Robotics: Science and Systems 2012
W11. Performance Comparison and Result Replication

Evaluation of Loop Detection in Visual SLAM

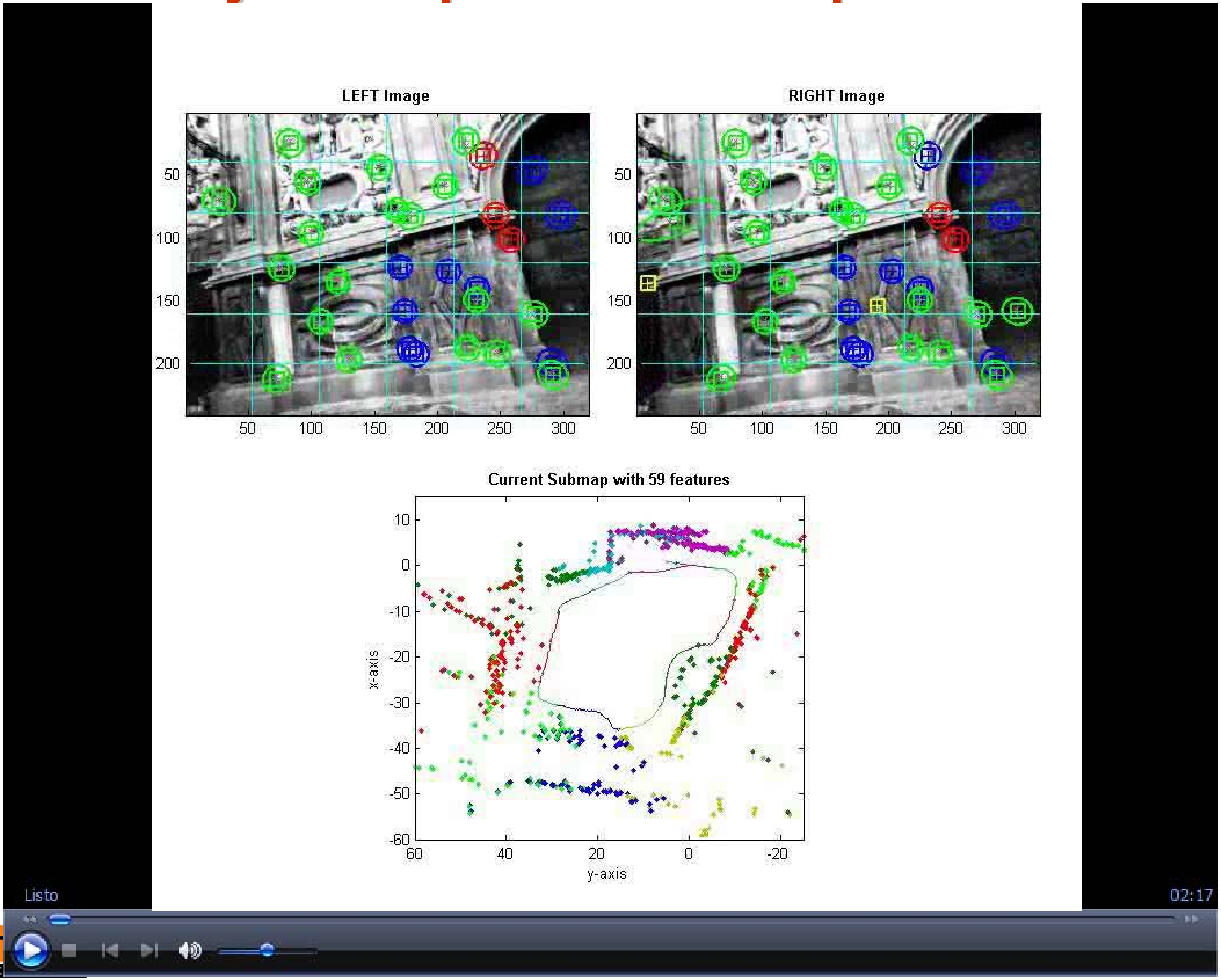
Dorian Gálvez-López, Juan D. Tardós
University of Zaragoza, Spain

Contributors: Cesar Cadena, José Neira, Lina Paz, Pedro Piniés

Outline

- Loop detection in Visual SLAM
- Our approach: Bags of Binary Words
- Evaluation of Loop Detection
- Conclusion

Why is Loop Detection Important?



Correct Map Topology and Geometry

Loop Detection Approaches

- Map to Map
  - Move the robot and build a local map
  - Match current local maps with previous local maps
    - » works for laser or sonar, too brittle for vision
- Image to Map
  - Build a visual feature map
  - Match features in the current image with map features
    - » Works well, but scales badly in large environments
- Image to Image (Appearance–Based)
  - Image features clustered into visual words (visual vocabulary)
  - For each image obtain a Bag-of-Words representation
  - Match BOWs of current and previous images
    - » Needs geometrical verification

Why is Loop Detection Difficult?

- Is this a loop closure?

Likely algorithm answer:

YES **YES** **TRUE POSITIVE**

Why is Loop Detection Difficult?

- Is this a loop closure?

Likely algorithm answer:

NO	**NO**	**TRUE NEGATIVE**
NO	**YES**	**FALSE POSITIVE**

Why is Loop Detection Difficult?

- Is this a loop closure?

Likely algorithm answer:

~~YES!~~	~~NO~~	~~FALSE NEGATIVE~~
NO	NO	TRUE NEGATIVE

Why is Loop Detection Difficult?

- Is this a loop closure?

Scene 1430

Scene 1244

Likely algorithm answer:

NO **YES** **FALSE POSITIVE**

Perceptual aliasing is common in some indoor scenarios

Why is Loop Detection Difficult?

- Is this a loop closure?

Scene 292

Scene 219

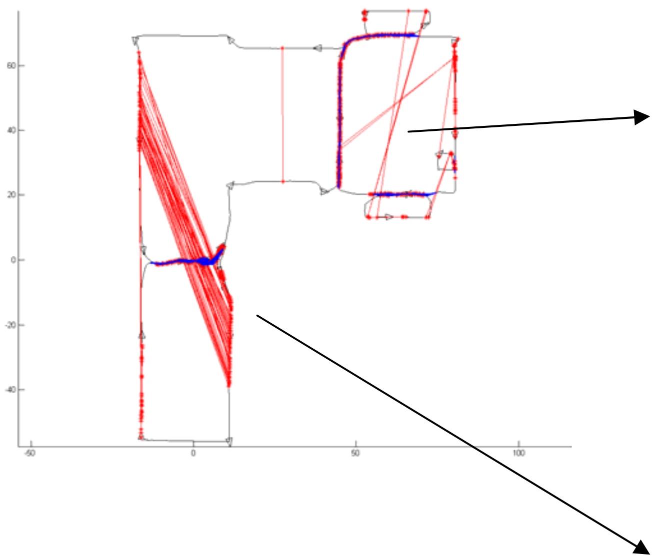
Likely algorithm answer:

NO **YES** **FALSE POSITIVE**

Specular perceptual aliasing!

False positives

BoW+epipolar



- False positives may ruin the map
  - But see two RSS 2012 papers that address this issue:
    - » Edwin Olson, Pratik Agarwal
    - » Yasir Latif, Cesar Cadena, José Neira,

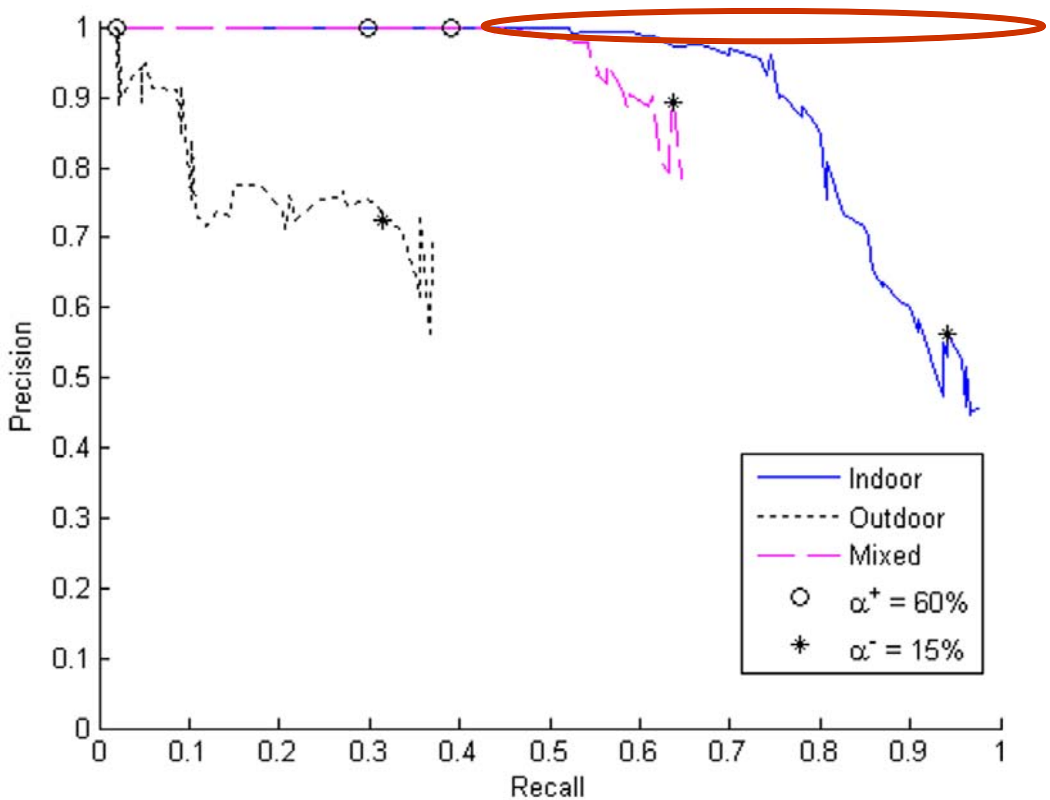
Common Metrics

$$\text{Precision} = \frac{\text{\# Correct detections}}{\text{\# Detections fired}} = \frac{TP}{TP + FP}$$

Desired: 100% precision,
No false positives

$$\text{Recall} = \frac{\text{\# Correct detections}}{\text{\# Existing Loops}} = \frac{TP}{TP + FN}$$

Desired: high recall,
Few false negatives



Ideal
working region

Juar

Outline

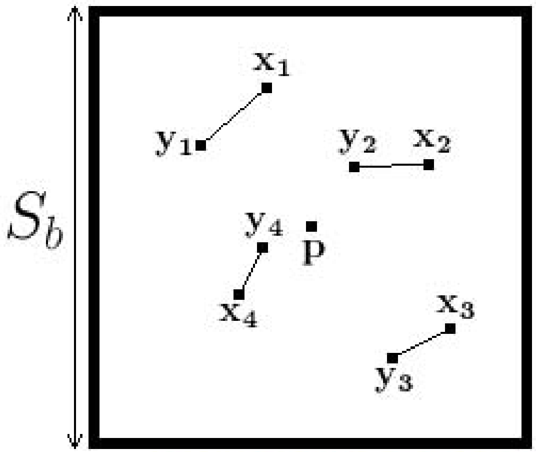
- Loop detection in Visual SLAM
- ➡ Our approach: Bags of Binary Words
- Evaluation of Loop Detection
- Conclusion

Bags of Binary Words

- Extract image features
  - FAST keypoint detector
  - BRIEF descriptor (binary)
- Convert into visual words
  - Binary version of the hierarchical vocabulary tree (Nister 2006)
  - Store the BOW representation of current image
- Search for matches with the previous images
  - Inverse index: which images contain some common word
- Check temporal consistency
  - with k previous matches
- Check geometric consistency: epipolar geometry
  - Direct index

BRIEF Binary Features

- BRIEF: Binary Robust Independent Elementary Features
  - Given a keypoint p, binary vector B of length L s.t:



Each bit, intensity comparison of two pixels:

$$B_i(\mathbf{p}) = \begin{cases} 1 & \text{if } \mathbf{p}+\mathbf{x_i} < \mathbf{p}+\mathbf{y_i} \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in [1..L]$$

Predefined random pixel coordinates:

$$\mathbf{x} = \mathcal{N}(0, \frac{1}{25} S_b^2), \quad \mathbf{y} = \mathcal{N}(\mathbf{x}, \frac{4}{625} S_b^2)$$

  - Computation time: 17 microseconds per keypoint

M. Calonder, V. Lepetit, C. Strecha, P. Fua: BRIEF: Binary Robust Independent Elementary Features. 11th European Conference on Computer Vision (ECCV), Heraklion, Crete. LNCS Springer, September 2010.

We use a patch of size S_b = 48 pixels and L = 256 bits

BRIEF Binary features

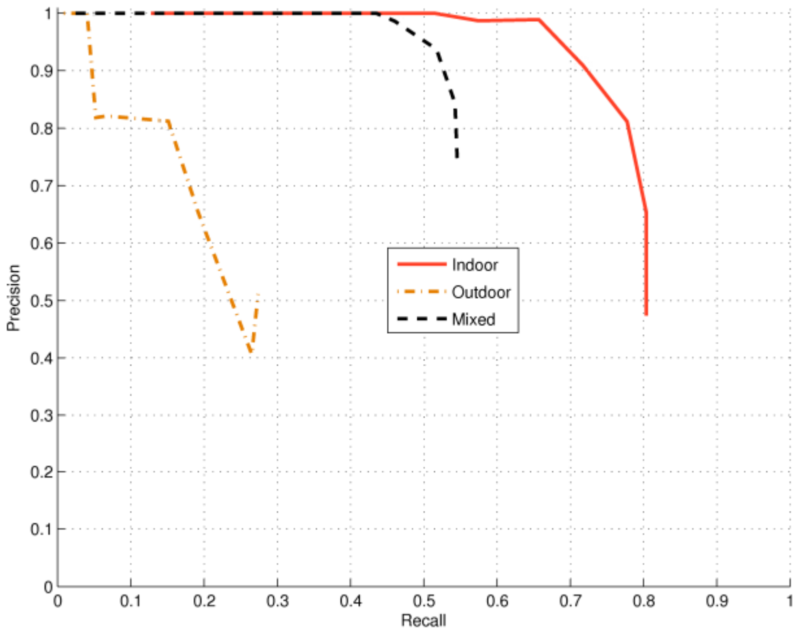
- Very fast to compute: 13ms per image
  - c.f. SURF: 100-400 ms
- Need less memory: 256 bits = 32 bytes
  - c.f. SURF of SIFT 64-128 bytes or floats
- Faster to compare: Hamming distance == xor
  - c.f. SURF or SIFT: Euclidean distance

- BUT not rotation and scale invariant

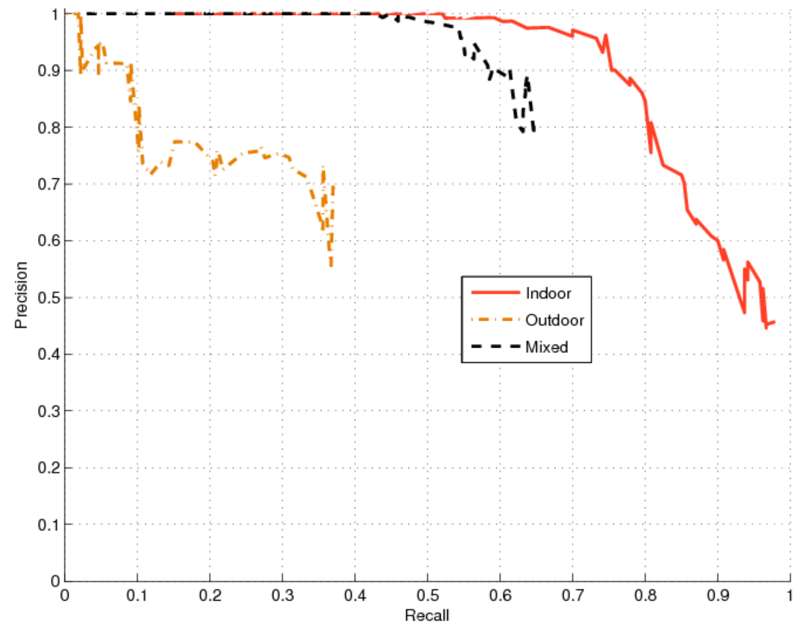
Are BRIEF features good for loop closing?

- BRIEF achieves results similar to SURF:

BRIEF



SURF*

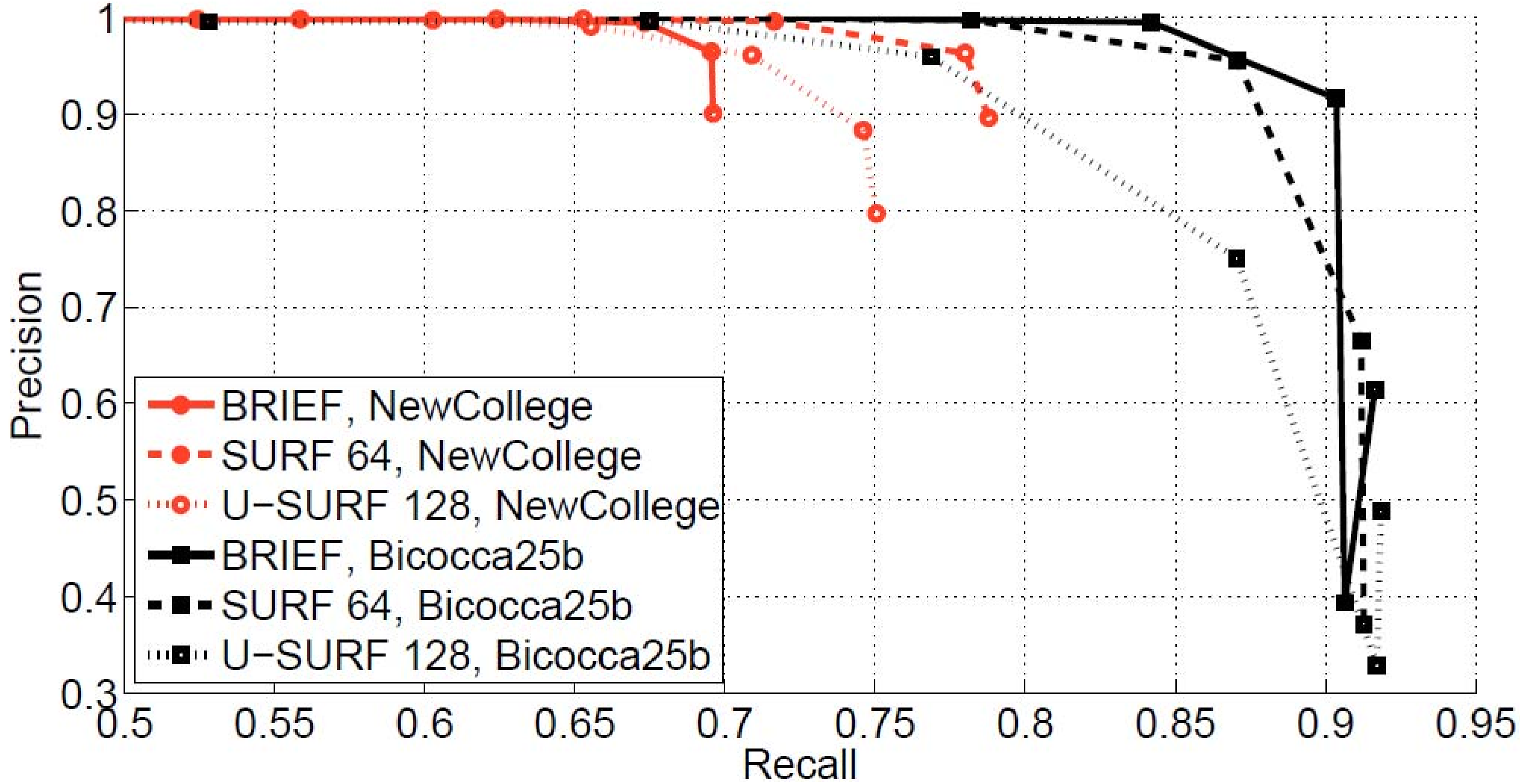


Without Geometrical Checking

C. Cadena, D. Gálvez-López, F. Ramos, J.D. Tardós, and J. Neira: **Robust place recognition with stereo cameras**. IROS 2010, pp. 5182–5189

Are BRIEF features good for loop closing?

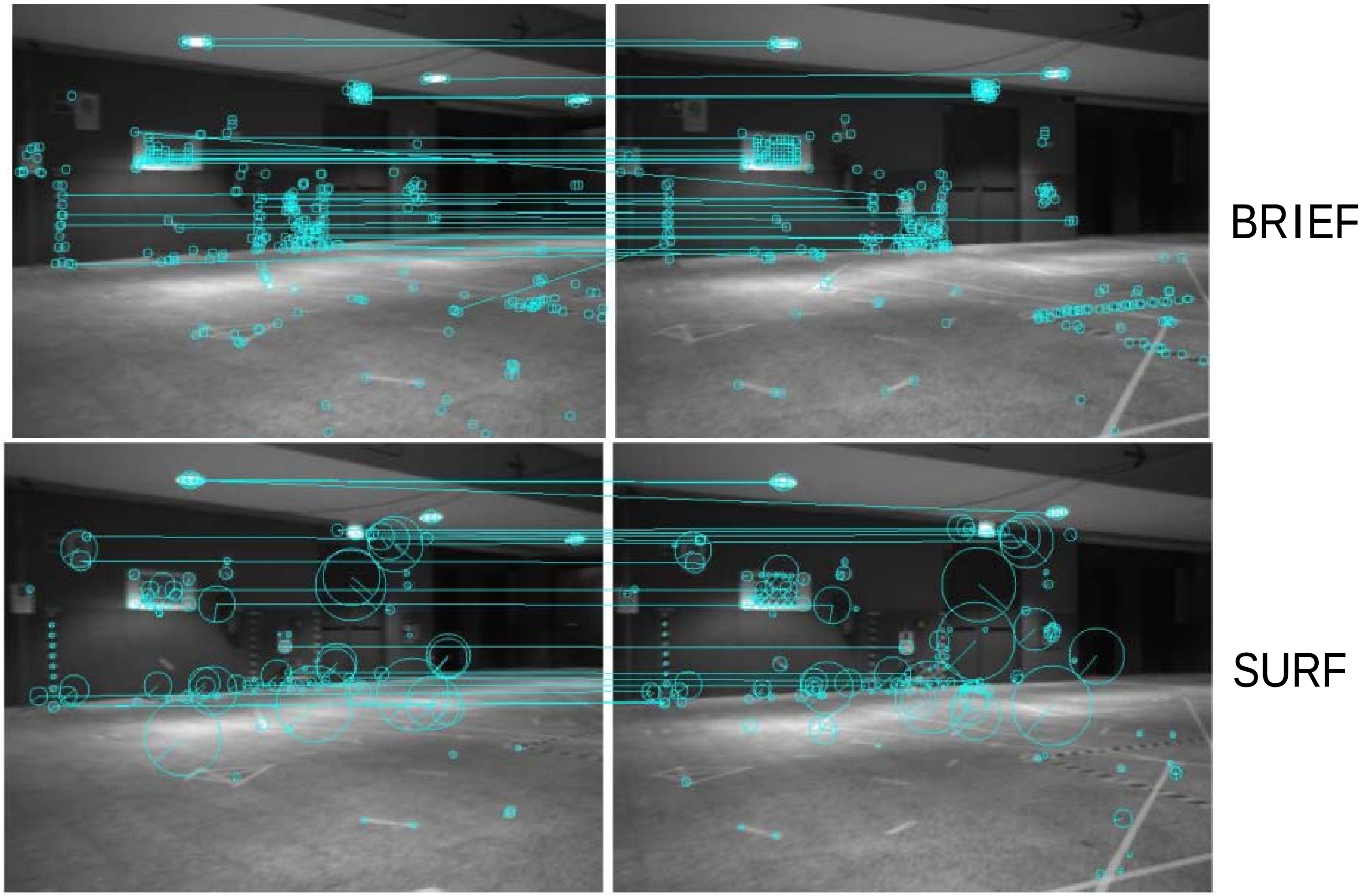
- BRIEF achieves results similar to SURF:



Without Geometrical Checking

BRIEF .vs. SURF

- Example of words matched by BRIEF and SURF:



BRIEF .vs. SURF

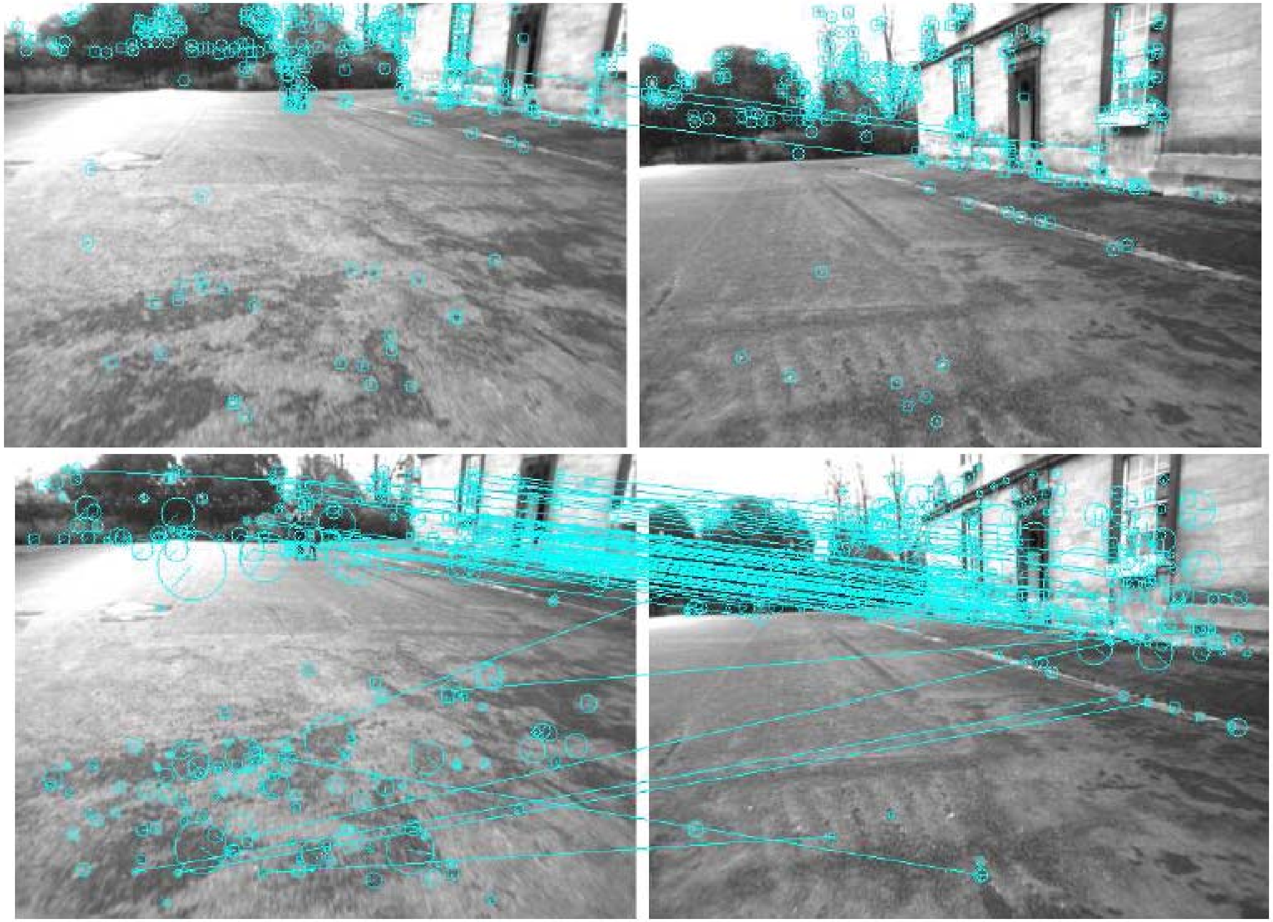
- Sometimes BRIEF works better

BRIEF

SURF

BRIEF .vs. SURF

- Sometimes BRIEF works worse

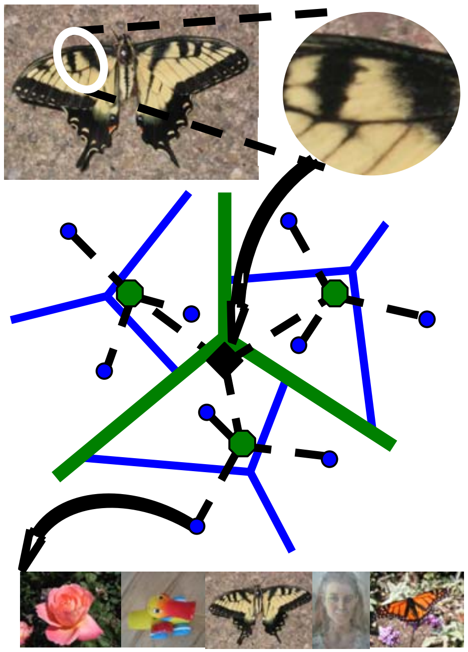


BRIEF

SURF

Bags of Binary Words

- Hierarchical vocabulary tree (Nister & Stewénius 2006)
  - Tree structure: branch factor 10, depth levels 6
  - Clustering with kmeans++
  - Created off-line



- Online:
  - Compute the BOW of current image

    $\mathbf{v}_k = (0,...0,\ v_k^i\ , 0,...0\ \ v_k^j\ , 0,\dots\ \ )$ tf-idf weights

  - Compare to previous images to find candidates

$$s(\mathbf{v}_1, \mathbf{v}_2) = 1 - \frac{1}{2}\left| \frac{\mathbf{v}_1}{|\mathbf{v}_1|} - \frac{\mathbf{v}_2}{|\mathbf{v}_2|} \right|$$

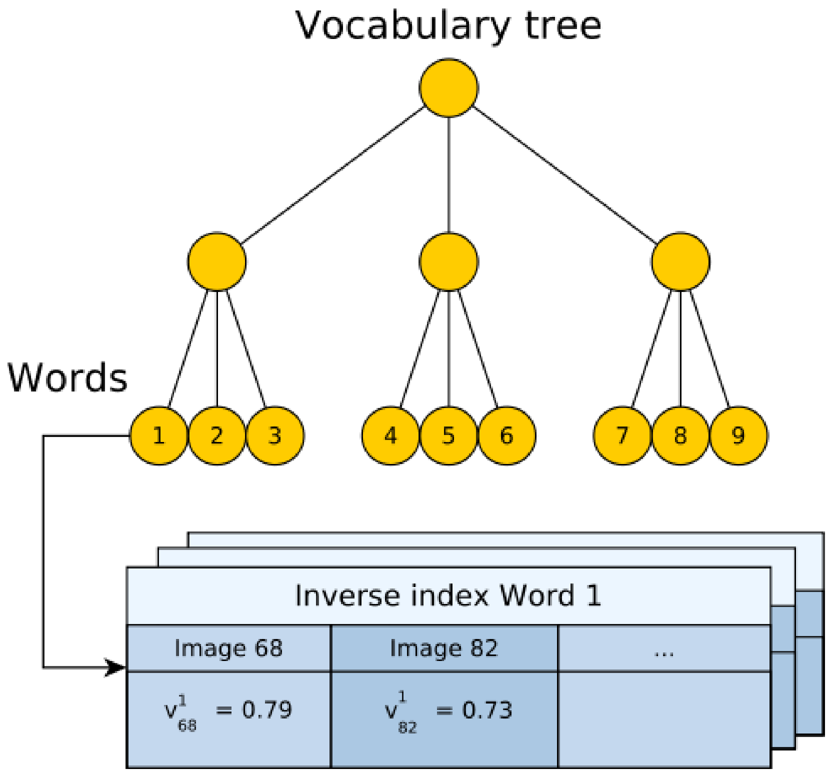
Image Similarity (L_1 norm)

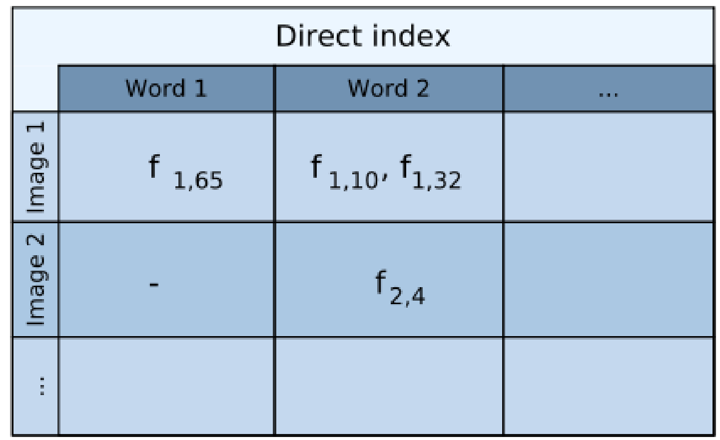
$$\eta(\mathbf{v}_t, \mathbf{v}_{t_j}) = \frac{s(\mathbf{v}_t, \mathbf{v}_{t_j})}{s(\mathbf{v}_t, \mathbf{v}_{t-\Delta t})}$$

Normalized Image Similarity

Image database

- Vocabulary tree + Inverse index + Direct index





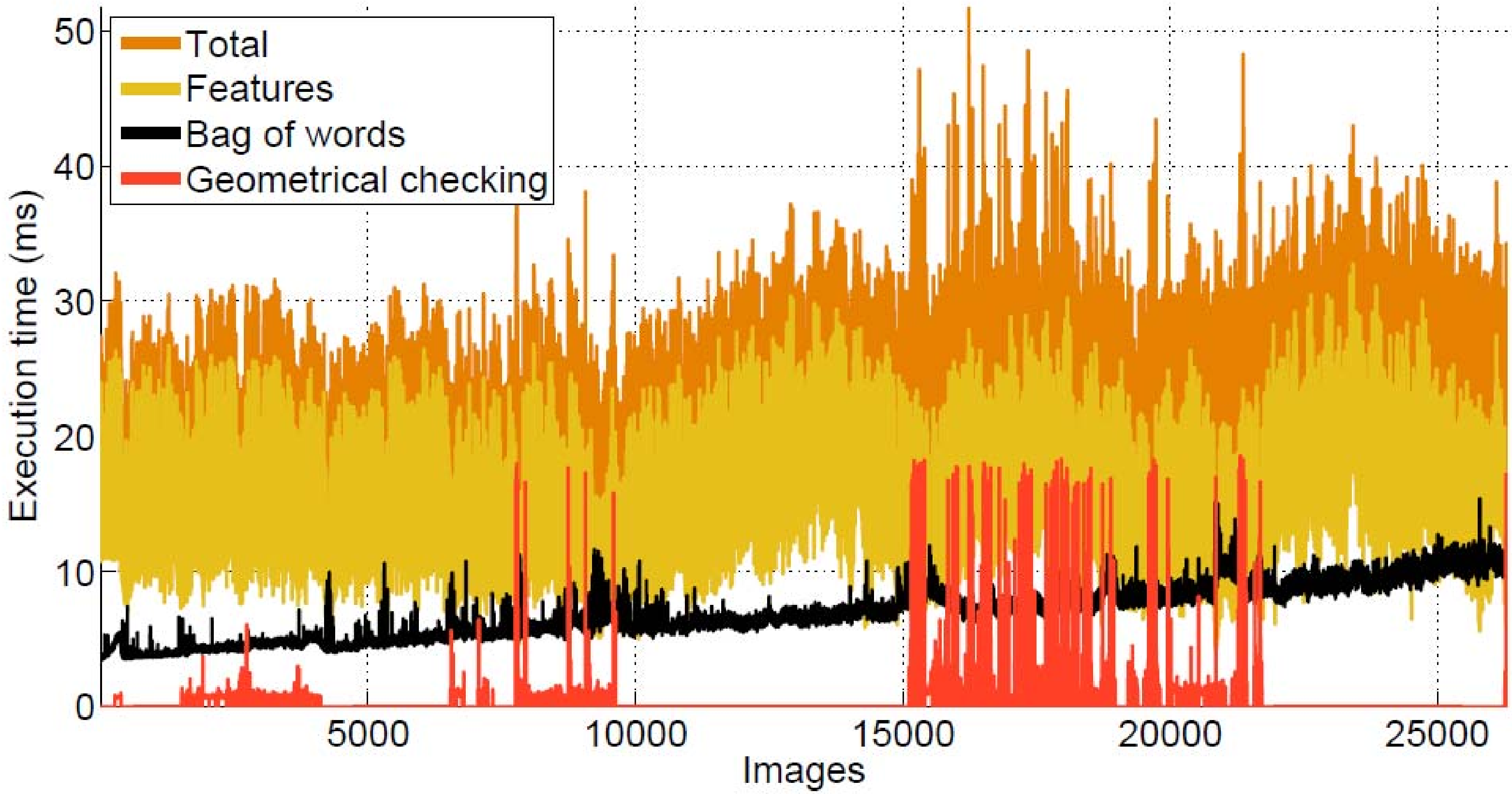
Only compare with images that have some word in common

Speed up correspondence search for verification of epipolar geometry

Very fast loop closing

- Execution time with 26K images:

  mean 21.6ms, max 52ms



One order of magnitude faster than previous approaches!!

Outline

- Loop detection in Visual SLAM
- Our approach: Bags of Binary Words
- ➡ Evaluation of Loop Detection
- Conclusion

Parameter tuning, how bad can it be?

TABLE IV
PARAMETERS

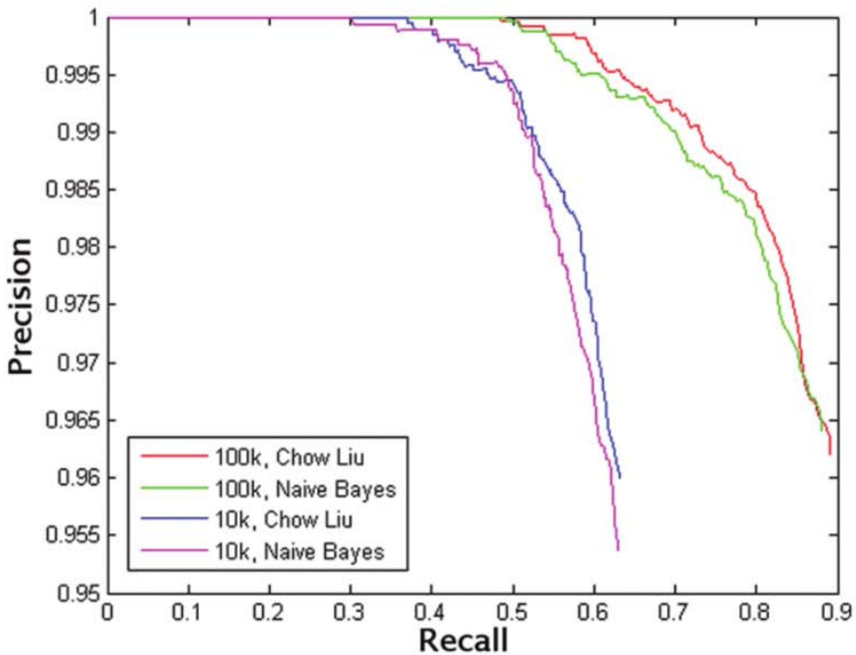
FAST threshold	10
BRIEF descriptor length (L_b)	256
BRIEF patch size (S_b)	48
Max. features per image	300
Vocabulary branch factor (k_w)	10
Vocabulary depth levels (L_w)	6
Min. score with previous image ($s(\mathbf{v}_t, \mathbf{v}_{t-\Delta t})$)	0.005
Temporally consistent matches (k)	3
Normalized similarity score threshold (α)	0.3
Direct index level (l)	2
Min. matches after RANSAC	12

TABLE IV
FAB-MAP 2.0—PARAMETERS FOR THE EXPERIMENTS

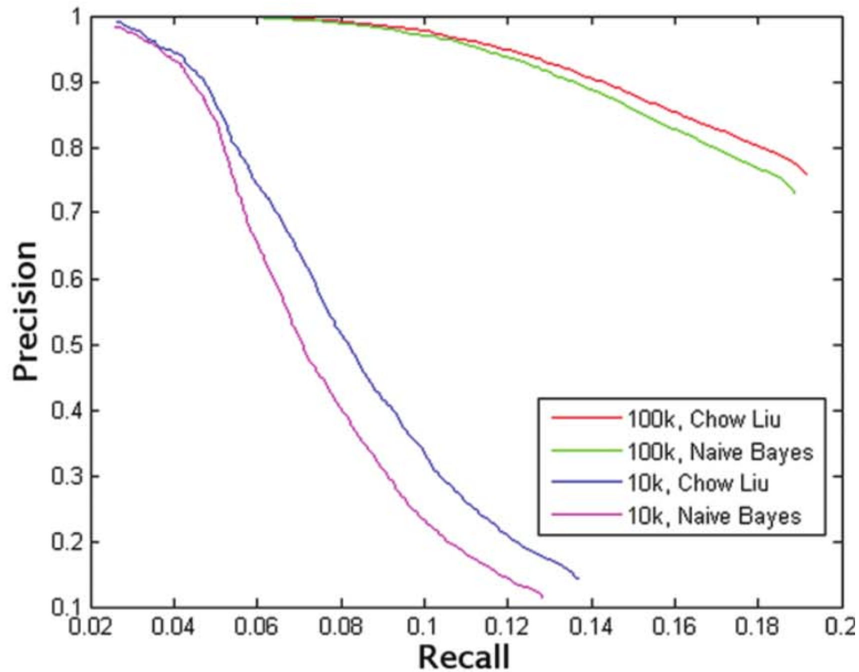
		Outdoor	Indoor	Mixed
	default	modified		
p	0.99	0.96	0.5	0.3
$P(\text{obs}\|\text{exist})$	039	0.39	0.31	0.37
$P(\text{obs}\|\neg\text{exist})$	0.05	0.05	0.05	0.05
$P(\text{newplace})$	0.9	0.9	0.9	0.9
σ	0.99	0.99	1.0	1.0
Motion Model	0.8	0.8	0.8	0.6
Blob Resp. Filter	25	25	25	25
Dis. Local	20s	20s	20s	20s

Precision-recall curves plot the performance as the main parameter changes

What's wrong with precision-recall curves?



(a)



(b)

- They tell us that for **some parameter value** the performance is good
- But is the parameter consistent across different experiments?

Avoid Overfitting

Usual Approach

Post-Tuning

Take an available dataset

Repeat

    Tune parameters

    Run your method on it

Until satisfied

Plot results

Write paper

Repeated Post-Tuning

Take several available dataset

For each dataset

    Repeat

        Tune parameters

        Run your method on it

    Until satisfied

    Plot results

End For

Write paper

OVERFITTING

Impossible to see the future is (Yoda 2002)

Proposed Approach

Avoid OverFitting

Take several dataset of **different** types

Some for training, some for evaluation (never peek into these)

Repeat

  Tune parameters

  Run on the **training** datasets

Until satisfied

Freeze parameters

For all datasets

  Run your method

  Plot results

End For

Write paper

And you can claim **robust** performance on a wide range of real scenarios

www.rawseeds.org

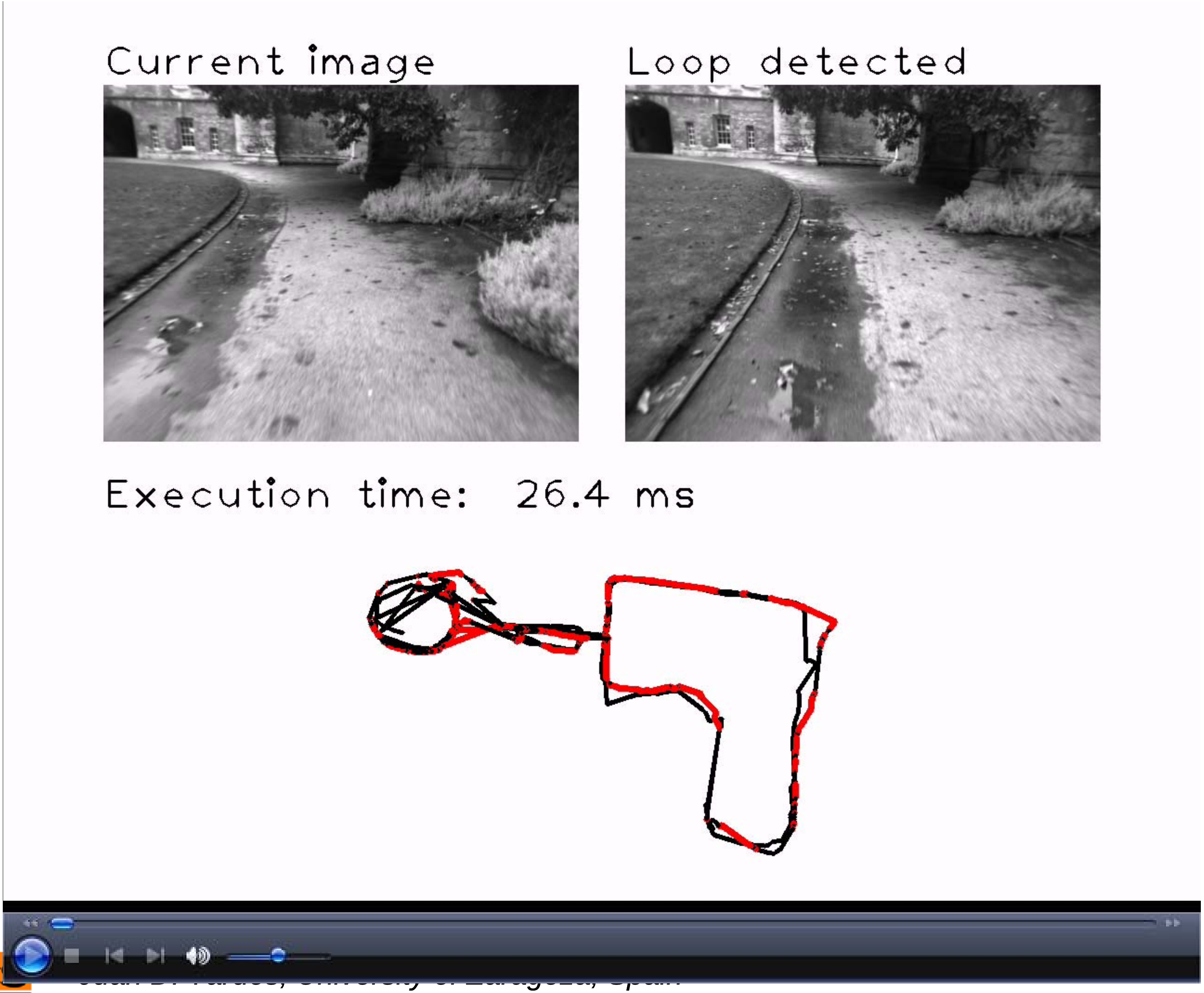
- Benchmark for SLAM algorithms
- Indoor and Outdoor multisensor datasets
  - Odometry and IMU
  - Sonar and Laser sensors: (Sick & Hokuyo)
  - Monocular, trinocular and panoramic vision
- Ground truth available
- Excellent benchmark for visual SLAM in the next years:
  - Size of datasets allows to test the scalability of the algorithms
  - GT allows to asses the accuracy
  - Challenging loop closings

3 Datasets for tuning, 2 for evaluation

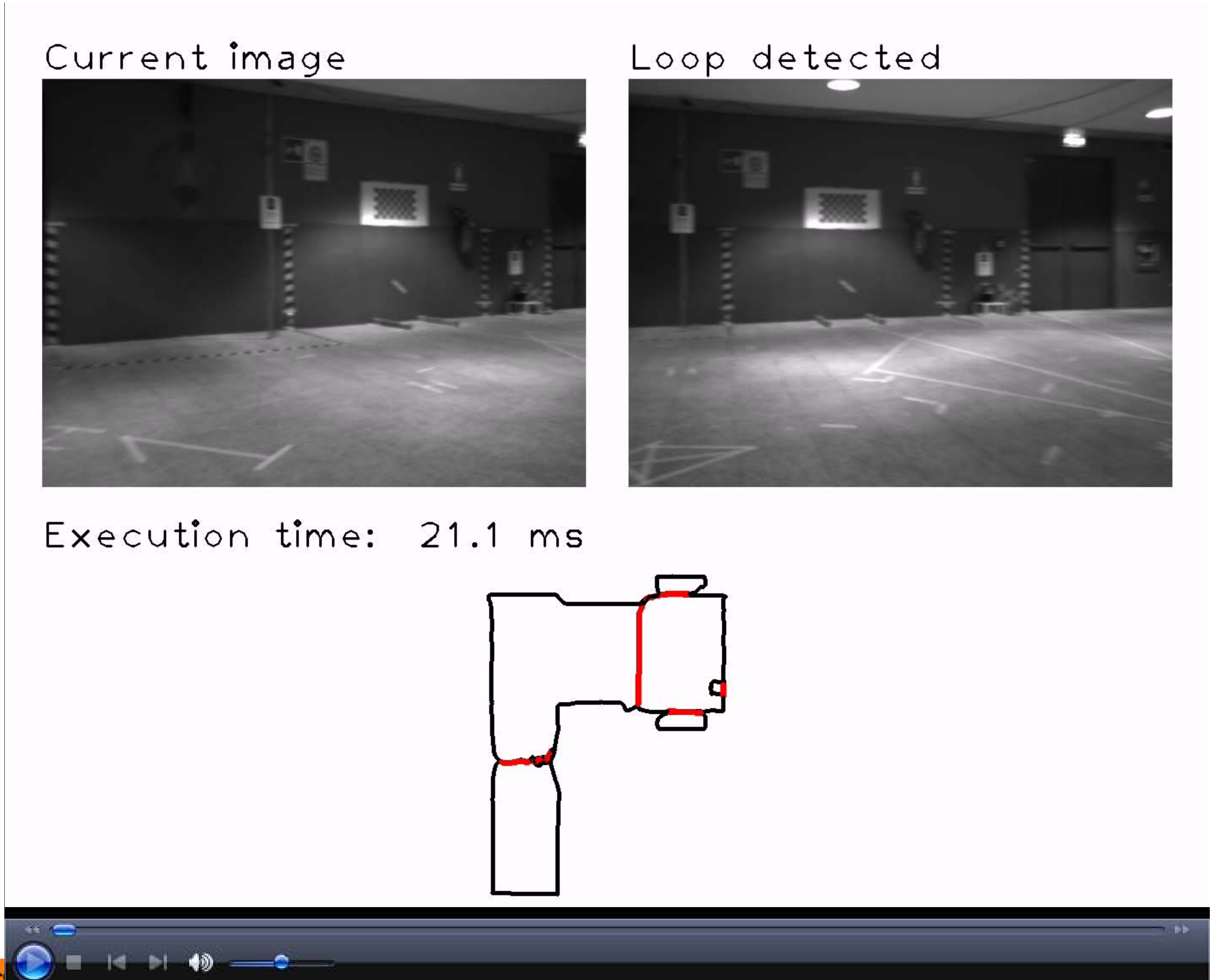
Dataset	Camera	Description	Total length (m)	Revisited length (m)	Avg. Speed $(m \cdot s^{-1})$	Image size (px × px)
New College [23]	Frontal	Outdoors, dynamic	2260	1570	1.5	512×384
Bicocca 2009-02-25b [24]	Frontal	Indoors, static	760	113	0.5	640×480
Ford Campus 2 [25]	Frontal	Urban, slightly dynamic	4004	280	6.9	600×1600
Malaga 2009 Parking 6L [26]	Frontal	Outdoors, slightly dynamic	1192	162	2.8	1024×768
City Centre [2]	Lateral	Urban, dynamic	2025	801	-	640×480




Example results: NewCollege



Example result: Rawseeds, indoor



Results

- No false positives, high recall:

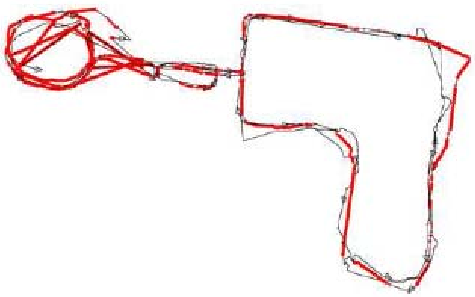
TABLE V
PRECISION AND RECALL OF OUR SYSTEM

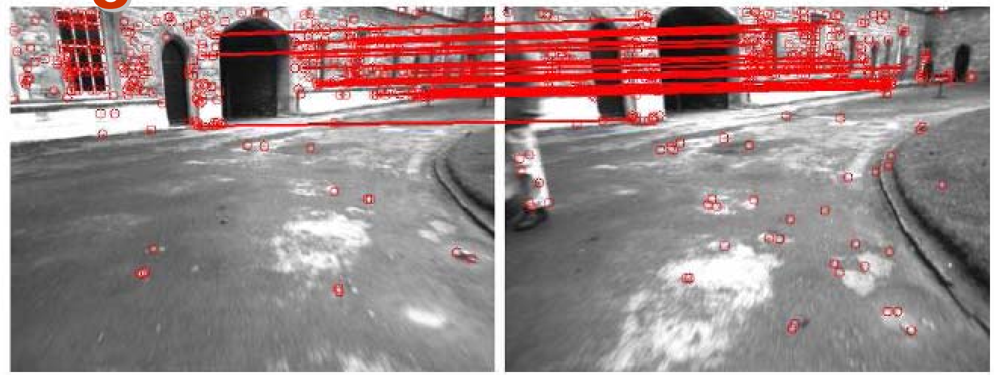
Dataset	# Images	Precision (%)	Recall (%)
NewCollege	5266	100	55.92
Bicocca25b	4924	100	81.20
Ford2	1182	100	79.45
Malaga6L	869	100	74.75
CityCentre	2474	100	30.61

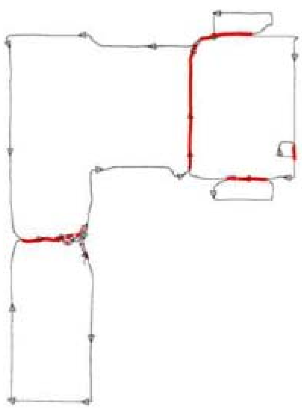
TABLE VI
PRECISION AND RECALL OF FAB-MAP 2.0

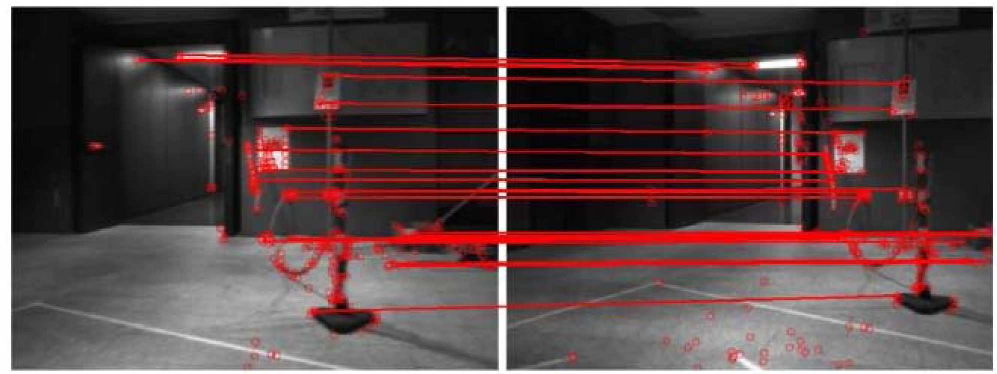
Dataset	# Images	Min. p	Precision (%)	Recall (%)
Malaga6L	462	98%	100	68.52
CityCentre	2474	98%	100	38.77

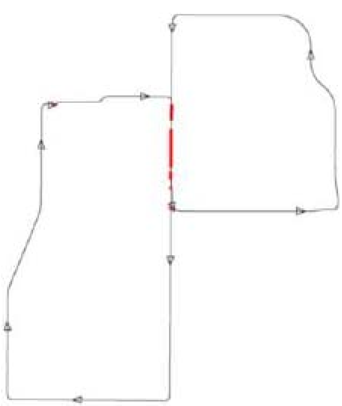
Tuning datasets

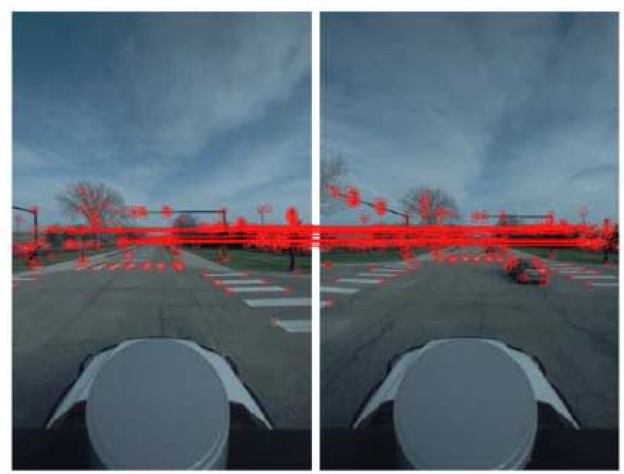




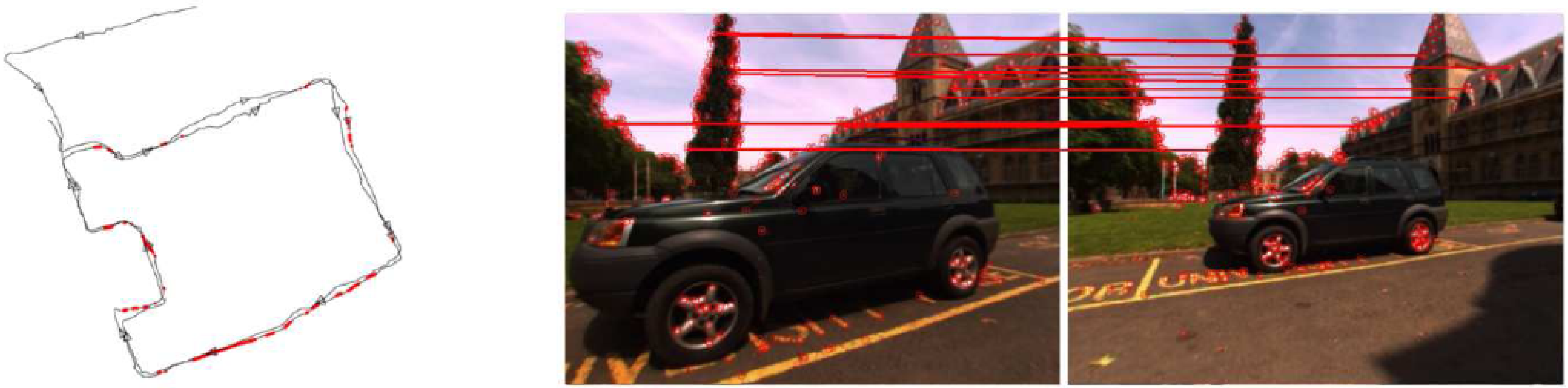








Validation Datasets



D. Gálvez-López, J. D. Tardós: **Bags of Binary Words for Fast Place Recognition in Image Sequences**. IEEE Transactions on Robotics, 2012 (in press)

Conclusions

- Loop detection with BRIEF features is:
  - One order of magnitude faster
  - Reliable for 2D camera motions
- Consistent results for diverse datasets, with the SAME parameters and vocabulary
- Big vocabularies speed-up matching
- But BRIEF lacks rotation and scale invariance
  - ORB, BRISK, …

Take-Home Messages

- Compare to previous approaches
- Evaluate the merit of each part of your algorithm
- Use available datasets, as diverse as possible
- Avoid over-fitting
  - Separate tuning and validation datasets
  - Don’t peek into the validation datasets
  - Report results with a fixed configuration for all datasets