Ensemble of Bayesian Filters for Loop Closure Detection

Mohammad Omar Salameh, Aziz Abdullah, Shahnorbanun Sahran

Pattern Recognition Research Group
Center for Artificial Intelligence Technology
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia

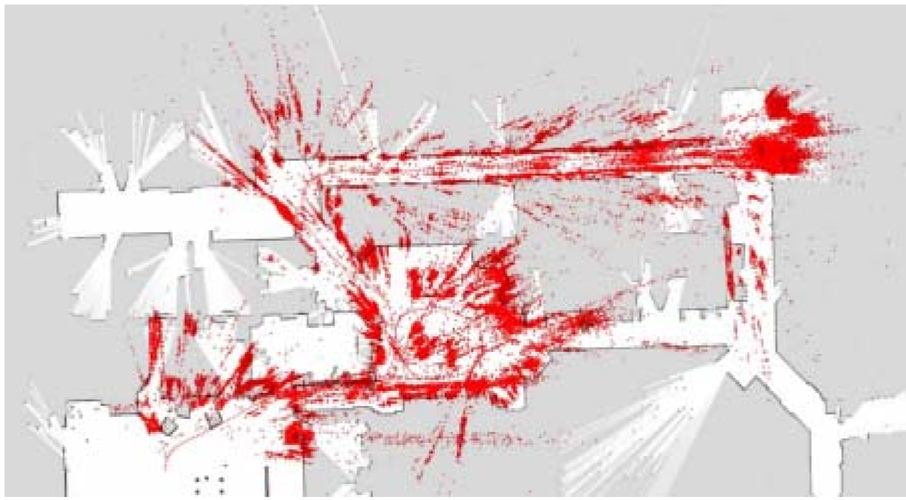
Presented by: Azizi Abdullah

Agenda

1. Introduction
2. Key Idea
3. Problem Statement
4. RTAB-Map (Real-Time Appearance-Based Mapping)
5. Mean Rule Ensemble Algorithm
6. Experiment and Results
7. Conclusion and Future Work

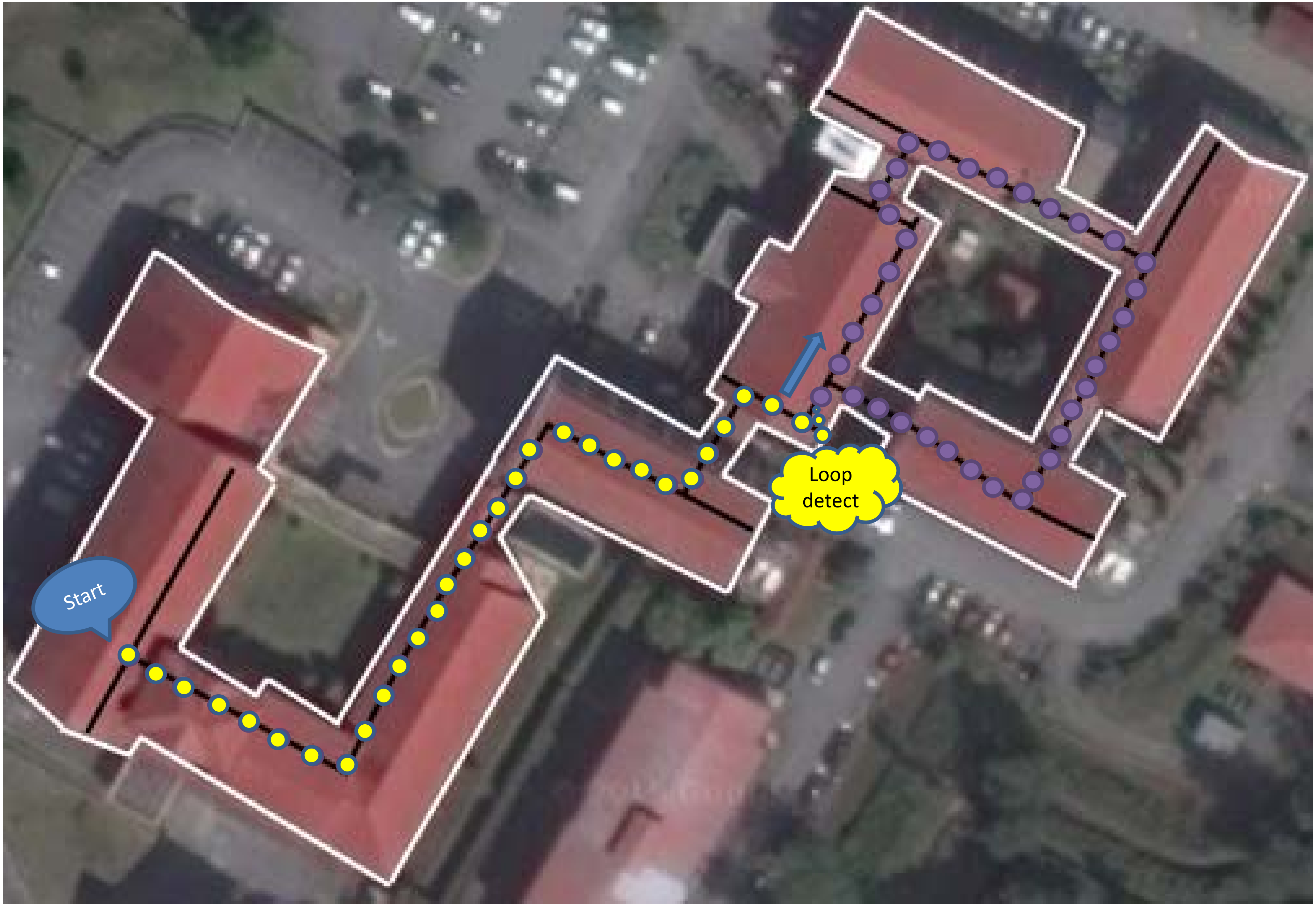
Introduction

- Visual Simultaneous Localization and Mapping (VSLAM) systems are widely used by mobile robots and/or autonomous vehicles to build up a map within an unknown environment using a vision sensor.



Introduction

- One of widely research topics in VSLAM is loop-closure detection.
- Loop-closure detection is a problem that require robots to recognize a previously visited place from current vision sensor measurement.
- Loop-closure detection also relevant for addressing the following problems in VSLAM:
  - Global localization
  - Robot kidnapping (a robot is moved by something it does not control)

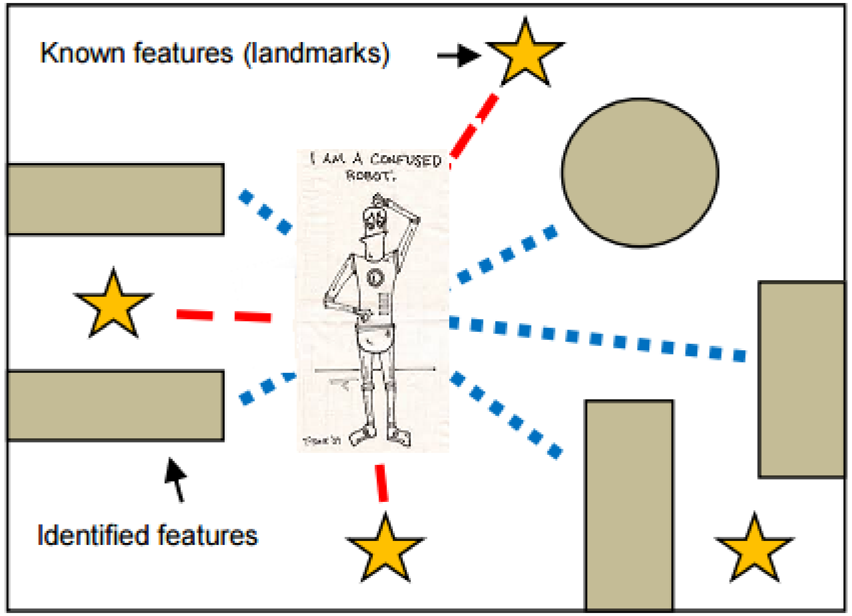

Start
Loop detect

Introduction

- Many algorithms have been proposed to solve loop-closure using vision sensors.
  - Angeli et. al. (2008): using a concept of visual words for landmark description and Bayesian filter to estimate the loop-closure probability
  - M. Labbé and F. Michaud (2011): using the bags of visual features for image description and Bayesian filter to estimate the loop-closure probability. Plus memory management unit for managing and storing previous locations namely RTAB-Map (Real-Time Appearance-Based Mapping)

Introduction

- It is common for the mobile robots to get confused of locations that have been visited as a map database of images (data frames) grows.
- Therefore, effective indexing and searching in the large image map database is needed and remains a challenge for VSLAM.



Problem Statement

- One of the important components in VSLAM is feature description for describing visual places such as local descriptor (SURF, SIFT) and global descriptor (color histogram).
- Single Descriptor vs Multi-descriptor: Content-based search and categorization have been shown to be effective in getting meaningful information from the query of digital images.
- Multi-Descriptor: A possible problem of the naive solution to create one large input vector which can increase problems of overfitting and hinder generalization performance.

Key Idea

- RTAB-Map (M. Labbé and F. Michaud, 2011)
  - Real-Time Appearance-Based Mapping for loop closure detection
  - Bags of Features (BOF) from key point detectors and descriptors
    - Vector E.g. SIFT/SURF
    - Binary E.g. Oriented FAST and Rotated BRIEF (ORB) – (a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to enhance the performance.)
  - Bayesian filter for loop-closure detection
  - Memory management
  - http://introlab.github.io/rtabmap/
- Ensemble of Bayesian filters (this work)
  - Combine multiple descriptor outputs using Bayesian filters for loop-closure detection

RTAB-Map

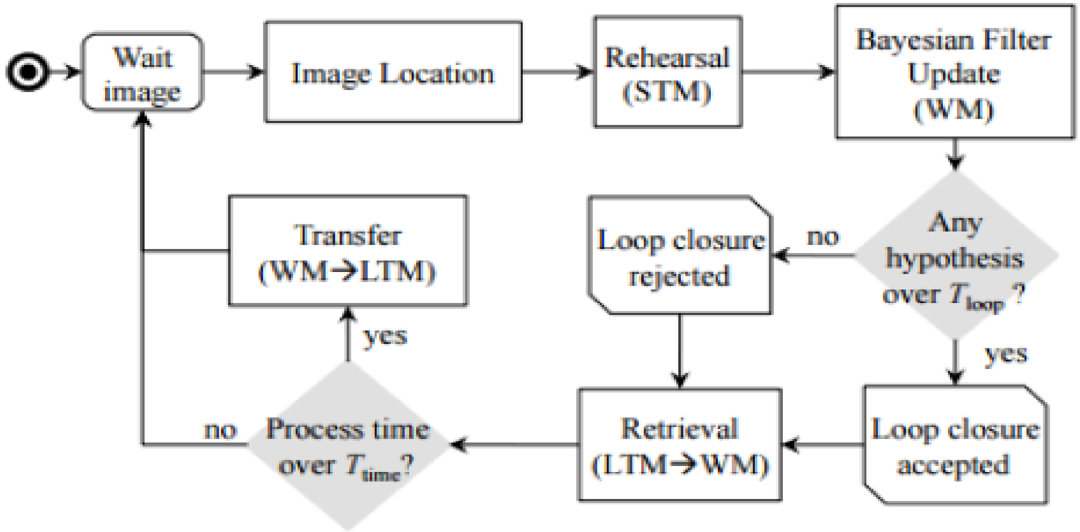
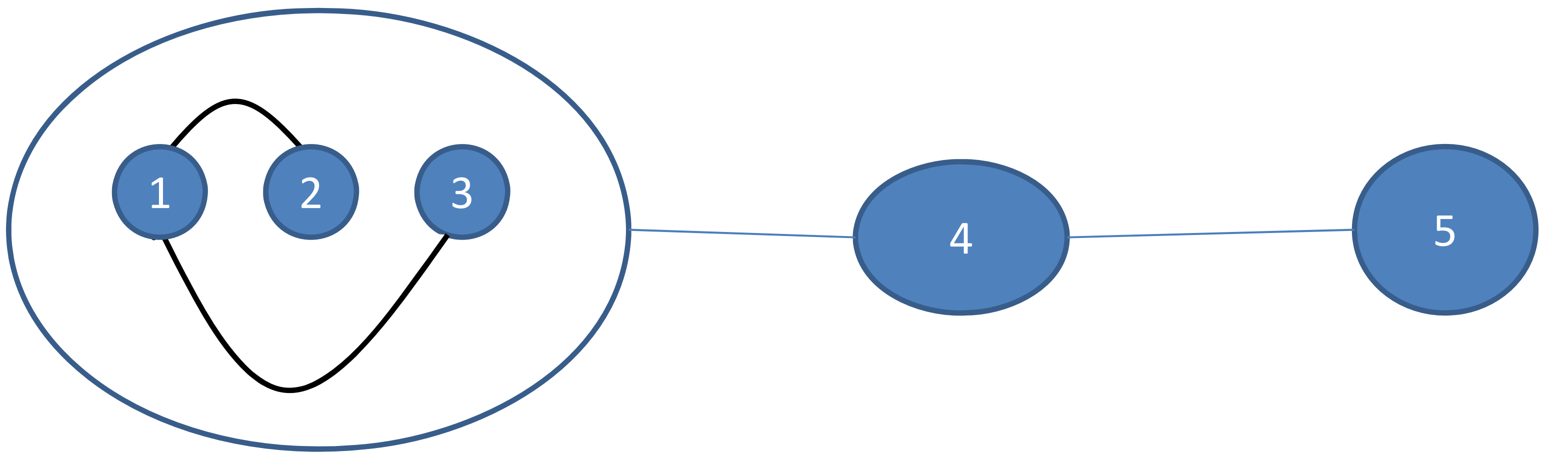


Fig. 1. Architecture of the RTAB-Map memory management loop closure detetcion steps.

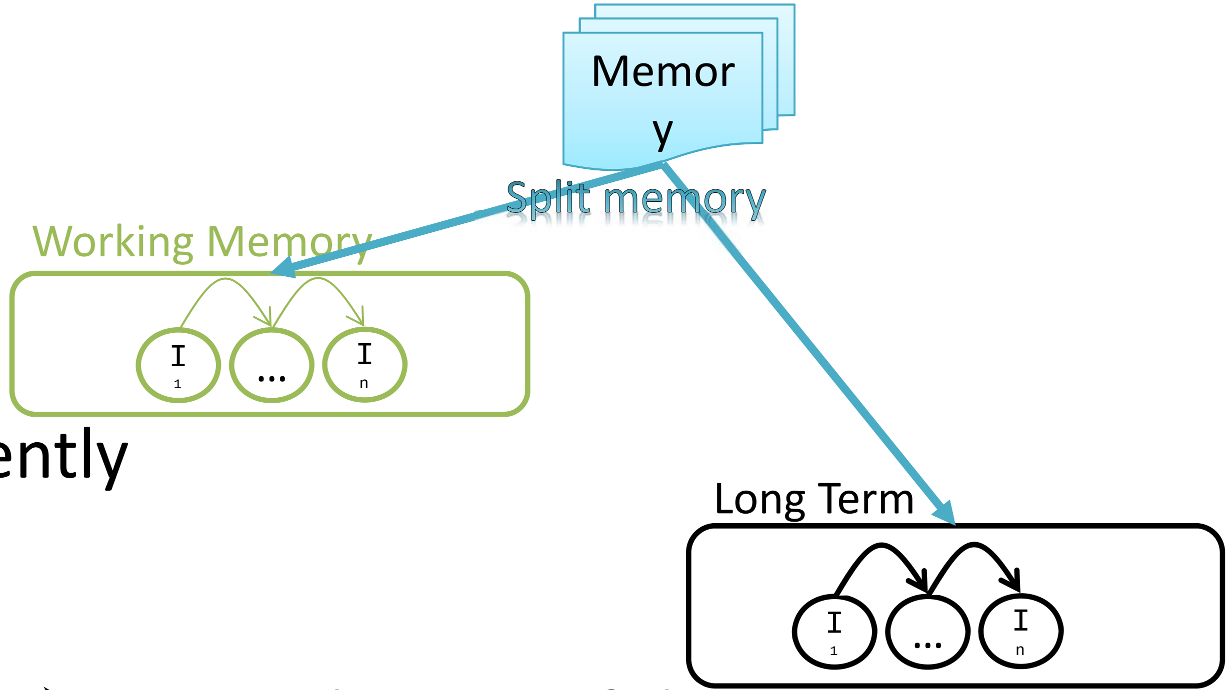
- **Sensor Memory (SM) –** Stores input information from vision sensor measurements.
- **Short Term Memory (STM)** – Stores selected features for visual description
- **Working Memory (WM)** – Keeps the most recent and frequent observed locations
- **Long Term Memory (LTM)** – Stores other observed locations. When a match is found between the current location and one stored in WM, associated locations stored in LTM can be remembered and updated

Memory management: STM



- Fixed size.
- Saved in RAM

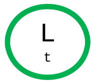
Memory management: WM and LTM

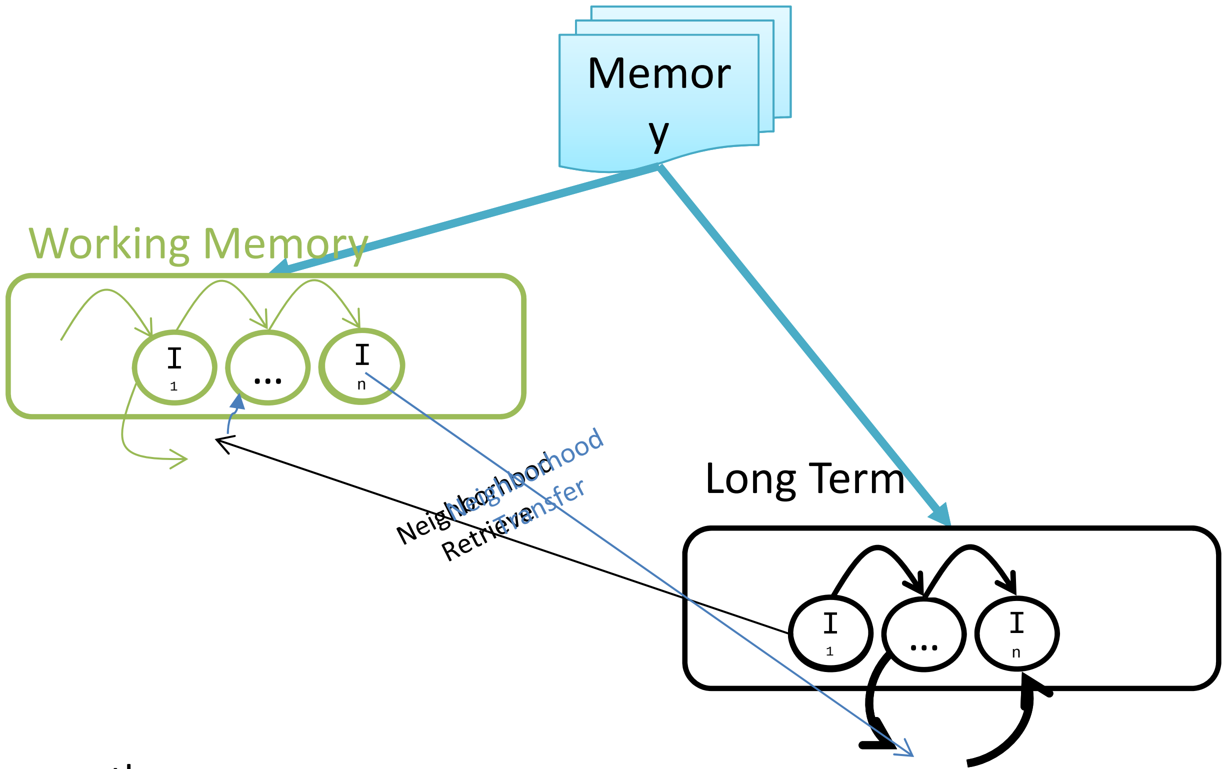


- Recent and frequently visited locations.
- Fixed size.
- Saved in RAM.

- Stores the rest of the map.
- locations in WM linked to neighborhood locations in LTM.
- Size unlimited.
- Saved on HDD.

Memory management: WM and LTM





- Current location L_t.
- Recognize L_t as new location or loop closure ?!!

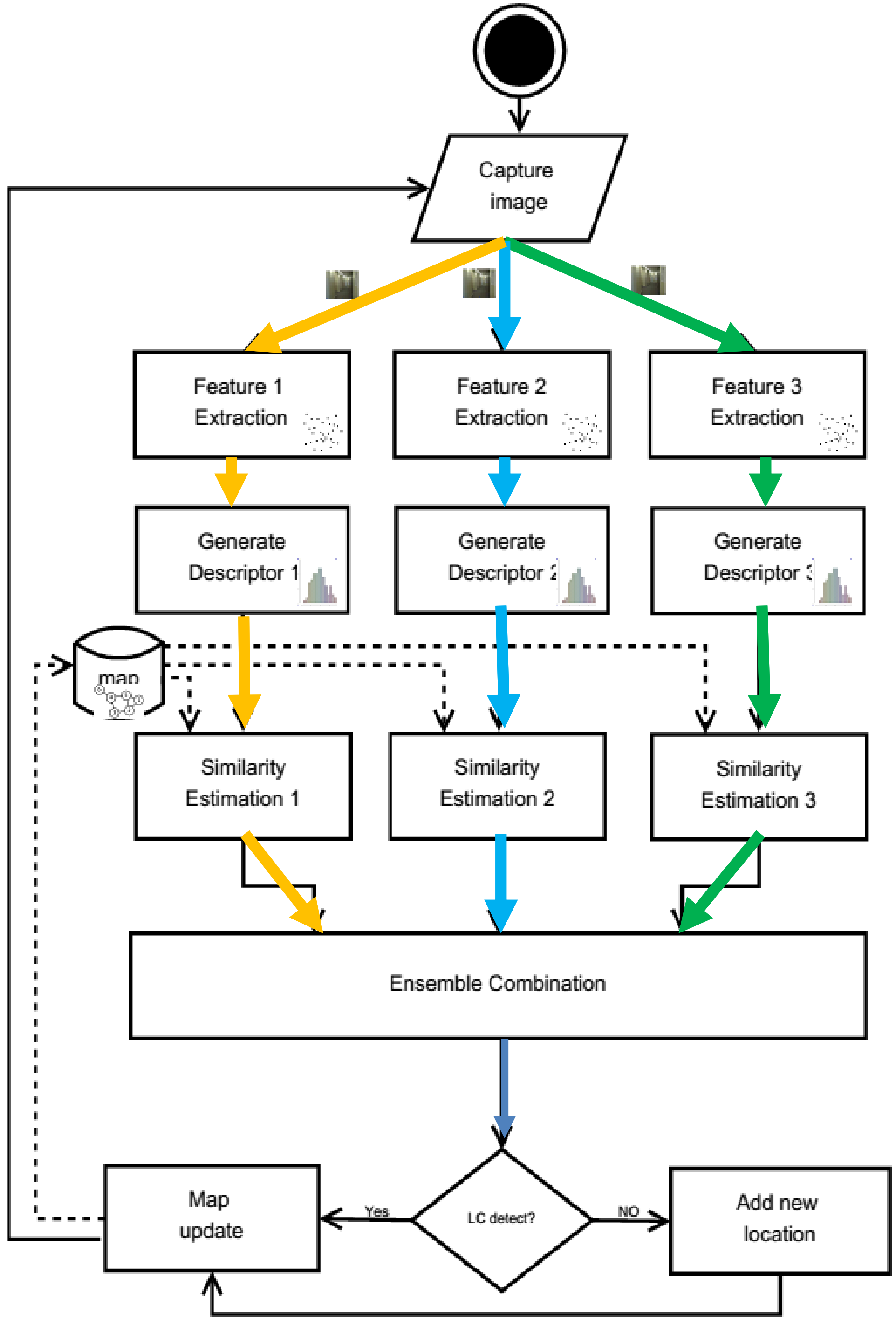
new
- Add L_t. to WM
- Linked L_t with L_{t-1}

Loop closure
- Add L_t. to WM
- Linked L_t with L_{t-1}
- Retrieve L_{t-1}neighborhood from LTM

- If WM size reaches the limit
- Then the oldest location and less frequently visited transfer to LTM.

Algorithm

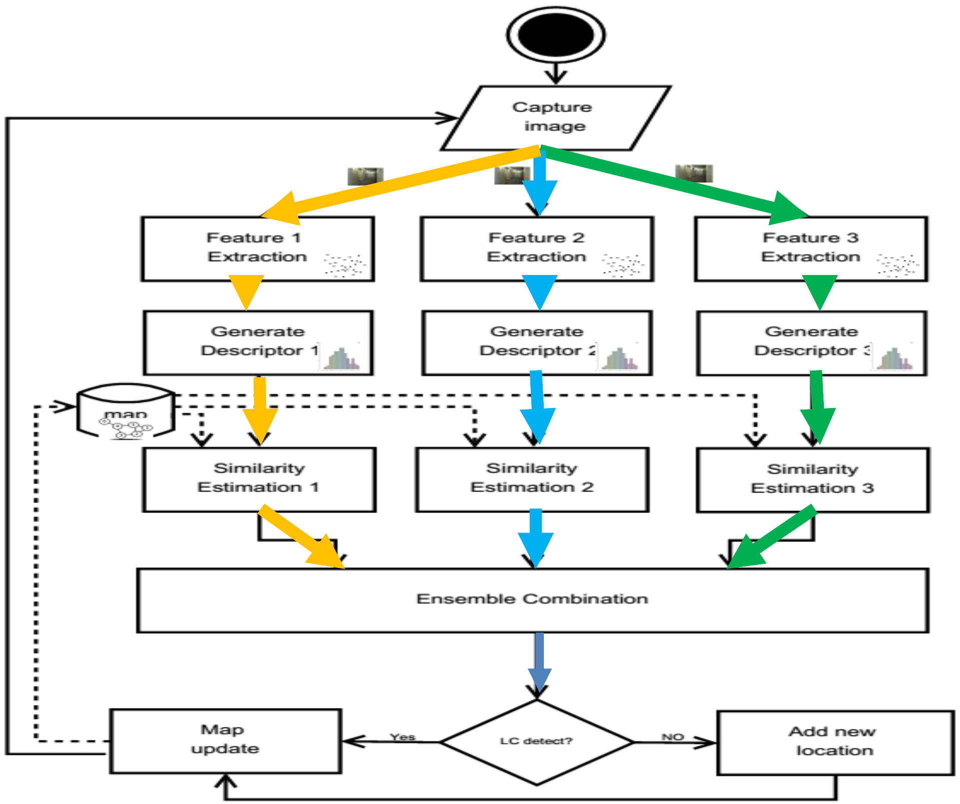
- The following steps describe the algorithm:
  1. A set of visual features “SURF, SIFT and ORB” extract from the same scene.
  2. Construct BOF for each keypoint descriptor.
  3. Each BOF become input to Bayesian filter.
  4. Ensemble learning is used to combine the Bayesian filter outputs and construct a single probability output for loop-closure detection.



Mean Rule Ensemble Learning

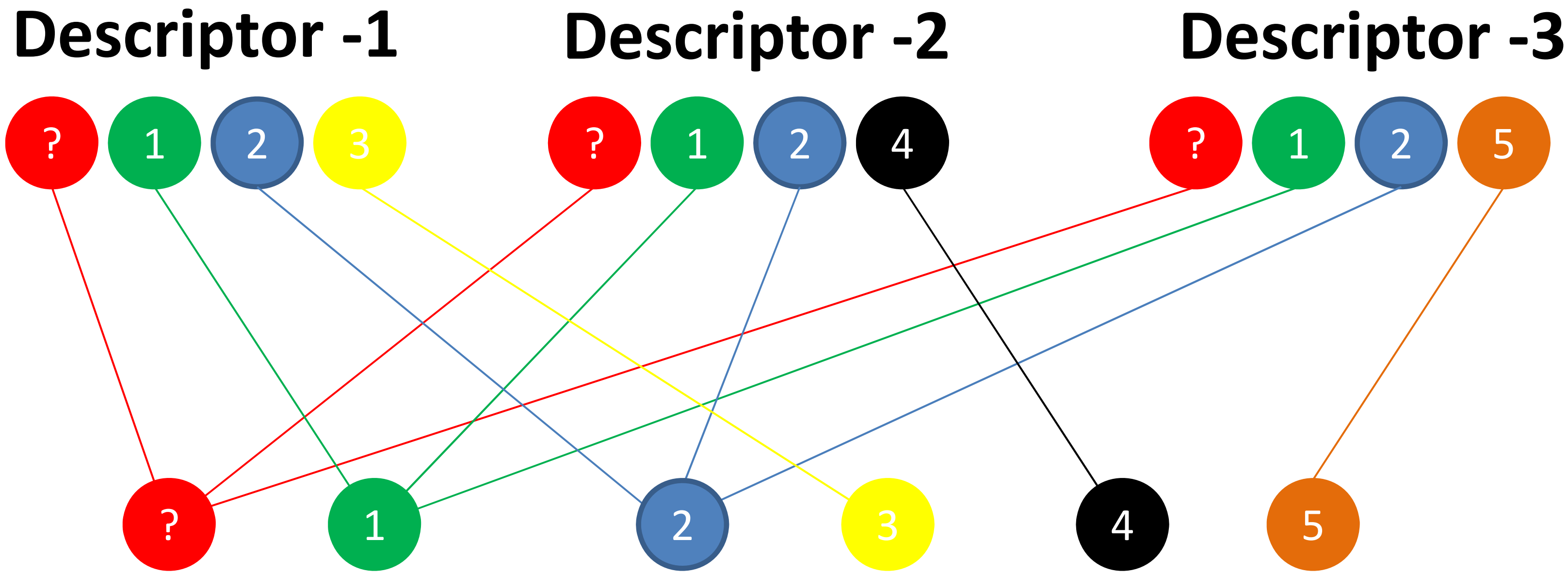
$$P_j^m(x^1, \ldots, x^n) = \frac{1}{n}\sum_{k=1}^{n} P_j^k(x^k) \qquad (1)$$

where x^k is the pattern representation (bag-of-features) of the k^{th} descriptor for location j of n different locations.



Mean Rule Ensemble Learning

- E.g.



Pb (?) > 0.08 = new location/place

Max(Pb(1) ,Pb(2) , Pb(3), Pb(4), Pb (5))

Experiment and Results

Data Sets

City Centre (*CiC*) dataset contains 1237 images of size 1280x480. The images captured outdoor in public roads for *2Km*

Dataset website:
http://www.robots.ox.ac.uk/~mobile/IJRR_2008_Dataset/

Experimental and Results

Precision-Recall

- *Precision* is the ratio of true positive loop closure detections to total detections: $$P = \frac{TP}{TP + FP}$$

- *Recall* is the ratio of true positive loop closure detections to the number of ground truth loop closures: $$R = \frac{TP}{TP + FN}$$

- The aim is to maximize the recall while precision = 100%

Experiment and Results

		Prediction	
		0	1
Actual	0	TN	FP
	1	FN	TP

Experiment and Results

Detection Results

Table 1: *single descriptor*, the recall performance observed at 100% precision

Dataset	CiC
$SURF$	82%
$SIFT$	81.28%
ORB	65.24%

Table 2: The combination descriptor SURF+SIFT+ORB, the recall performance observed at 100% precision

Dataset	CiC
$Ensemble\ Learning$	84.67%

Comparing with other approaches using the same dataset (*CiC*):

- FAB-MAP get 37%[9].
- PIRF-Nav2.0 get 80%[17].
- RTAB-Map get 81%[19].

Conclusion and future work

- The ensemble loop closure detection improves the robustness and the efficiency of a detection system.
- In the future work:
  - Test on more complex datasets
  - Real time navigation on Turtlebot

Thanks