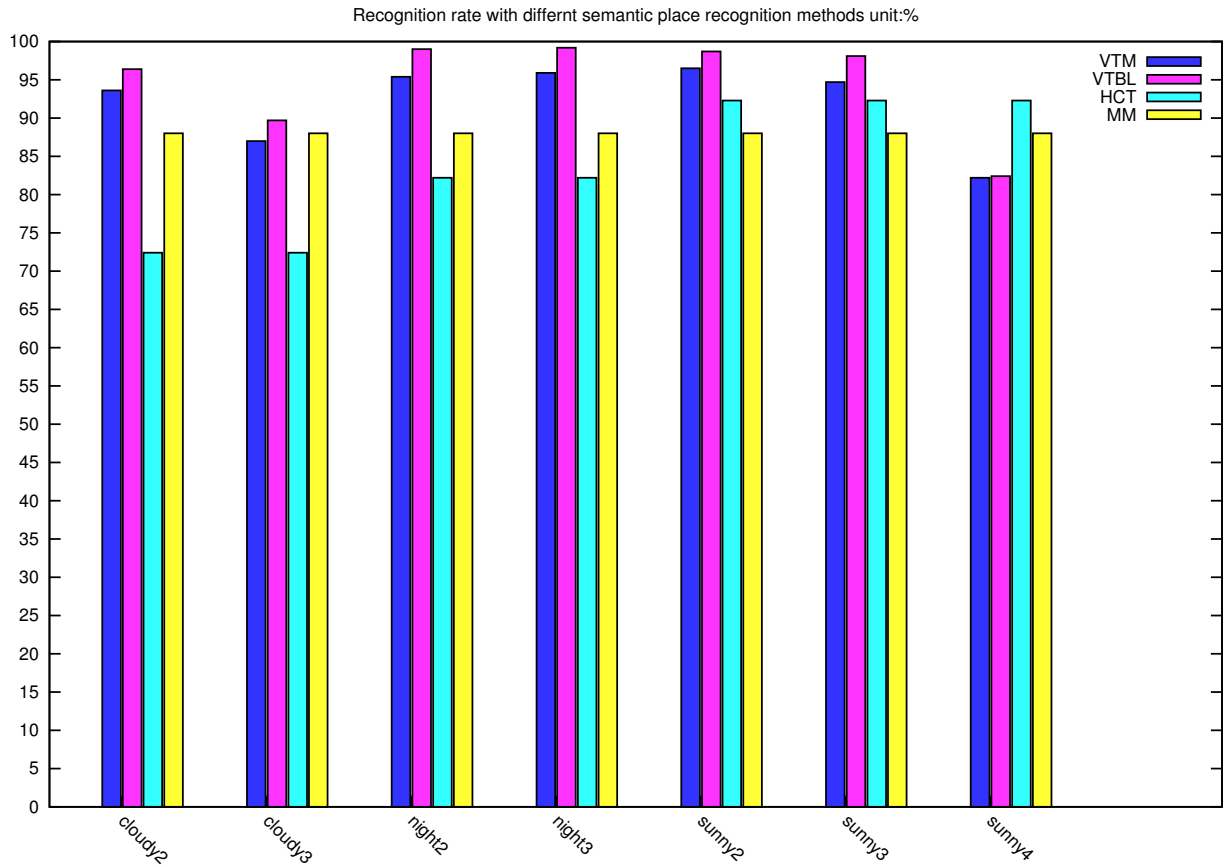


Video



./video/TPR.mp4

Comparison with alternative methods

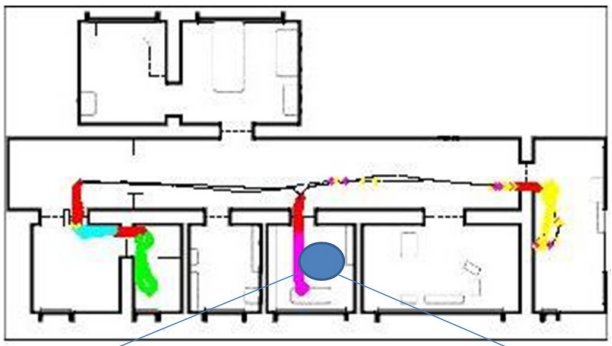


Conclusion

- Experimentally verified a hypothesis that misclassification in Semantic Place Recognition happens mostly in “landmark positions”.
- Proposed a supervised learning method “BoLTS” for detecting “landmarks”.
- This “landmark detection” approach is based on Time Series data, differen from most of other landmark detection approaches.
- A robust unsupervised method for detecting “landmark positions” will be useful for realizing “Semantic SLAM”.
- “Semantic SLAM” is a promising research topic developing together with “Semantic Mapping” and “Point Clouds”.

Future work

Now let's come back to the question, "what if we only label the transition(landmark)"?
What if we can devise an unsupervised learning method for landmarks?
What if the model of each semantic place can be obtained online? (Online learning of Graphical Models)



There is a gap between
human understanding and robot understanding

Thank you!

checkout `http://web.ics.purdue.edu/~yuanl/projects` for the matlab code.

M. Cummins and P. Newman.
FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance.
The International Journal of Robotics Research, 27(6):647–665, June 2008.

J. Knopp, J. Sivic, and T. Pajdla.
Avoiding confusing features in place recognition.
In *European Conference on Computer Vision*, volume 6311, pages 748–761. Springer Berlin / Heidelberg, 2010.

A. Pronobis, O. Martinez Mozos, B. Caputo, and P. Jensfelt.
Multi-modal semantic place classification.
The International Journal of Robotics Research, 29(2-3):298–320, 2010.

A. Ranganathan, E. Menegatti, and F. Dellaert.
Bayesian inference in the space of topological maps.

IEEE Transactions on Robotics, 22(1):92 – 107, Feburary 2006.


M.-L. Wang and H.-Y. Lin.
An extended-hct semantic description for visual place recognition.
The International Journal of Robotics Research, 30(8), July 2011.


Z. Zivkovic, B. Bakker, and B. Krose.
Hierarchical map building using visual landmarks and geometric constraints.
In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2480–2485, 2005.