6-DoF Semantics-Aware Condition- and Viewpoint-Invariant Visual SLAM

Confirmation of Candidature

Sourav Garg

Principal Supervisor : Michael Milford

Associate Supervisor: Niko Sunderhauf

Contents

- Introduction
- Literature Review
- Research Questions
- Research Plan
- Work Progress
- Conclusion

Introduction

Social

Domestic

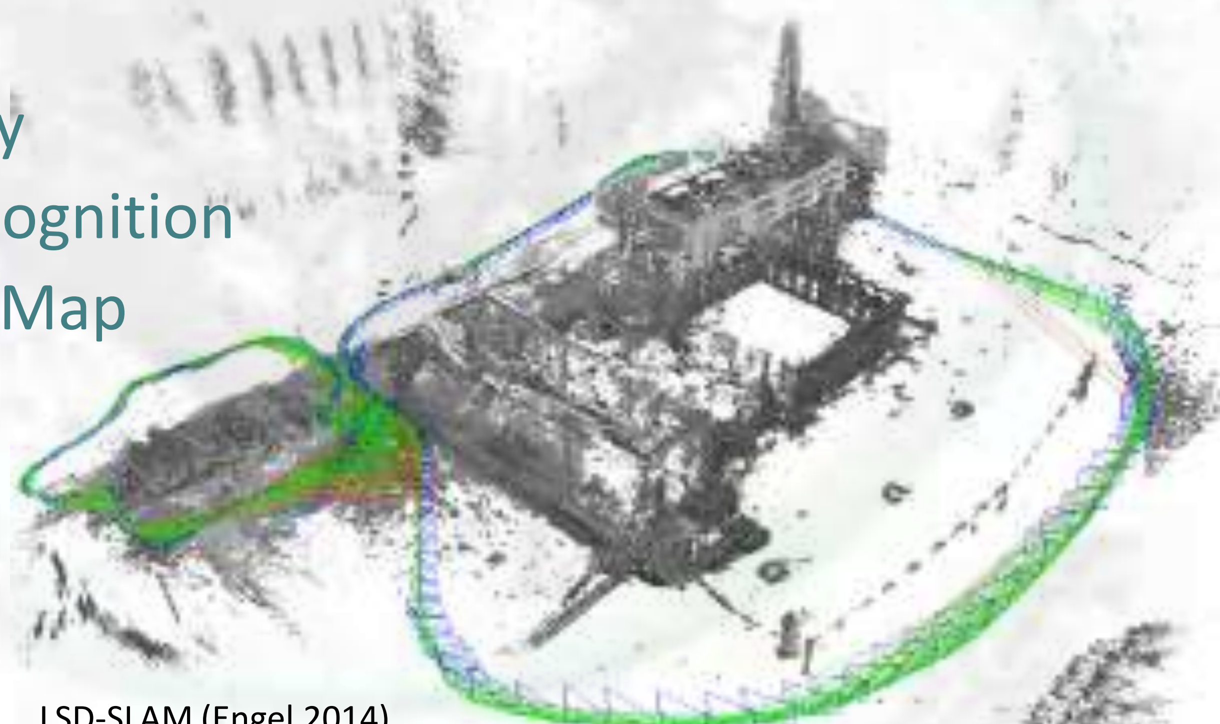
Industrial

SLAM: Is it solved?

- Simultaneous Localization and Mapping
- Solved (Thrun and Niera, 2010):
  - For Static Environments
  - Using Laser Rangefinder
- Unsolved:
  - Vision Only
  - Extreme Appearance Change
  - Dynamic Objects
  - Semantic Understanding

Visual SLAM

- Use only Visual cues
  - Monocular or Stereo Vision using Digital Cameras
- Components:
  - Visual Odometry
  - Visual Place Recognition
  - Representation Map

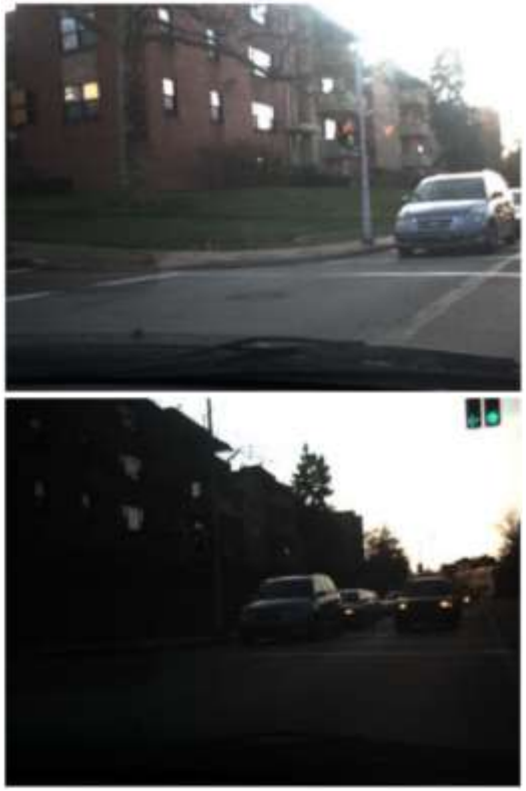


LSD-SLAM (Engel 2014)

Visual Place Recognition

- Image Representation
  - Local Features – SIFT, SURF, BRIEF, ORB etc.
    - Bag of Words in FAB-MAP, ORB-SLAM etc.
  - Global Representation – ConvNet, Bit-Planes, HoG etc.
    - Direct Matching in LSD-SLAM, SeqSLAM etc.
- Challenges
  - Perceptual Aliasing
    - Different places appear similar
  - Conditions Variations (Weather, Season, Time of day)
    - Same place appears different
  - Viewpoint Variations

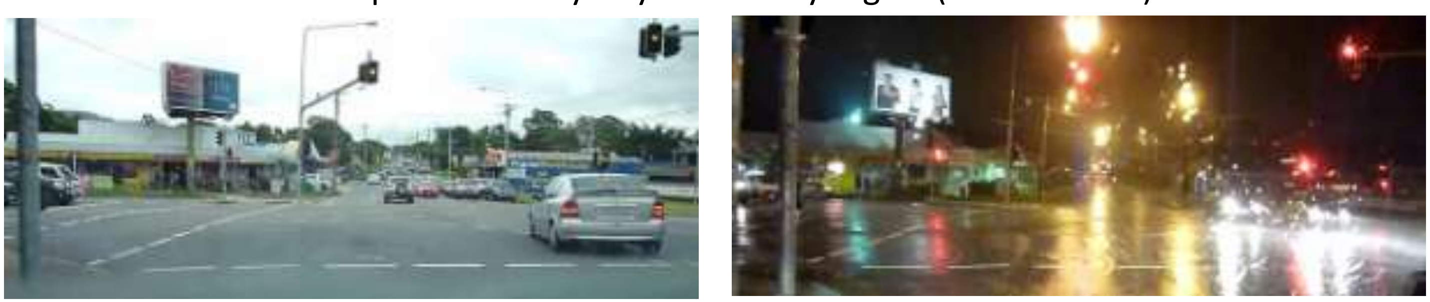
Environmental Conditions



Non-linear Intensity Deformation by Automatic Camera Settings (Alismail 2017)

Seasonal Variations in Nordland (Sunderhauf 2013)

SeqSLAM - Sunny Days vs Stormy Nights (Milford 2012)



Perceptual Aliasing

Different Places with Similar Appearance (Cummins 2007)

Place Recognition – Review

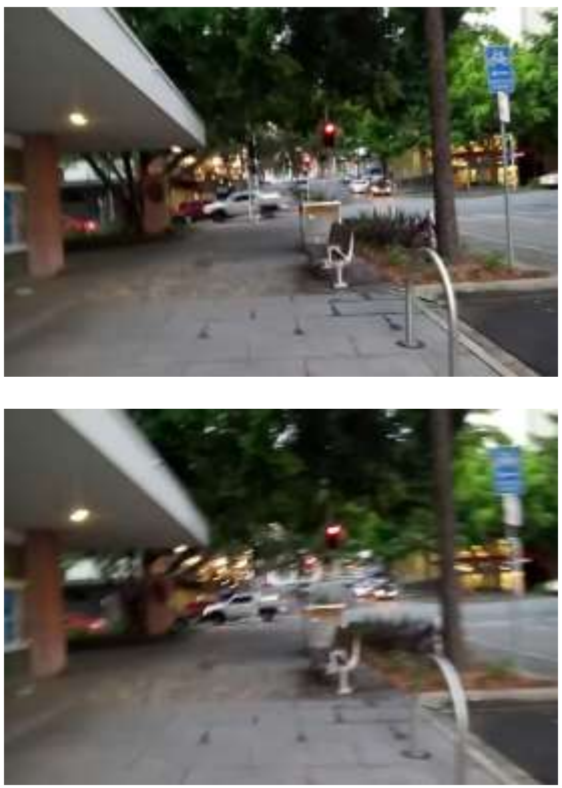
- Local Image Features
  - Robust to viewpoint changes
  - Prone to appearance variations
- Global Representations
  - Robust to appearance variations
  - Prone to viewpoint variations
- Deep-learned features
  - Robust in both cases
  - Lack relative 3D pose estimation

Visual Odometry

- Matching consecutive images for Egomotion estimation
- Methods
  - Local Feature Tracking
    - FAST, SIFT, ORB etc. based feature correspondence
  - Direct Whole Image Matching
  - Deep VO
    - PoseNet, Deep Tracking, Sfm-Net etc.
- Challenges
  - Local Appearance variations
  - Low light environment
  - Rapid Camera Motion
  - Motion Blur
- VO Failure is catastrophic for Visual SLAM

Some Challenges in Visual Odometry

Motion Blur



Low Light Environment

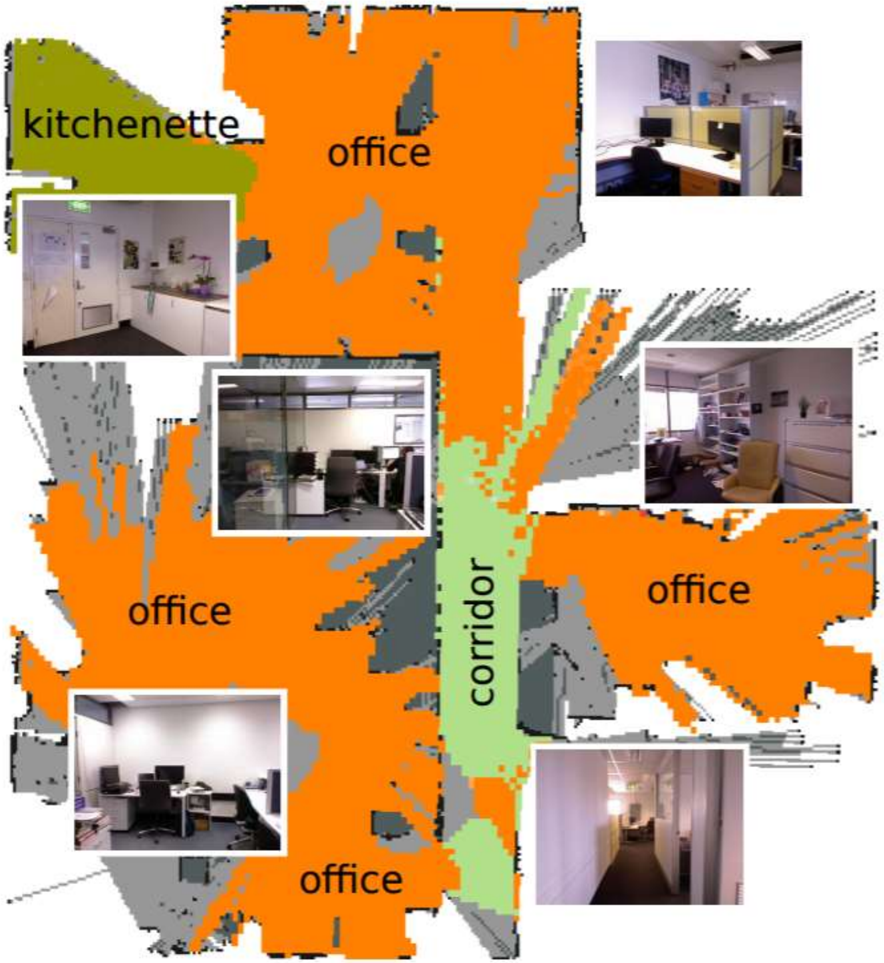
Illumination Variation

(Alismail 2016)

Representation Map

- Types of Maps
  - Topological, Metric or Topometric
- 3D Reconstruction of Environment
  - 6-DoF Visual SLAM systems
  - Sparse, Semi-Dense or Dense Reconstruction
- Semantics in Maps
  - Meaningful interaction with environment
- Aim: Metric Map with Semantic Topology

Probabilistic Semi-Dense Mapping
(Mur-Artal 2015)

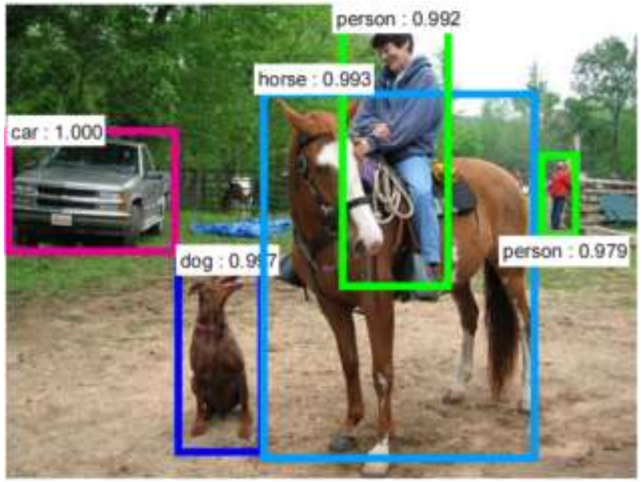


Place Categorization and Semantic Mapping
(Sunderhauf 2016)

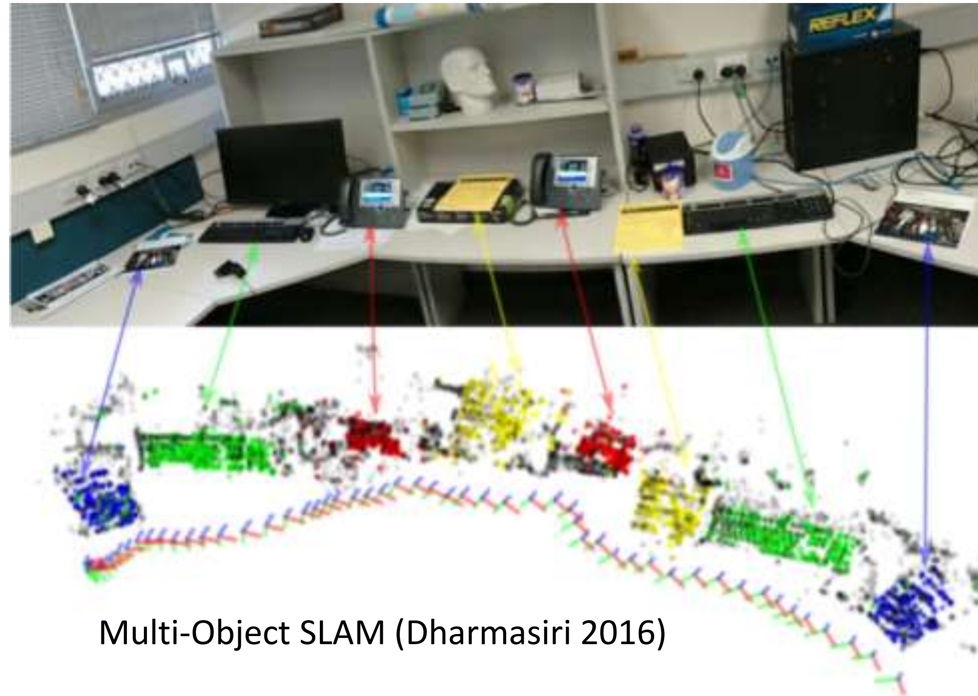
Semantics in Visual SLAM

- Semantics within an Image
  - Object-Oriented
    - Deep-learned Object Recognition
    - Sparse or Dense Object Segmentation
  - Semantic Structures like wall, ceiling, floor etc.

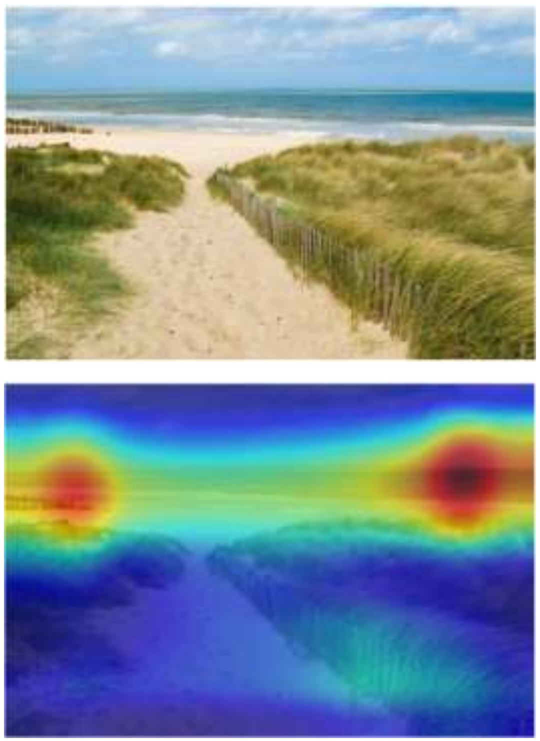
- Place Semantics
  - Place-centric vs Object-centric Approach
  - Place Categories and Attributes
    - Structural – Train Station, Kitchen, Bedroom etc.
    - Transient – Night, Rain, Snow etc.



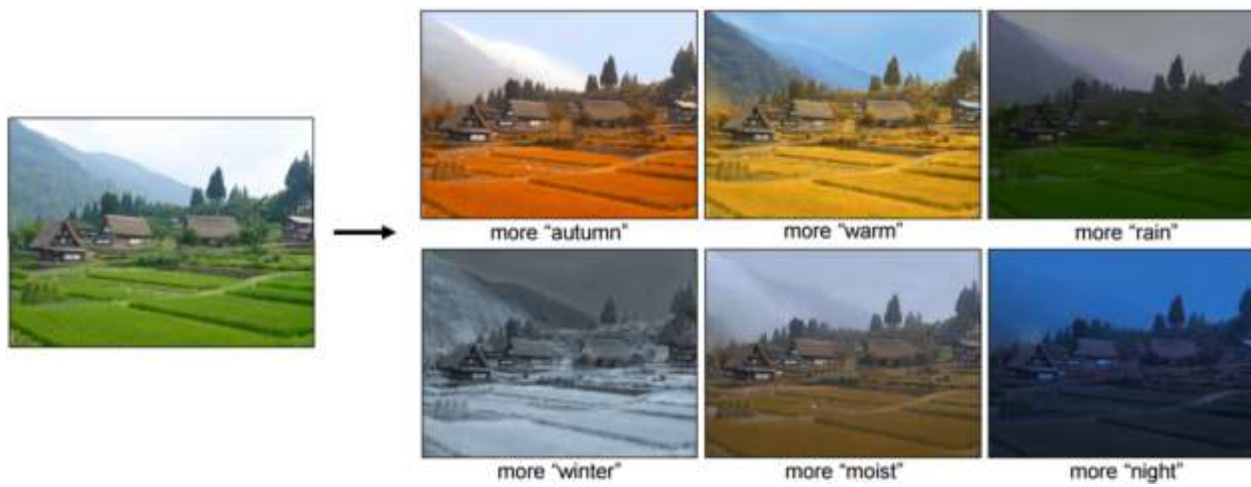
Object Detection Using Faster R-CNN
(Ren 2015)



Multi-Object SLAM (Dharmasiri 2016)



Scene Recognition Using Places
Database (Zhou 2014)



Transient Attributes for Outdoor Scene Editing
(Laffront 2014)

Research Gap

- Place Recognition
  - Viewpoint- and Condition-Invariance
  - With 3D Relative Pose Estimation
- Visual Odometry
  - Robustness to Environmental Conditions
  - Hybrid Approach for Rapid Camera Motion
- Representation Map
  - Semantic Maps on top of Geometric Maps
- Semantics in Visual SLAM
  - An integrated system where both can benefit

Problem Statement

How can we develop a general purpose 6-DoF semantics-aware condition- and viewpoint-invariant visual SLAM system?

Research Questions

- How can we characterize a visual SLAM system against variations in environment and egomotion, and make it more robust for both visual place recognition and visual odometry?

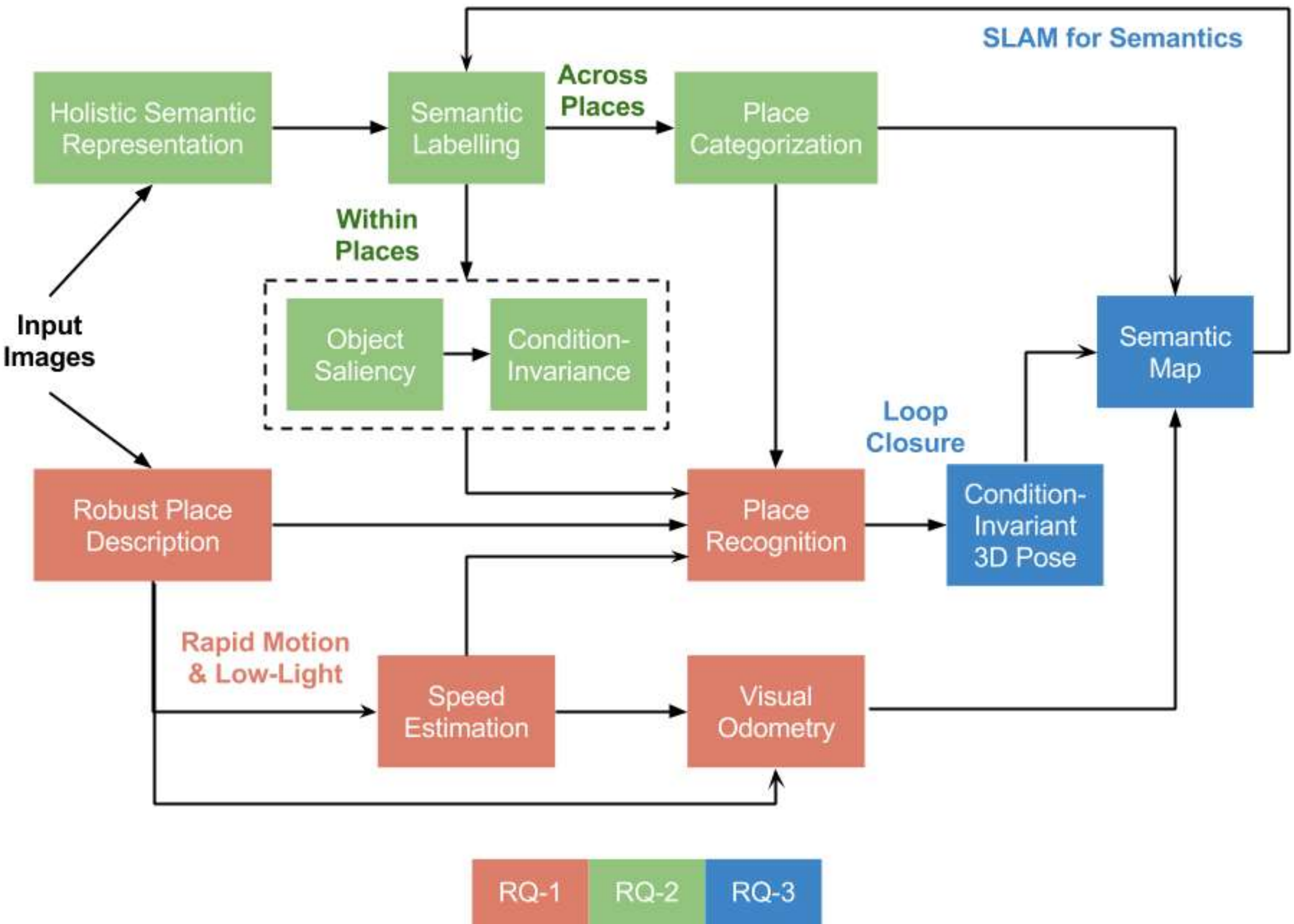
- How can semantic information related to images add value to a visual SLAM system?

- Can we develop a general purpose 6-DoF semantics-aware SLAM system robust to condition and viewpoint variations.

Research Plan

- Characterizing and Adapting Visual SLAM system
  - Visual Odometry and Place Recognition
  - Adapting to Environment
  - Egomotion Estimation
- Utilizing Semantic Information
  - Holistic Semantic Representation
  - Semantics Across Places
  - Semantics Within Places
- 6-DoF Semantics-Aware Visual SLAM
  - Condition-Invariant 3D Relative Pose Estimation
  - Semantics for SLAM
  - SLAM for Semantics

Research Plan



Characterizing and Adapting Visual SLAM

- Characterizing Viewpoint and Condition Invariance
  - Place Recognition – SeqSLAM
  - Visual Odometry – ORB-SLAM
- Handling Variations within the Environment
  - Transitions in Environment
    - Indoor/Outdoor, Naturally-Lit/Artificially-Lit etc.
  - Environment Segmentation
    - Improve Place Recognition
- Egomotion Estimation in Low Light and Rapid Motion
  - Speed Estimation for Visual Odometry
  - Speed-Normalized Data Sampling for Place Recognition

Utilizing Semantic Information

Holistic Semantic Representation

- *Place-Centric* Semantics (Zhou 2014)
  - Environment Type
    - Outdoor
  - Categories
    - Harbor -0.44
    - Dock – 0.33
    - Boat Deck – 0.10
  - Attributes
    - Open Area
    - Natural Light
    - Far-Away Horizon
    - Man-Made

- *Object-Centric* Semantics (Jia 2014)
  - Vessel – 1.26
  - Craft – 1.17
  - Vehicle – 0.80
  - Ship – 0.76

- Transient Attributes (Laffront 2014)
  - Time of Day
    - Sunrise/Sunset – 0.59
  - Weather
    - Sunny/Direct Sun – 0.39
    - Clouds/Overcast – 0.27
  - Season
    - Summer – 0.34
    - Winter – 0.07

Semantic Segmentation *Across* Places

- Semantic Map
  - Segment Environment
    - Office Space - Cubicles, Canteen, Corridor, Restroom etc.
    - Vehicle on Roads - Urban Canyon, Highway, Tunnel etc.
- Place Categorization and Place Recognition
  - Reduce Search Space
    - Coarse Semantic Localization
  - Seamless Transitions Within Environment
    - Indoor/Outdoor, Naturally/Artificially Lit, Bland/Cluttered

Indoor-Outdoor Seamless Localization

Semantic Segmentation *Within* Places

- Object Semantics
  - Deep CNN Object Recognition
  - Learn New Object Classes Online
- Place Semantics
  - Transient Environmental Conditions
  - Condition-Invariant Place Recognition
- Object/Patch Saliency *Within* Place
  - Dynamic Objects
  - Condition-Invariance

Object Oriented Semantic Mapping (Sunderhauf 2016)

6-DoF Semantics-Aware Visual SLAM

- Condition-Invariant 3D Relative Pose
  - Condition-Invariant Place Description
    - Exploring Bit-Planes Descriptor
- Semantics for SLAM
  - Improved Localization
  - Meaningful Maps
  - Scale to 3D
- SLAM for Semantics
  - Improve Semantic Labelling

An example of the LBP descriptor evaluated on a 3 × 3 neighborhood, which results in a 8−channel bit-plane descriptor. Unlike the classic LBP descriptor, the bit-plane descriptor can be employed within a multi-channel LK framework using a sum of squared differences (SSD) cost measure. (Alismail 2016)

Work Progress

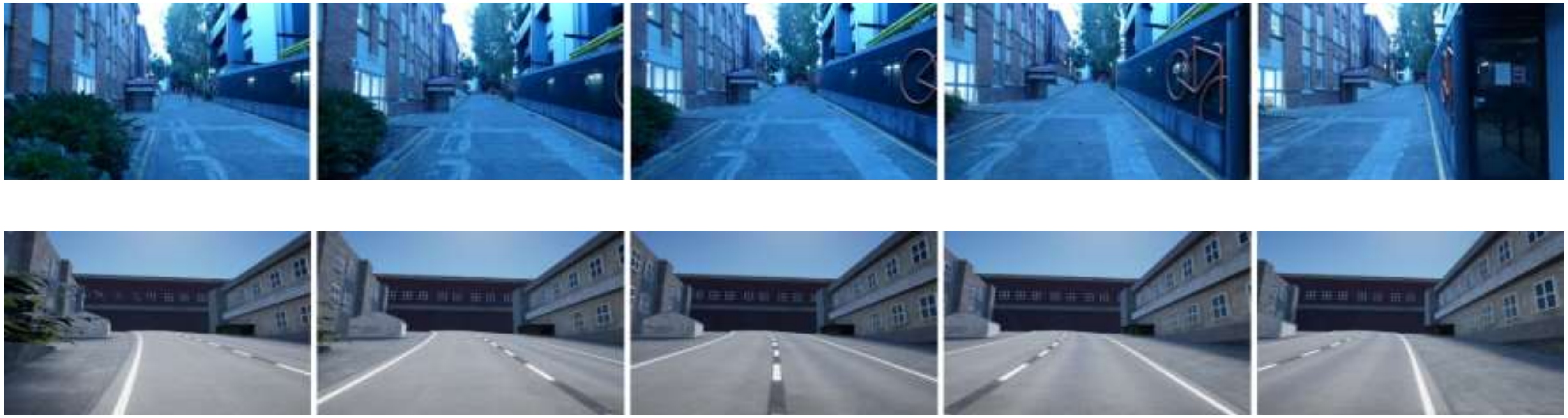
- Performance Evaluation using High-Fidelity Simulation
  - Place Recognition – SeqSLAM
  - Visual Odometry – ORB-SLAM

- Semantic Segmentation of Environment
  - Transitions in Environment
  - Improve Condition-Invariant Place Recognition

- Egomotion Estimation in Unfavourable Conditions
  - *Hybrid* Visual Odometry
  - Speed-Normalized Data Sampling for Place Recognition

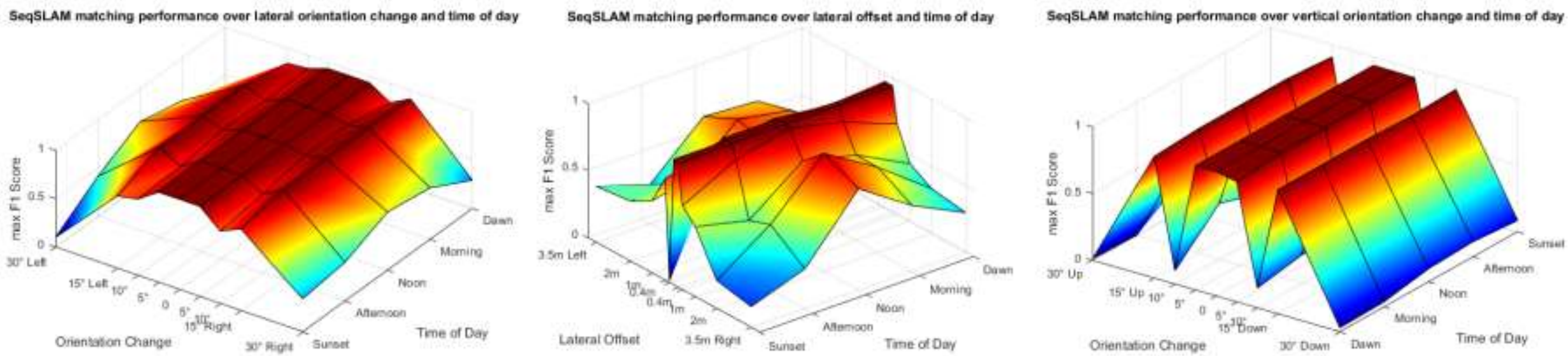
Characterizing Visual SLAM - Progress

- Performance Evaluation Using High-Fidelity Simulation
  - Viewpoint- and Condition-Invariance
    - Place Recognition using SeqSLAM
    - Visual Odometry using ORB-SLAM

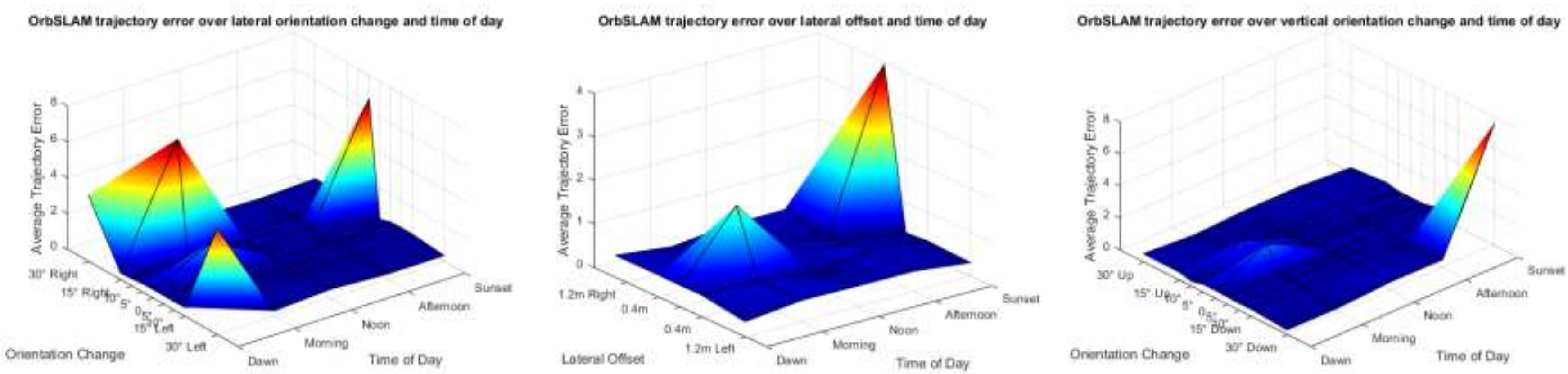


(Skinner 2016)

(Skinner 2016)



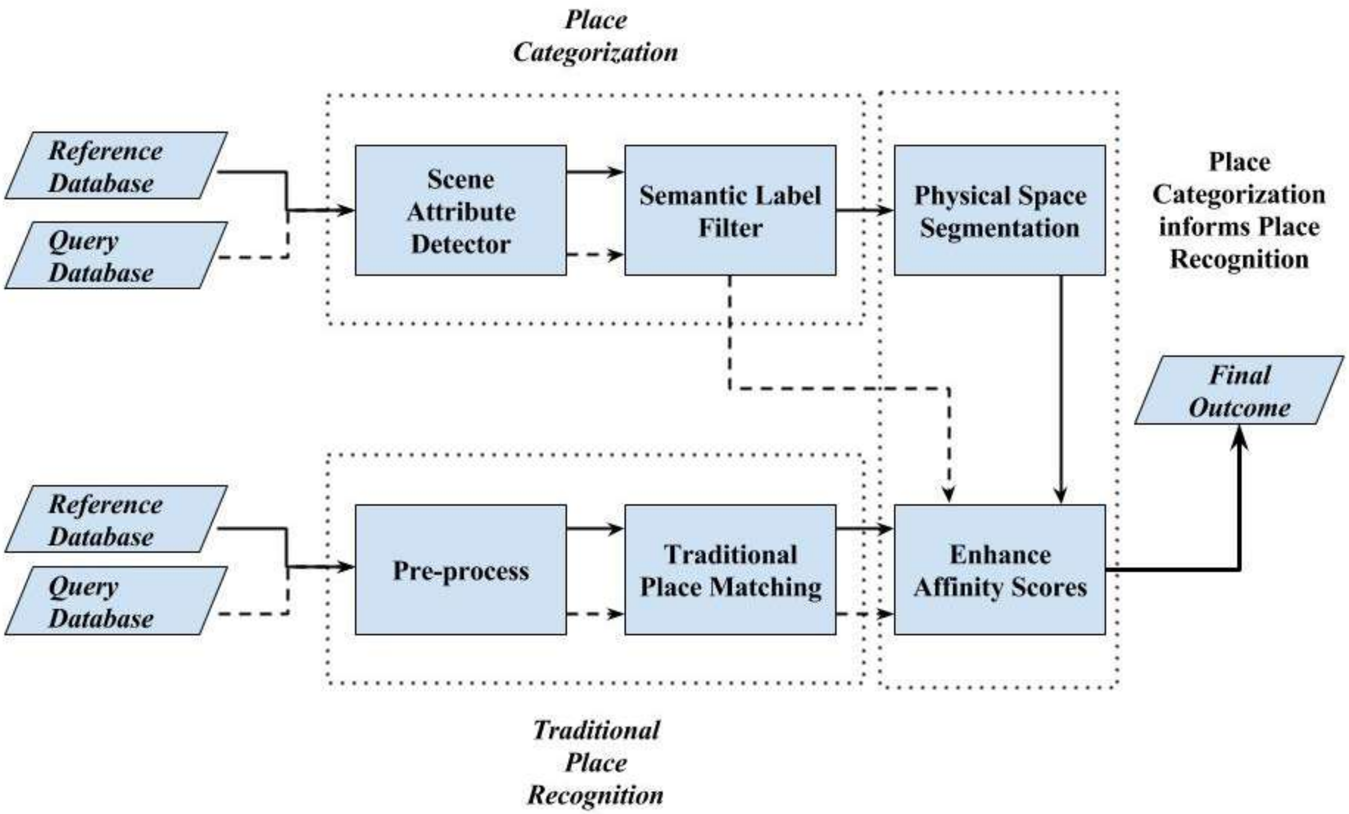
SeqSLAM – Place Recognition

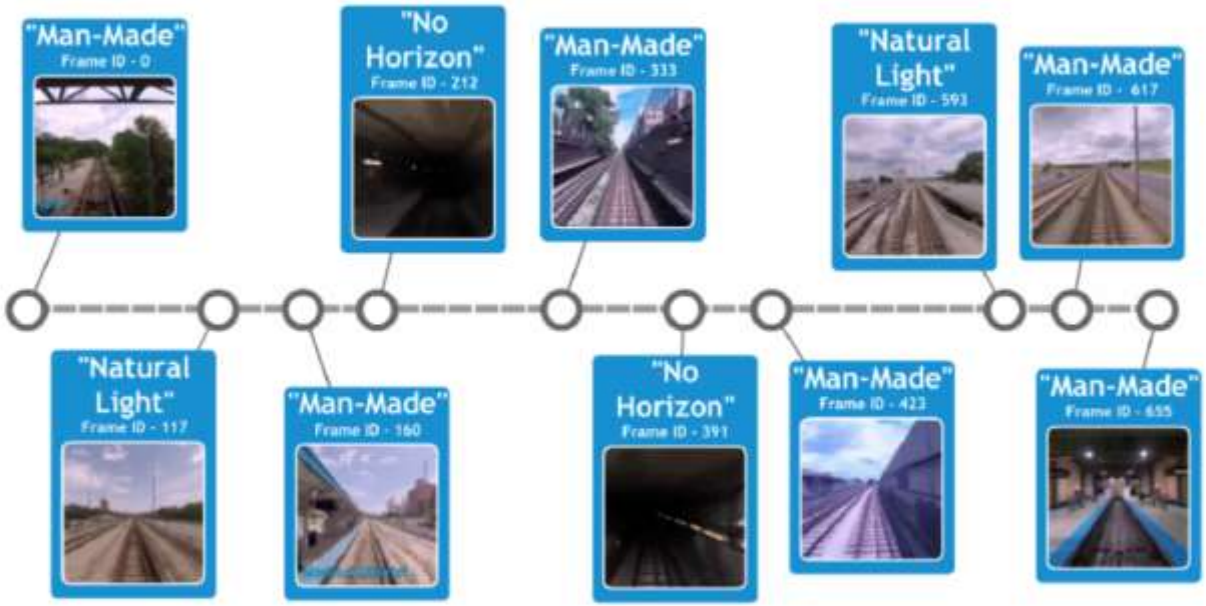


ORB-SLAM – Visual Odometry

Adapting to Environment

- Semantic Segmentation of Environment
- Improving Place Recognition





Semantic Labels for Segmented Environment

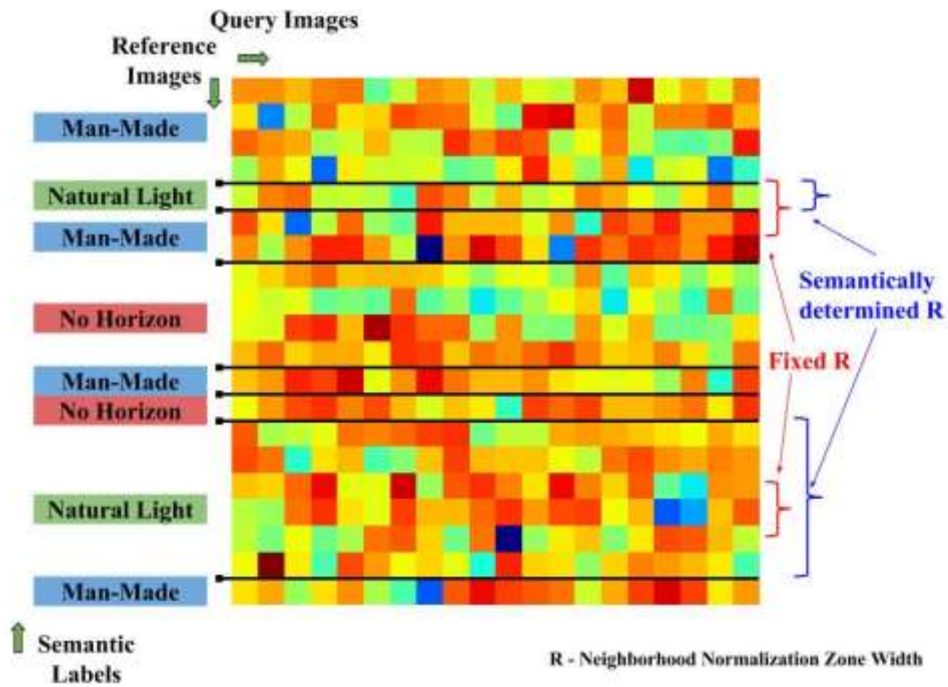
Extreme Variations in Appearance Within Environment

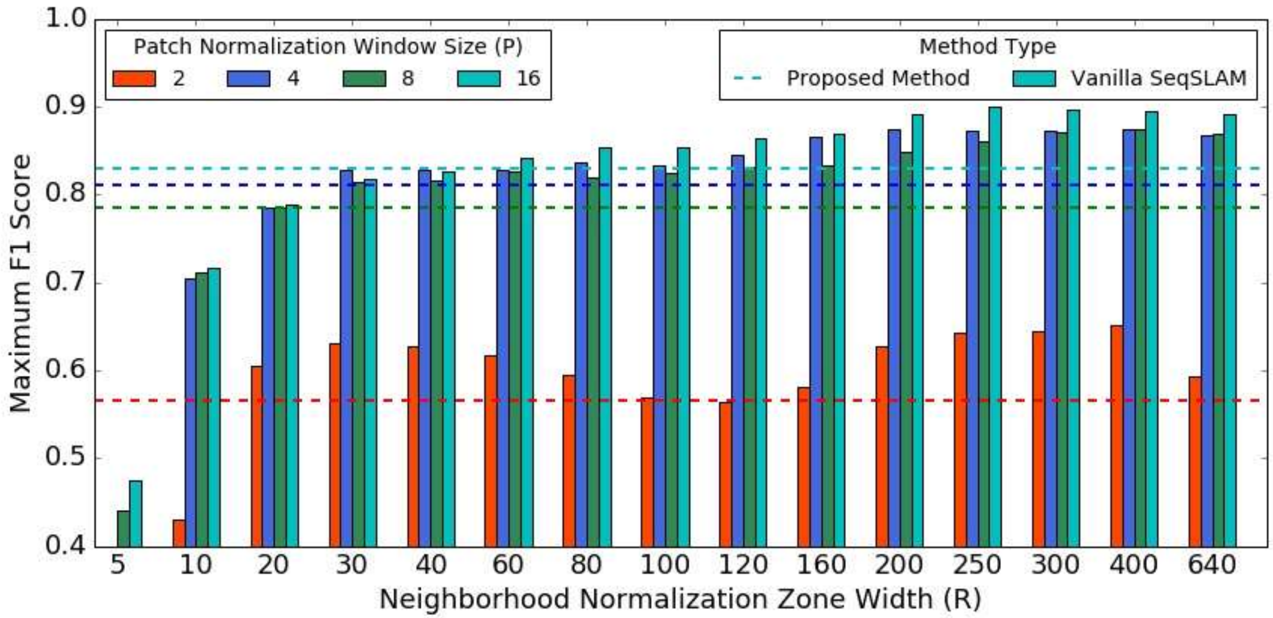
Ground Truth

Vanilla SeqSLAM

Proposed Approach

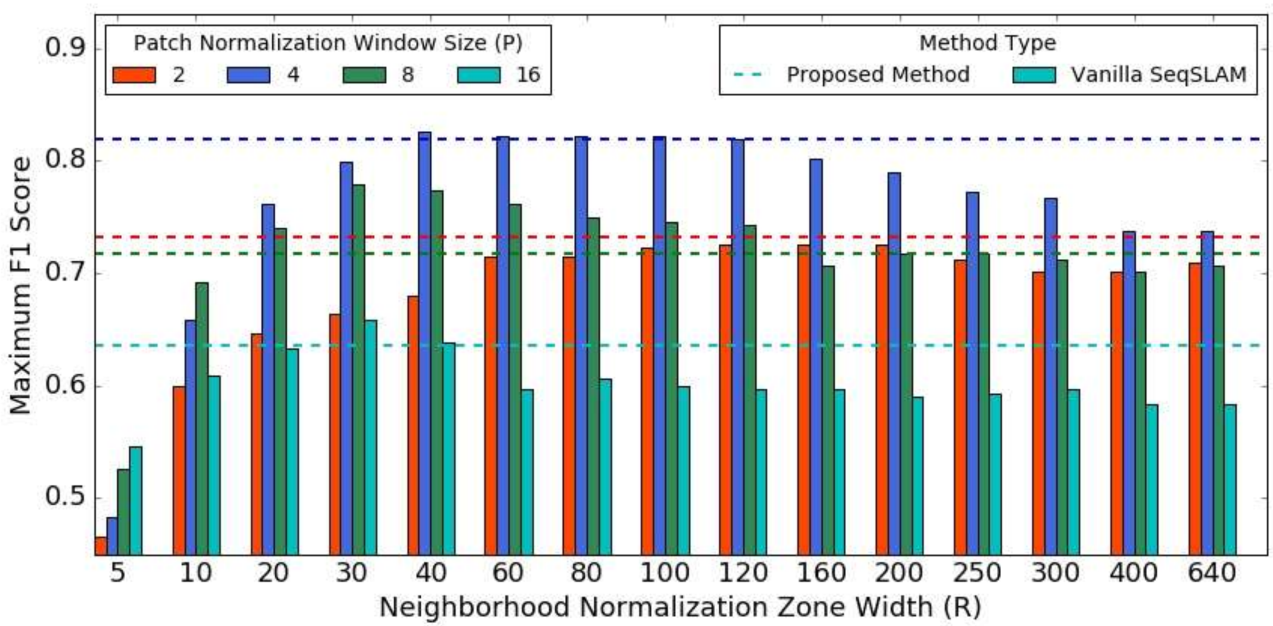
- Performance Changes
  - More pronounced with extreme variations within environment
  - Global Conditions
    - If similar – Global minima generally works
    - If different – Depends on Query Image
  - Semantics later replaced by segmentation based on image matching score




Patch Normalization Window Size (P)
2
4
8
16
Method Type
Proposed Method
Vanilla SeqSLAM
Maximum F1 Score
1.0
0.9
0.8
0.7
0.6
0.5
0.4
5
10
20
30
40
60
80
100
120
160
200
250
300
400
640
Neighborhood Normalization Zone Width (R)

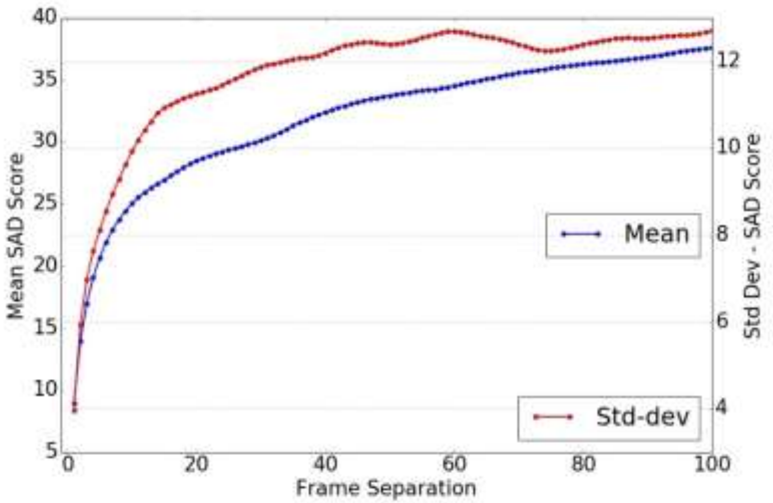

Kedzie-Homan
cta


Kedzie-Homan
cta

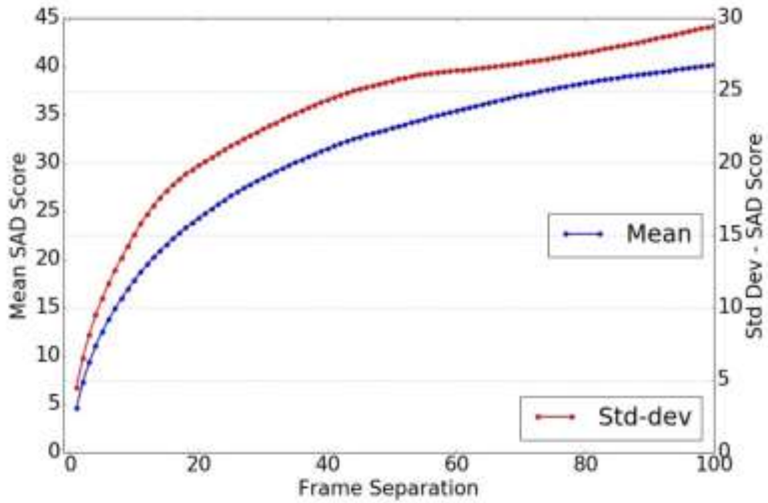

Patch Normalization Window Size (P)
2
4
8
16
Method Type
Proposed Method
Vanilla SeqSLAM
Maximum F1 Score
0.9
0.8
0.7
0.6
0.5
5
10
20
30
40
60
80
100
120
160
200
250
300
400
640
Neighborhood Normalization Zone Width (R)

Egomotion for Unfavourable Conditions

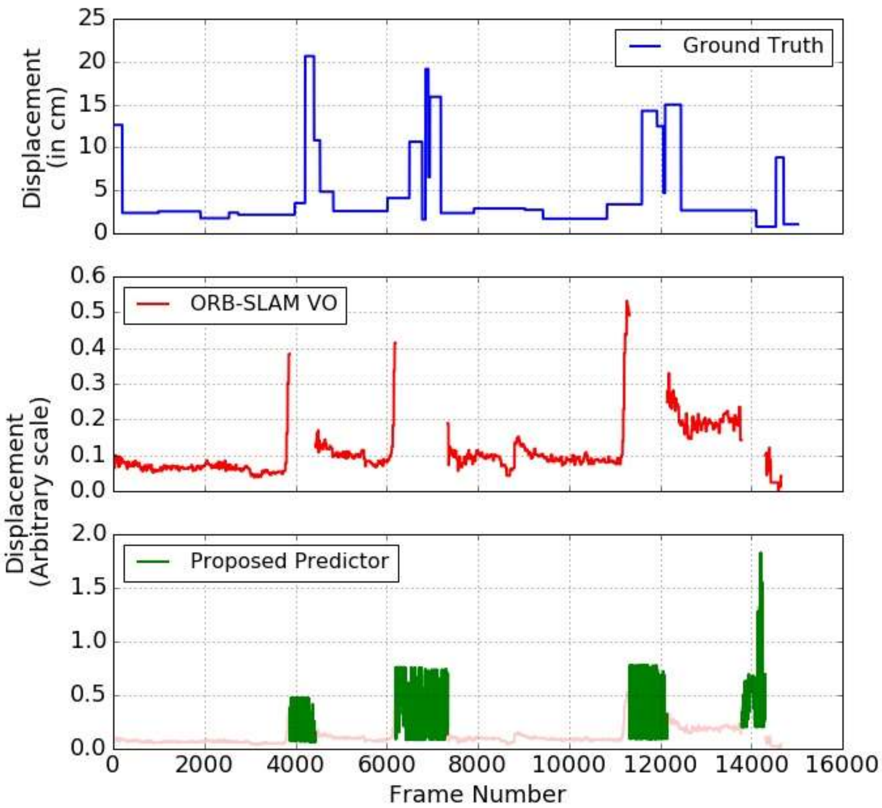
- Speed Estimation
  - Low Light and Rapid Motion
  - Approximate Linear Relationship between Appearance and Geometry (Lukas and Kanade 1981)
  - Image Matching Score vs Frame Separation



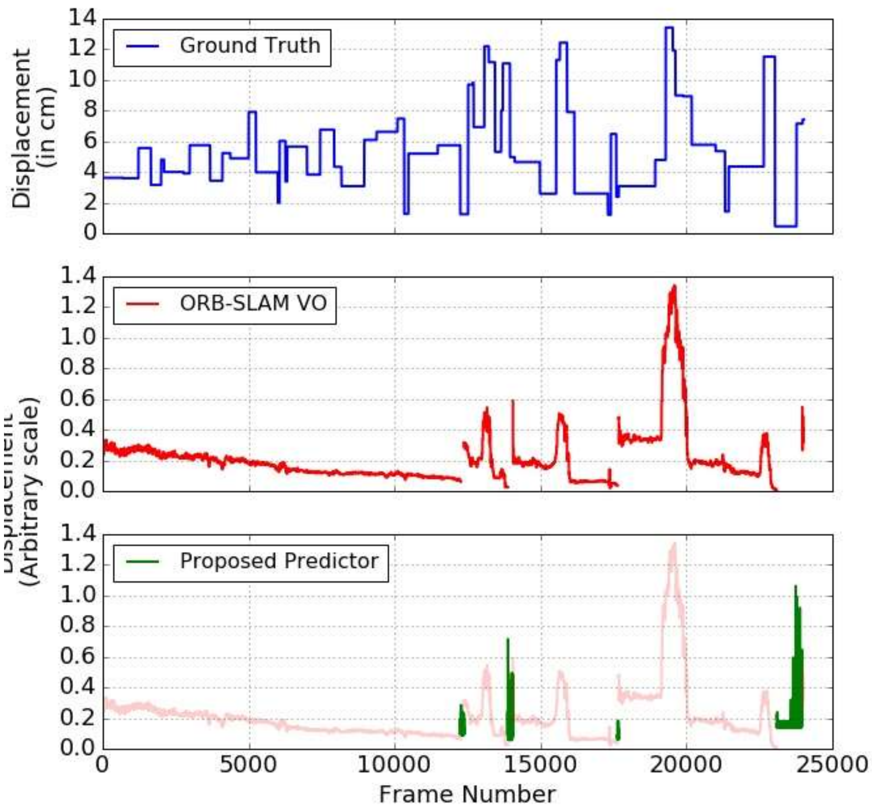
Campus Indoor-Outdoor Dataset



Home Indoor-Outdoor Dataset



Night Traverse

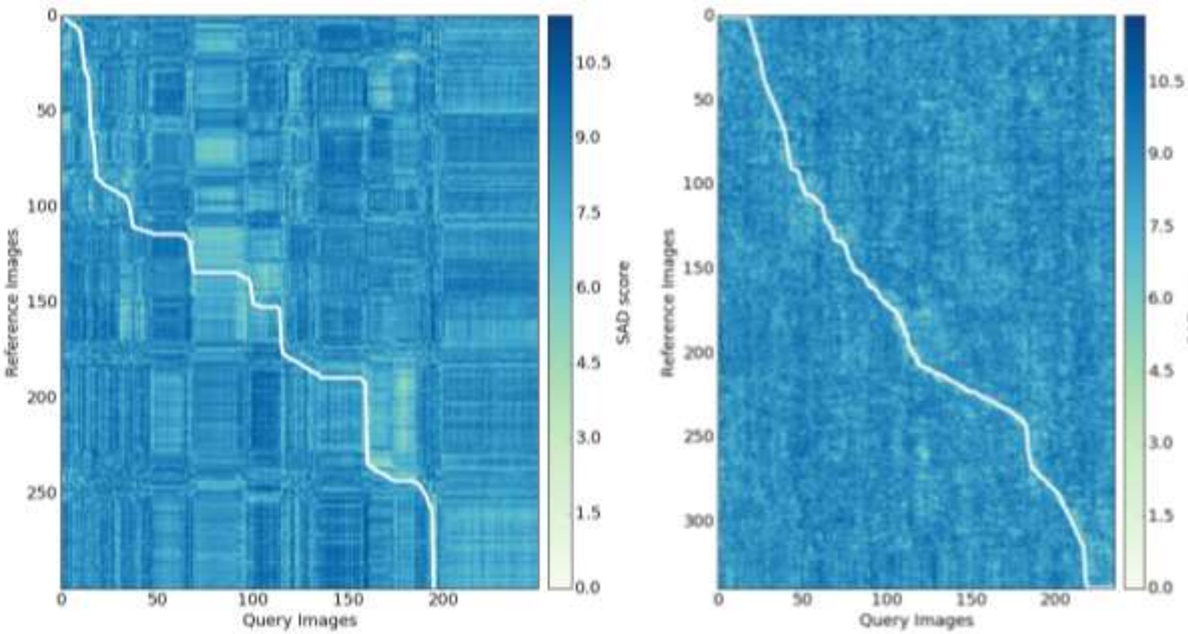


Day Traverse

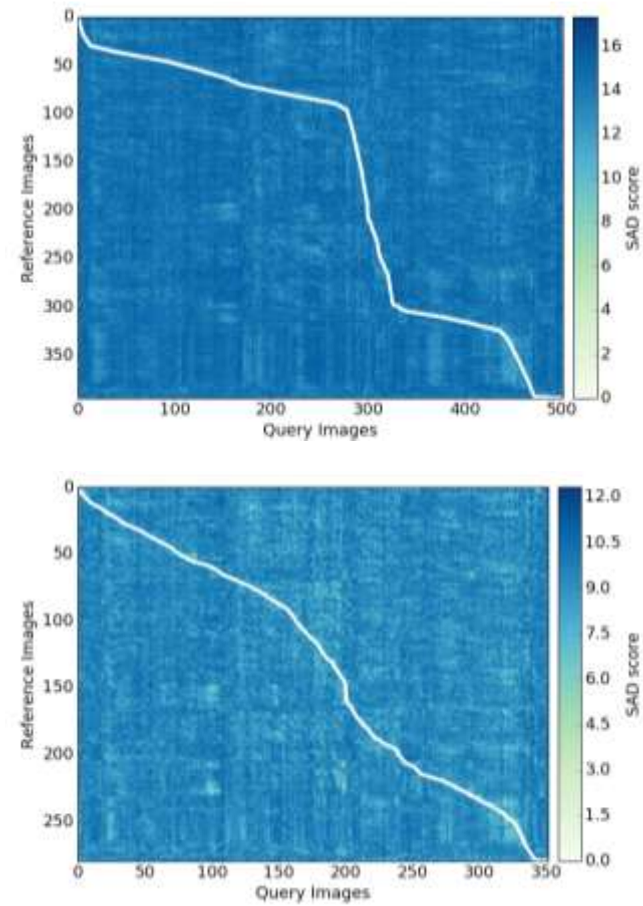
Kelvin Grove On-Foot Dataset featuring Low Light Environment and Rapid Motion

- Image Matching Using
  - Patch-Normalized Images with SAD Score
  - Bit-Plane Descriptor with SSD Score
- Polynomial fitting - SAD score and Frame Separation
- Challenges
  - Depends on Environment
  - Is 1-D
- State-of-the-art VO failure reasons
  - Less features in low-light for ORB-SLAM
  - Mainly rapid motion for LSD-SLAM
- Possible Uses
  - Hybrid approach with state-of-the-art
  - Adapting parameters with change in speed

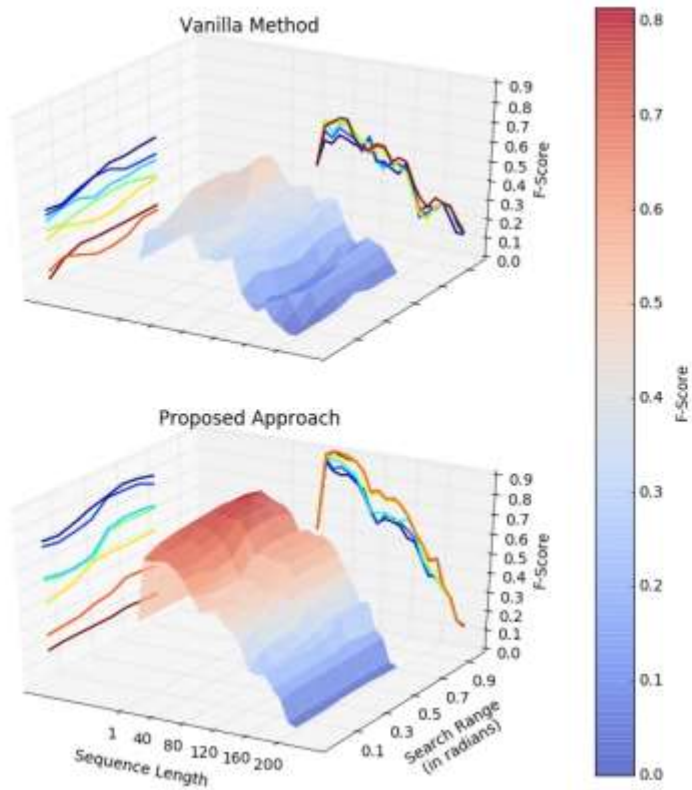
- Improving Place Recognition
  - Speed-Normalized Data Sampling
  - Ground Truth Trajectory Comparison

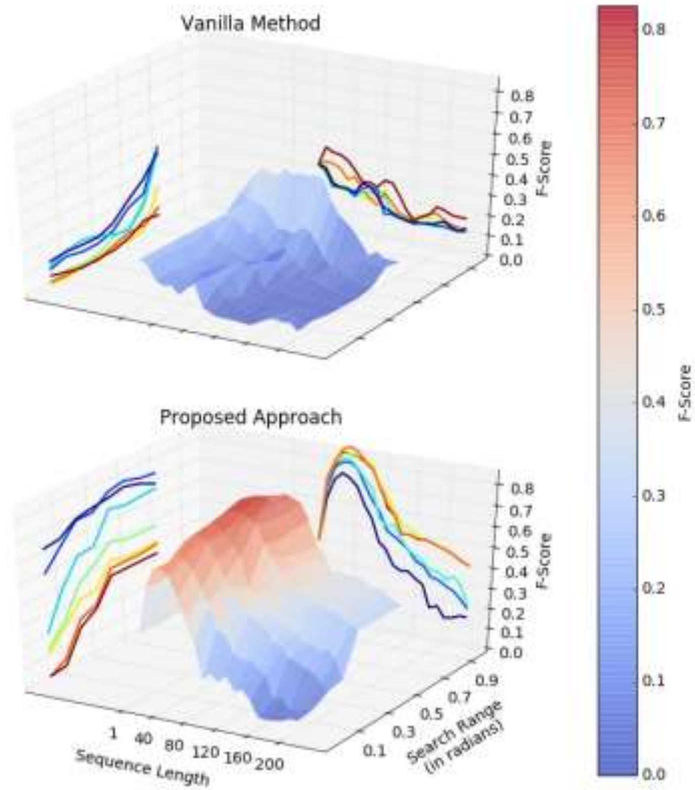


Vehicle Moving in Heavy
Traffic Conditions



Varying Pedestrian
Motion on Footpath





Vehicle Moving in Heavy Traffic Conditions

Varying Pedestrian Motion on Footpath

Thank You!