VISUAL EXPLORATION ALGORITHM THAT USES SEMANTIC CLUES

Aravindhan K Krishnan , K Madhava Krishna
Robotics Research Center
IIIT-Hyderabad

Why vision for exploration?

- Data obtained from a camera is far higher that that of a single laser scan
- More data gives more discriminative power for localization
- Very much suited for topological mapping
- Closer to how human beings navigate

Previous Work on Vision Based Exploration

Sim and Little

- Compute depth of all points in stereo image pair
- Build an occupancy grid map
- Find frontiers

What if depth is inaccurate?

Santosh and Supreeth

- Segment image as floor and obstacles
- Identify boundary between floor and obstacles
- Identify gaps as frontiers

What if there are open spaces?

New strategy based on vision to overcome these pitfalls?

Will the use of semantics help in exploration

This works explores the use of semantics in Vision Based Exploration

This work is different from Semantic Mapping by Paul Newman et al

Newman et al focus on building a 3D map and using laser and camera data

After the map is built, semantic labels are added

This work is different from Semantic Mapping by Paul Newman et al

We use semantic information for the exploration process.

Hence form a three level map with, topological, metric and semantic information

Sensors Used

Stereo Camera

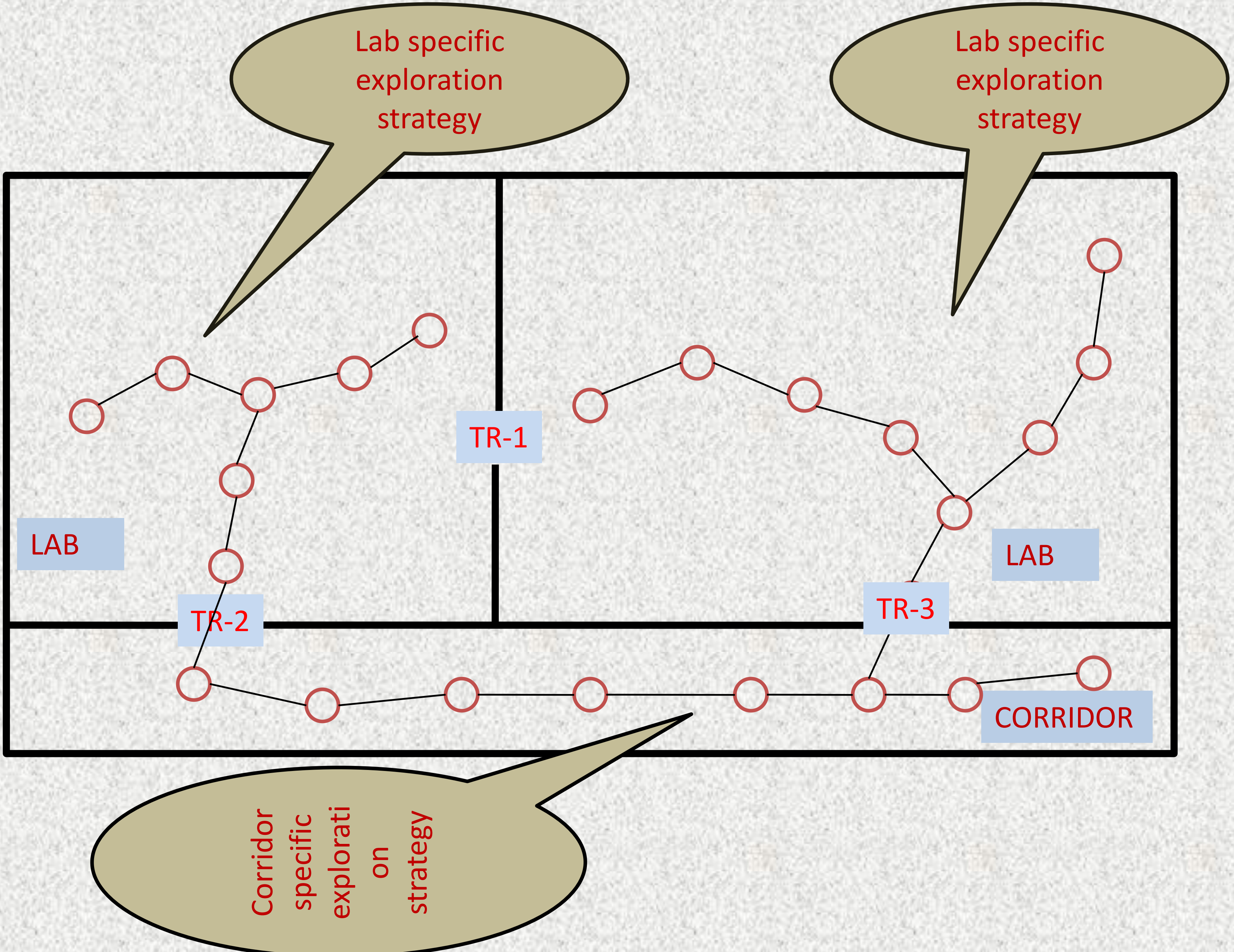
- For computing depth which is a part of our exploration strategy

SICK Laser

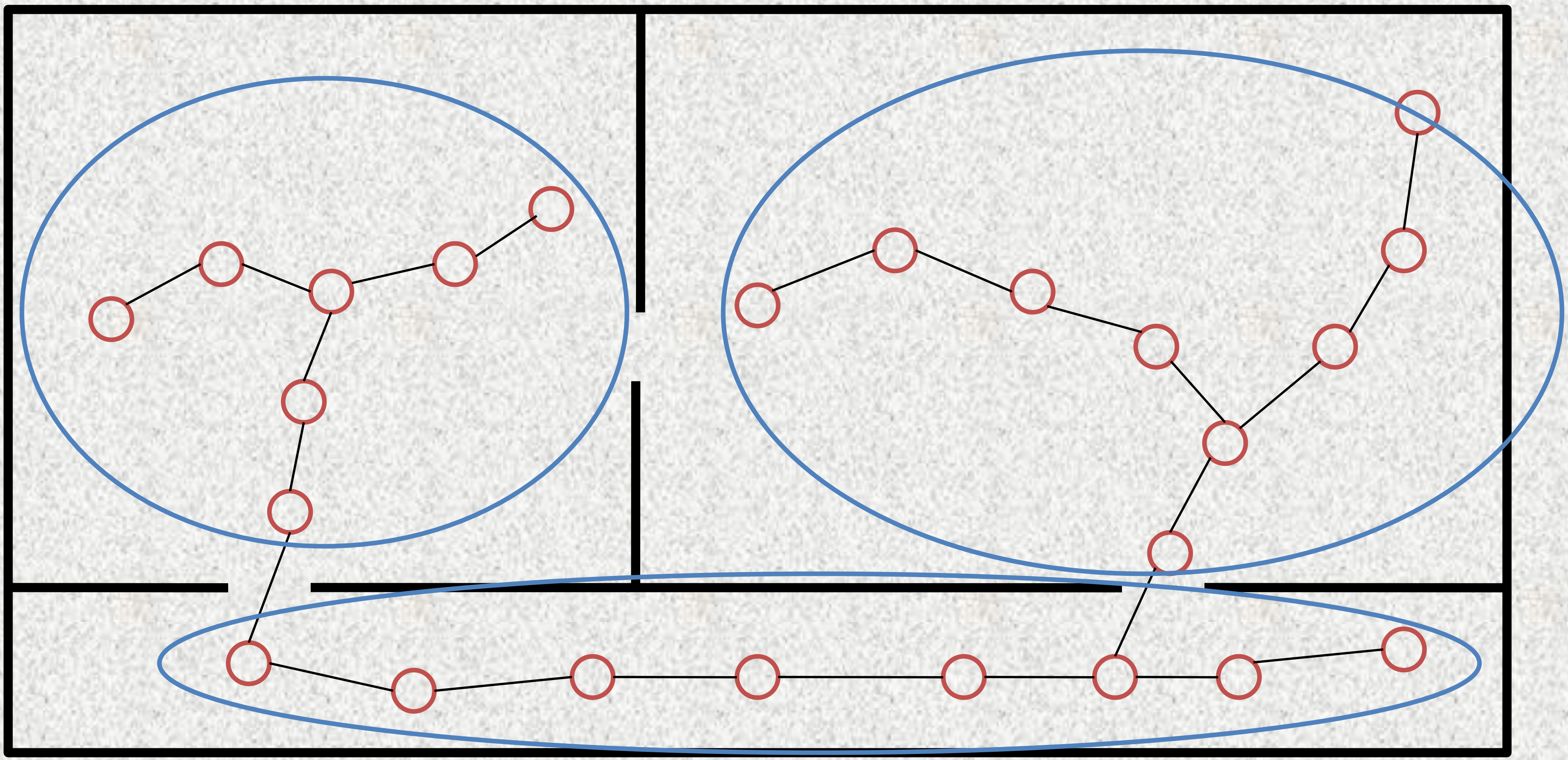
- Used to find doorways (or) Transition Regions
- Used for obstacle avoidance by enabling VFH

Semantic Exploration Outline

- The robot first explores the current semantic construct using a combination of global and local decision making strategies
- In the process, the robot identifies the doorways as transition points to move out of the room
- After exploring a semantic construct, the robot moves through the transition region and starts exploring the new semantic construct
- This process is continued until all transition regions are visited, simultaneously building a hybrid map of three levels
- SVM + Visual Bag Of Words used for semantic labeling
- Laser data used to identify doorways


Lab specific exploration strategy
Lab specific exploration strategy
TR-1
LAB
LAB
TR-2
TR-3
CORRIDOR
Corridor specific explorati on strategy

3 Level Hybrid map



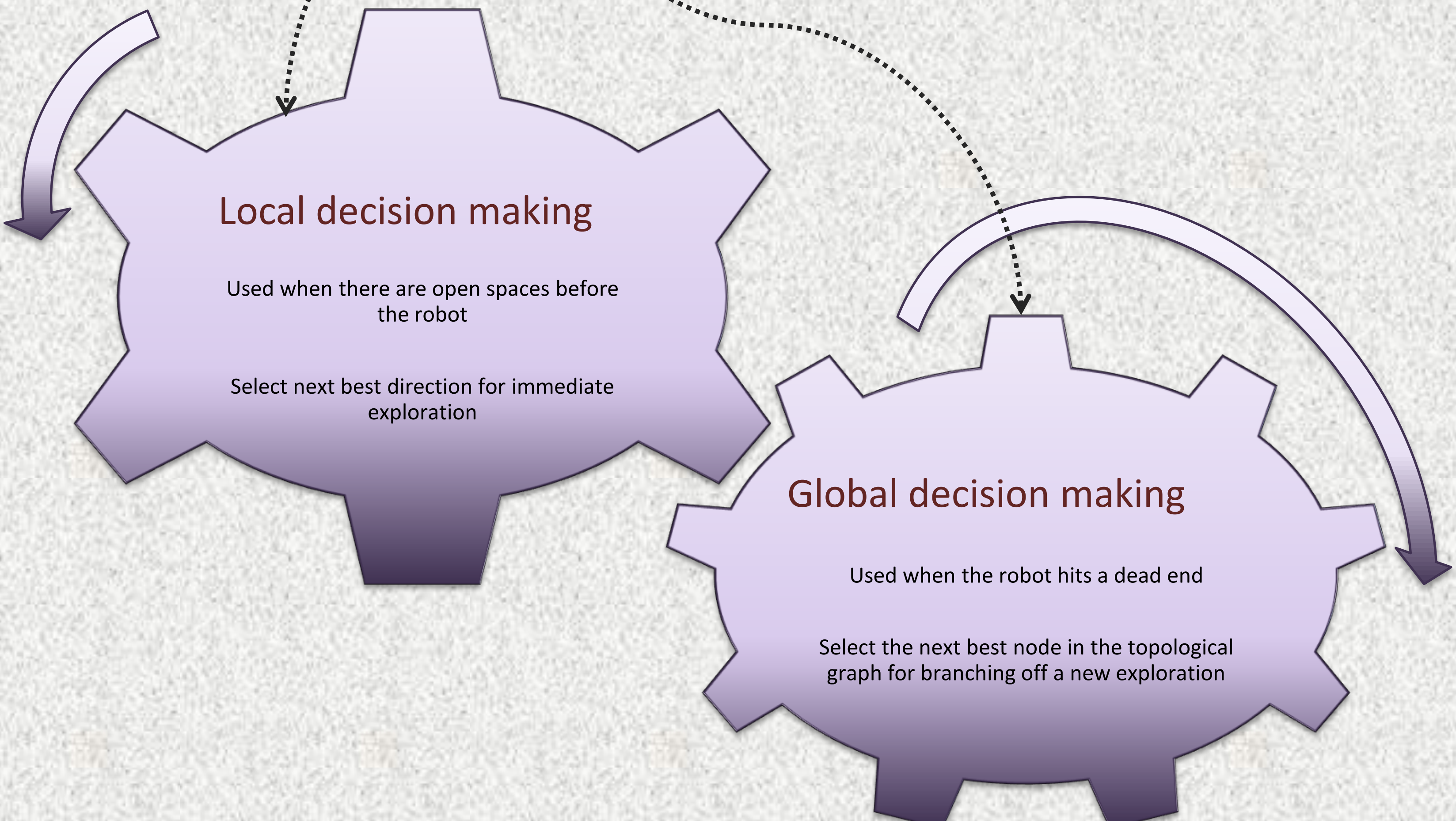
Semantic map

Topological map

Metric map

Our approach : Two part strategy for **exploring a particular semantic construct**

Basic idea : To find the next best direction or the next best node to move at every decision point



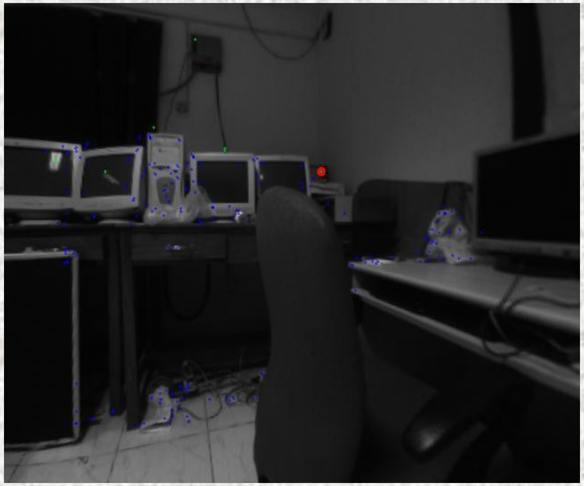
Local decision making

- Aimed at increasing information of neighboring nodes to the current node in the graph
- Information is defined in terms of the feature association across images in different nodes
- Information about a feature is proportional to the number of times it gets associated across different nodes
- To increase information, the robot has to move in a direction, so that it can see the features it has seen already
- The strategy is to move towards the farthest feature
- Assumption - The farthest feature lies in the proximity of other far features
- Hence, moving towards the farthest feature leads to seeing the same far features in the next node, this increasing information in neighboring nodes

Local Decision Making – Far feature based strategy

- Farthest feature in bright red
- Robot orients and moves towards that feature

- Features seen at closer depth (change of color)
- Features getting associated resulting in increase in information



- All features becoming near (blue)
- Dead end
- Global decision making is invoked

Global decision making

- Finds the next best node to branch off a new exploration
- Decision taken globally when
  - There are no far features in the current view
  - Large fraction of features obtained in the current location have been seen multiple times across different nodes
- Weights are assigned to nodes using global feature table
- Global feature table
  - Maintains count of features (*cnt(i)* – count of feature i)
  - Maintains the minimum depth at which the far feature was seen

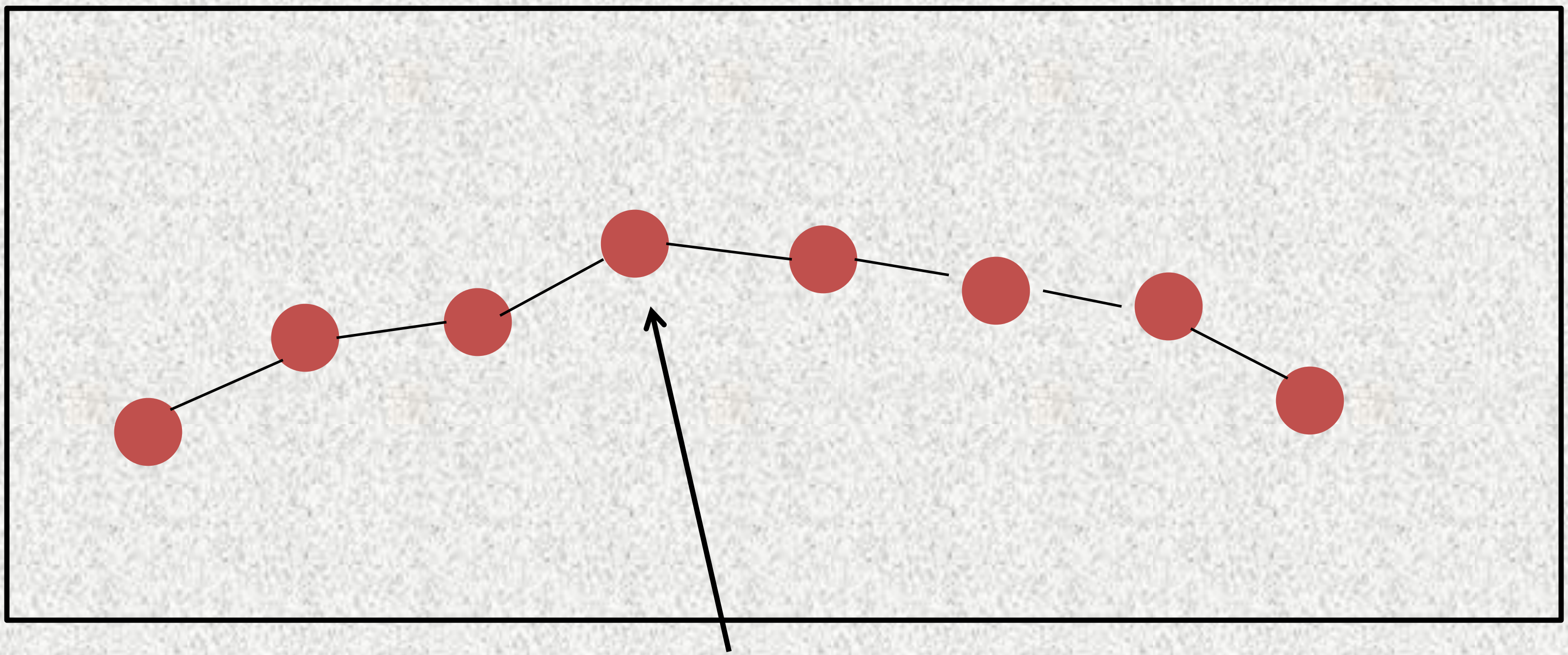
$$weight = \sum_i \alpha * cnt(i), \sum_j \beta * depth_j$$

- Select a node which has the least weight

Shoots up when features are seen repeatedly

Shoots up when features are near, or were seen near already

Global Feature Table

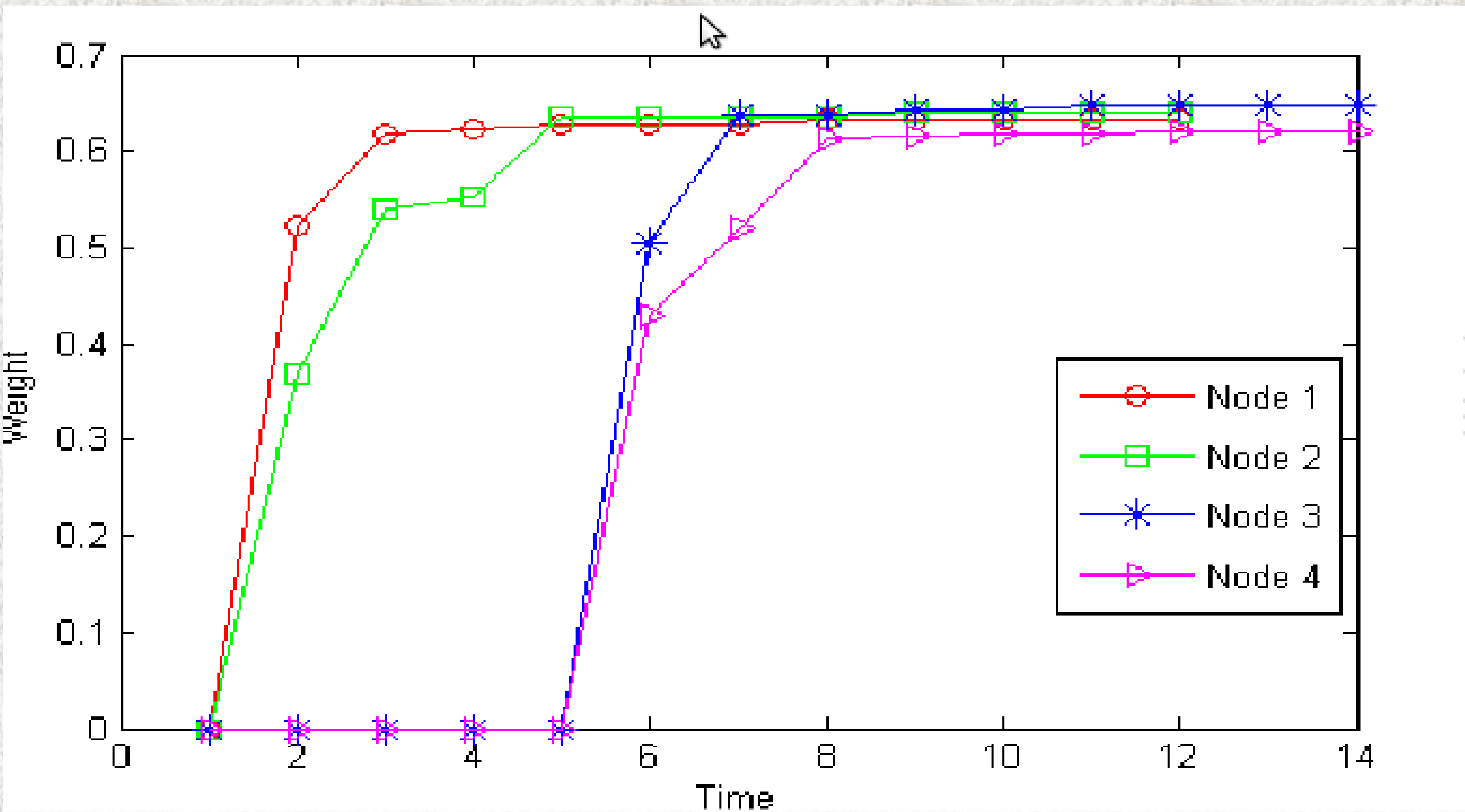


Robot hits a dead end

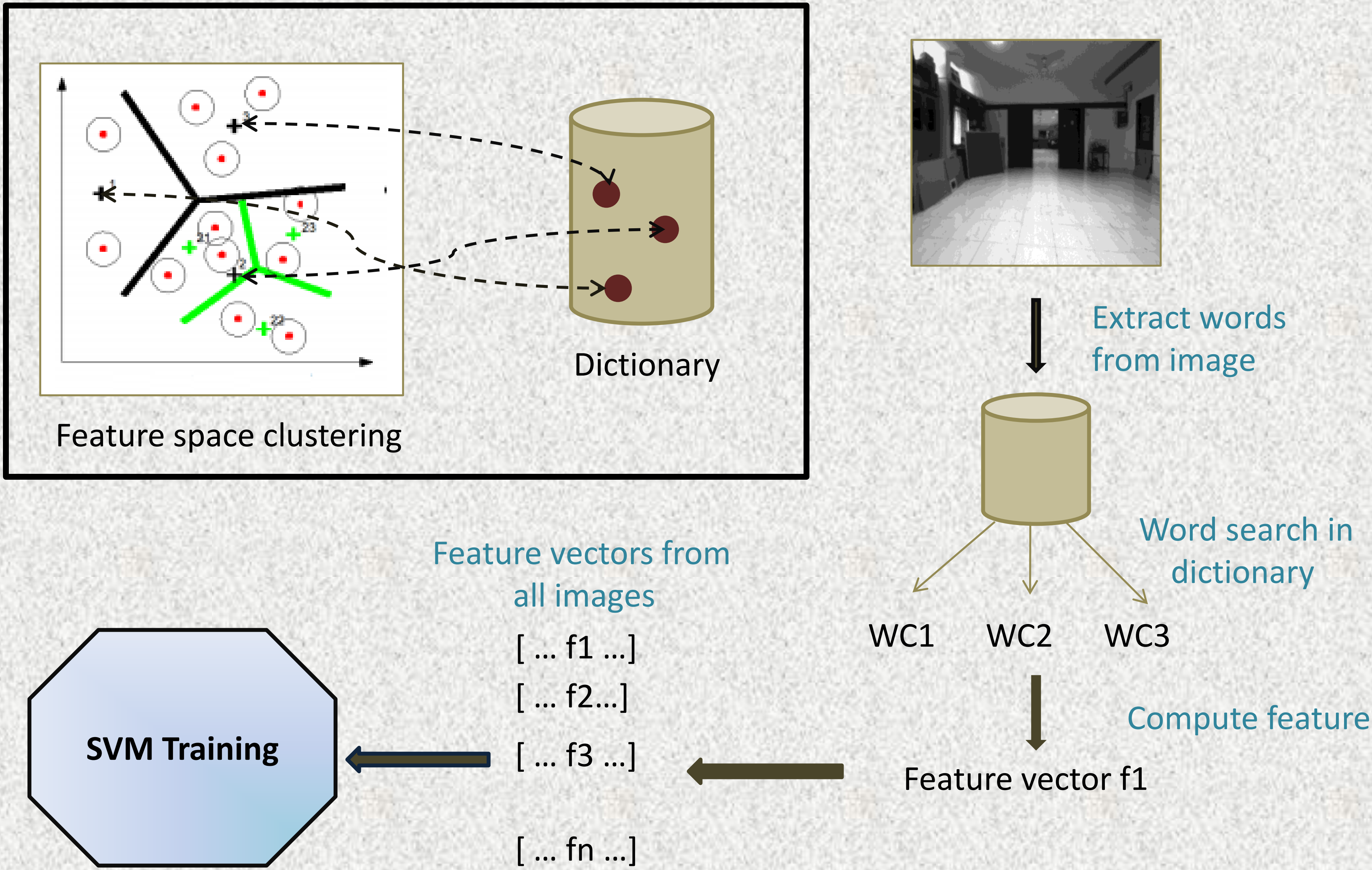
Find the next best node

Branches off new exploration

Weight change graph



VBOW for semantic understanding



Classification results

(a) Pr(LAB)=0.45 (b) Pr(LAB)=0.45 (c) Pr(LAB)=0.46 (d) Pr(TR)=0.44

(g) Pr(CORR)=0.85 (h) Pr(CORR)=0.85 (i) Pr(CORR)=0.85 (j) Pr(LAB)=0.45

(m) Pr(LAB)=0.45 (n) (e) Pr(TR)=0.40 (f) Pr(LAB)=0.45

Detecting Transition regions

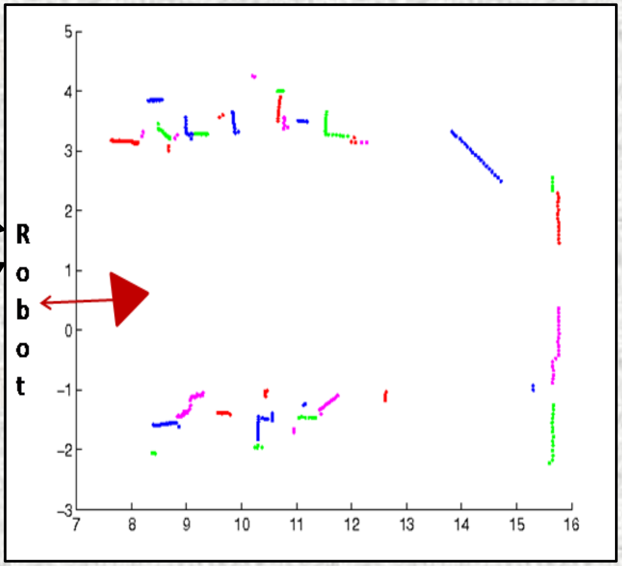
Cluster points in scan and fit line segments

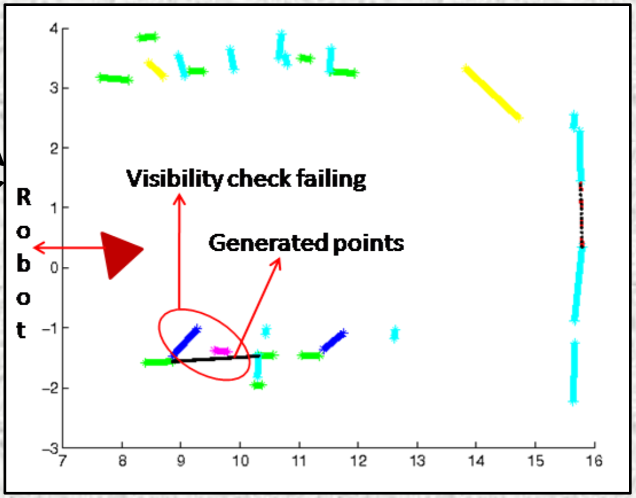
Group line segments based on slope

Connecting line segments between adjacent pair of line segments are considered possible transition regions

Do a visibility check on the "possible transition regions" and filter them

Do a final check using VBOW + SVM framework





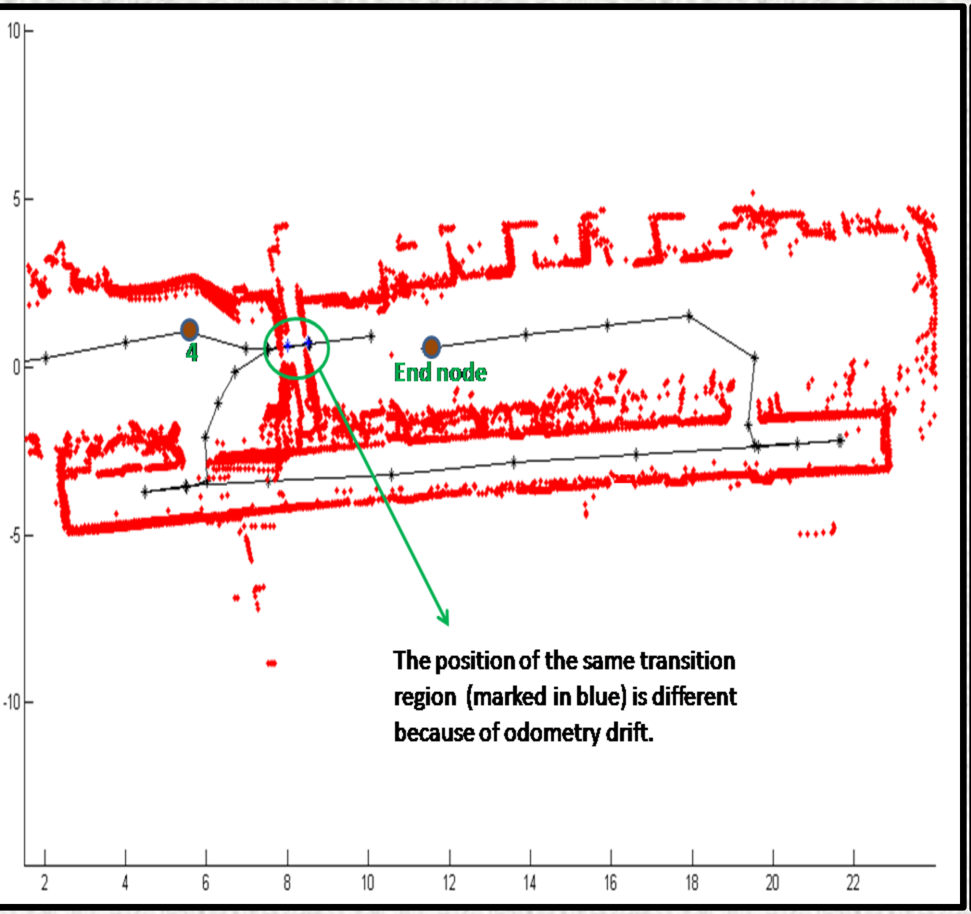
Detecting Transition regions

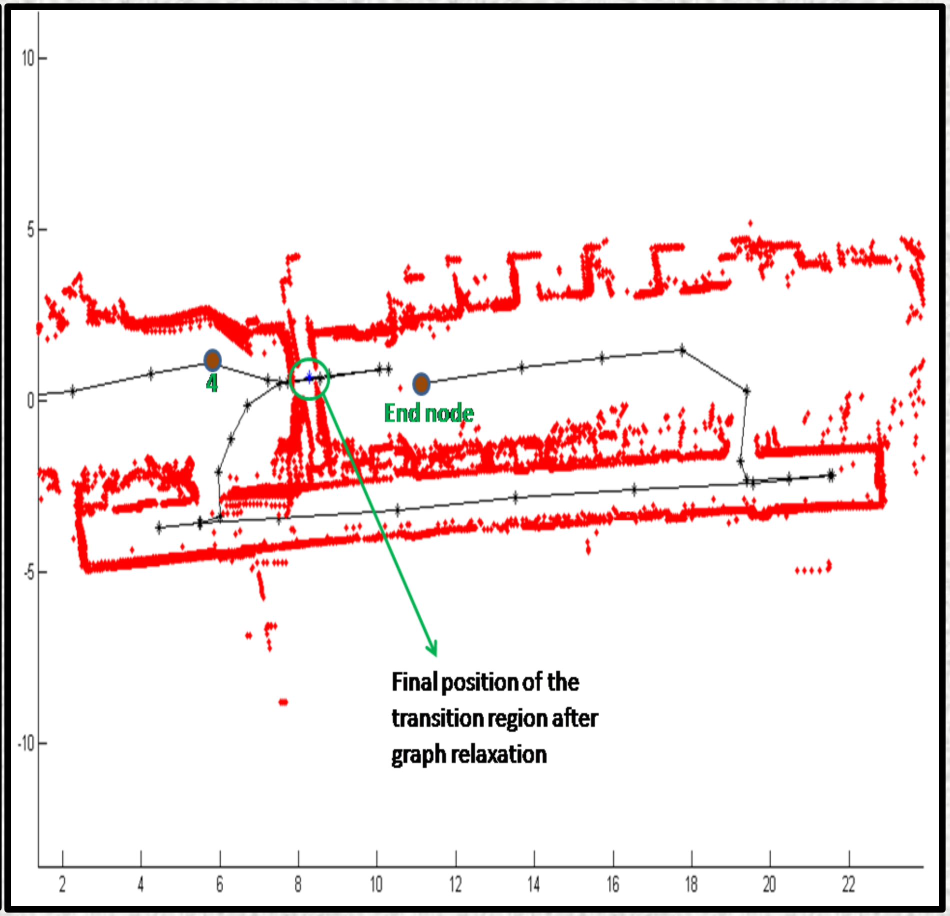
This method of finding doors is not very robust.
It works very well only in geometric environments
Burgard et al have used ML techniques for this

Loop Detection at Transition regions

- Each Transition Region(TR) is denoted by its midpoint.
- Whenever a TR is detected, the uncertainty of the robot position is projected over the TR along with the measurement uncertainty
- When TRs are seen again, they are associated with already seen ones using nis distance. If the nis distance is within a range then visual clues are used to check if they are the same
- Loop is thus detected Graph relaxation is run to close the loop

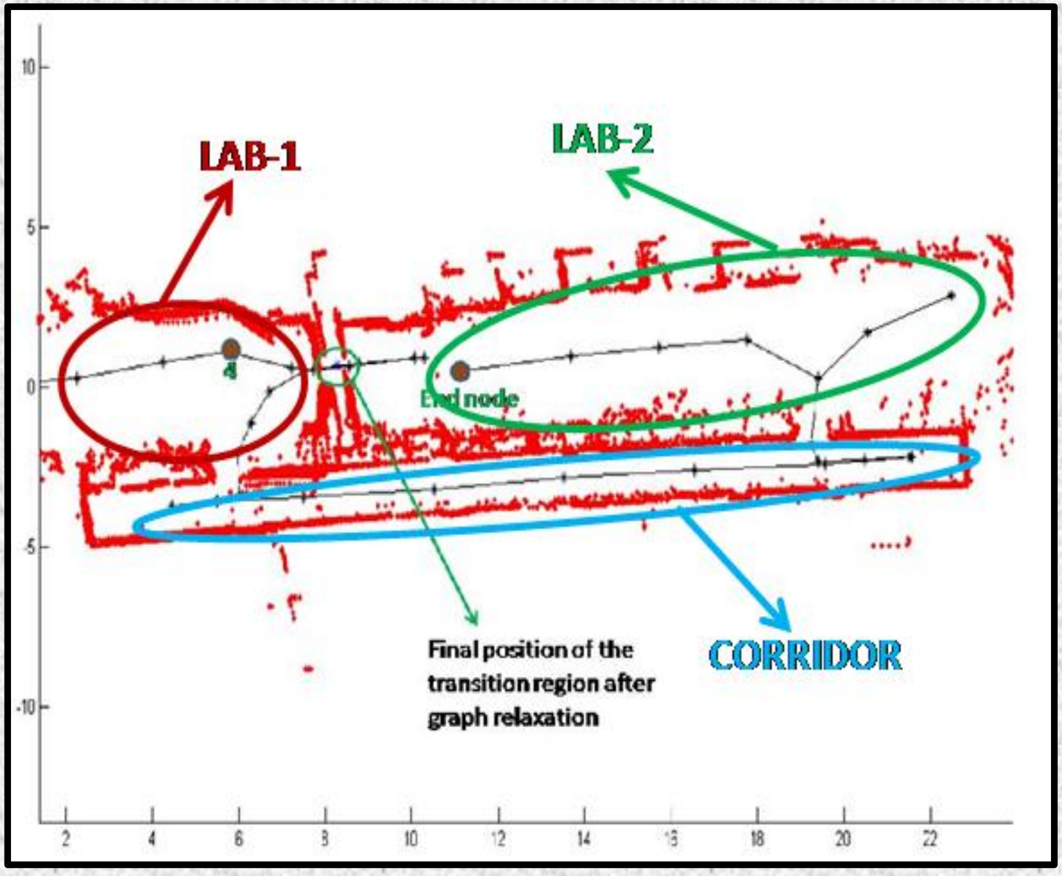
Loop closure

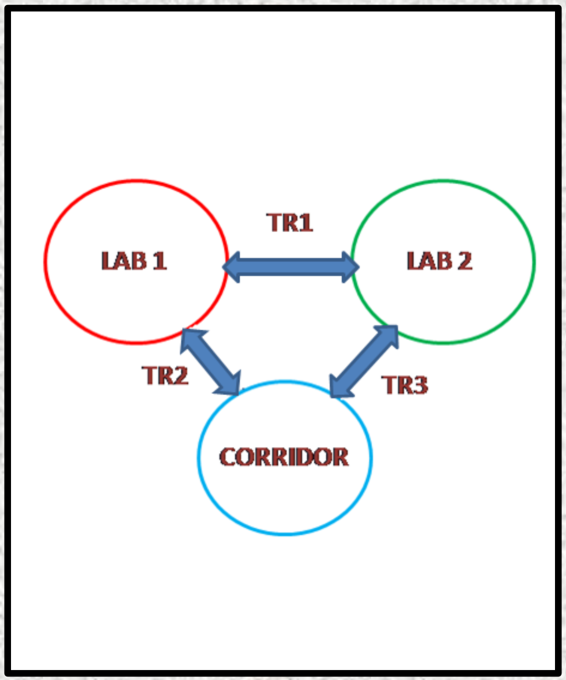




Semantic map constructed

- Once the robot explores a particular semantic construct and moves out of a transition region, all nodes in the semantic construct are labeled with the same class label

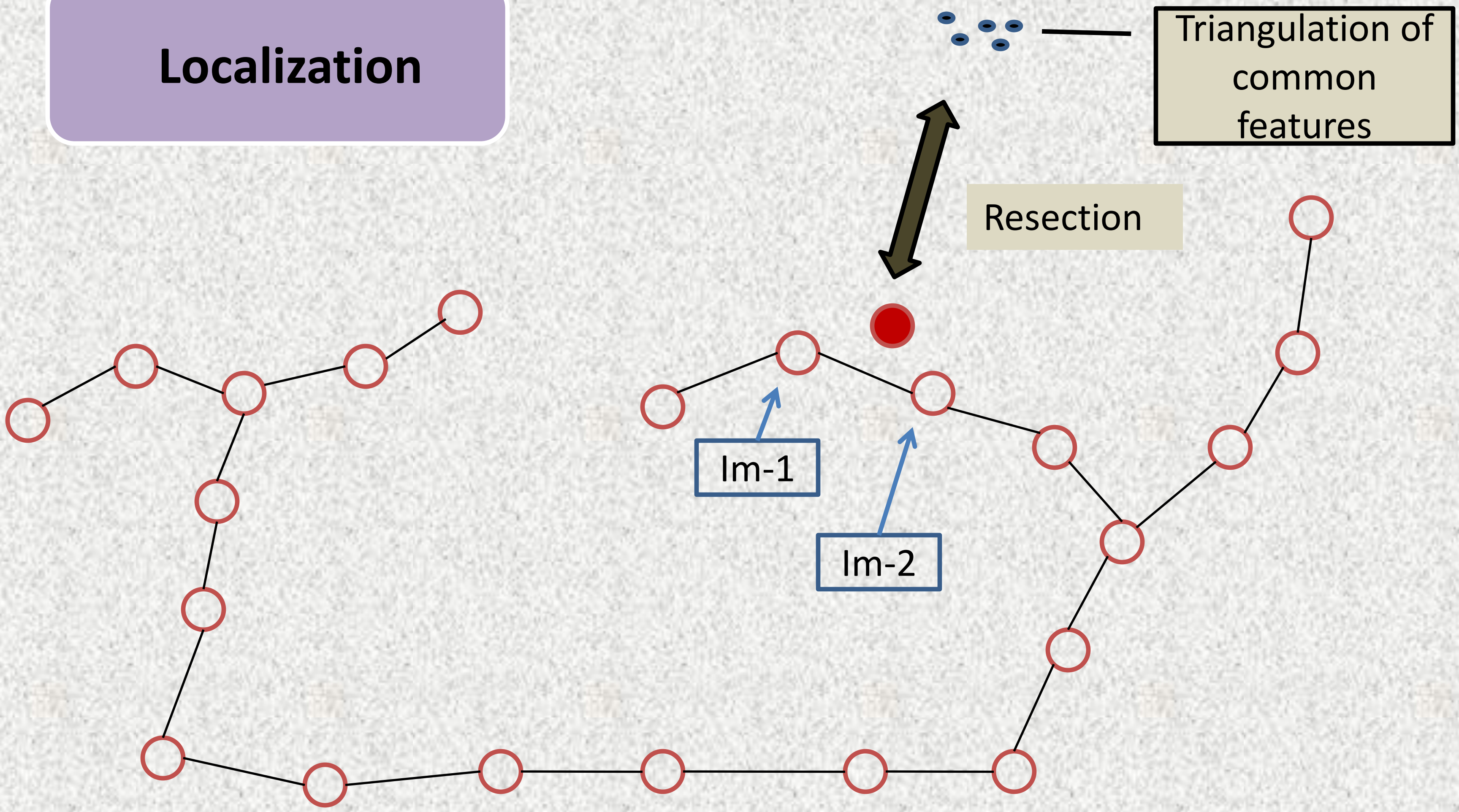




Localization

- The closest image(Im1) to the current image(Im3) is retrieved based on feature matches
- The image from the adjacent node(Im2) is also retrieved
- Matching points are triangulated
- The robot is localized with respect to Im1 by estimating pose by PnP with the triangulated points and the matching matching points in Im3

Localization



Advantages of the method

- Semantic exploration more closer to the way human beings explore
- Exploring a semantic construct completely ensures that the pose error is bounded and contained for a particular semantic construct
- Loop detection at transition regions is faster (No need to compare the currently acquired image to all the images acquired previously)
- Moreover, we could also have a semantic construct specific exploration strategy

Future Work

- Include corridor intersection as loop closure agents and do loop closure for a larger environment
- Try out other graph relaxation methods