ASLAM Final Report

Stephen Se
MDA
13800 Commerce Parkway
Richmond, B.C.
Canada
V6V 2J3

Defence R&D Canada

Contract Report
DRDC Suffield CR 2013-071

March 2013

Canada

ASLAM Final Report

Prepared By:
Stephen Se
MDA
13800 Commerce Parkway
Richmond, B.C.
Canada
V6V 2J3

Contractor's Document Number: ASLAM-RP-53-4753
PWGSC Contract Number: W7702-115043/A
CSA: Jack Collier, 403-544-4871

Defence Research and Development Canada – Suffield

Contract Report

DRDC Suffield CR 2013-071

March 2013

IMPORTANT INFORMATIVE STATEMENTS

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Abstract

The objective of the Appearance-based Simultaneous Localization And Mapping (ASLAM) project is the research and development of an Appearance-Based Simultaneous Localization and Mapping system for day/night operations in indoor and outdoor environments. These algorithms would perform place recognition based on sensor data gathered from an Unmanned Ground Vehicle (UGV) as it travels through the environment. When the vehicle returns to a previously visited scene, the ASLAM algorithm would recognize the scene, update its internal representation, report this to the UGV, and finally provide information to aid in closing the loop with geometric Simultaneous Localization And Mapping (SLAM).

This final report describes the work done in this project, including the components of the developed system and the results from the various field trials.

Résumé

L'objectif du présent contrat est la recherche et le développement d'un système de localisation et de cartographie en temps réel basé sur l'apparence pour les opérations menées de jour et de nuit, à l'intérieur comme à l'extérieur. Ces algorithmes doivent effectuer une reconnaissance de l'endroit basée sur les données recueillies par le capteur de l'UGV alors que celui-ci se déplace dans un environnement donné. Lorsque le véhicule revient sur une scène déjà visitée, l'algorithme ASLAM reconnaît la scène, met à jour sa représentation interne, la communique au UGV et, enfin, dispose d'un mécanisme pour fermer la boucle à l'aide du SLAM géométrique.

Le présent rapport final présente les travaux effectués, y compris les éléments du système conçu et les résultats de divers essais sur le terrain.

This page intentionally left blank.

ASLAM

Final Report (Task 4.3)

March 22, 2013

Stephen Se

MDA Systems Ltd.

PWGSC Contract Title:	Appearance Based SLAM for Indoor/Outdoor Urban Terrain
MDA Project Title:	6024 - ASLAM
MDA Document Number:	ASLAM-RP-53-4753
Contract No.:	W7702-115043/A
Project Duration:	October 20 2010 – March 31 2013
DRDC Technical Authority:	Mr. Jack Collier (403) 544-4871

13800 Commerce Parkway
Richmond, B.C., Canada, V6V 2J3
Telephone (604) 278-3411
Fax (604) 231-2753

THIS PAGE INTENTIONALLY LEFT BLANK

ASLAM

Final Report (Task 4.3)

March 22, 2013

Stephen Se

MDA Systems Ltd.

13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

PWGSC Contract Title:	Appearance Based SLAM for Indoor/Outdoor Urban Terrain
MDA Project Title:	6024 - ASLAM
MDA Document Number:	ASLAM-RP-53-4753
Contract No.:	W7702-115043/A
Project Duration:	October 20 2010 – March 31 2013
DRDC Technical Authority:	Mr. Jack Collier (403) 544-4871

Prepared By: Stephen Se

Stephen Se Mar 20, 2013
Signature and Date

Project Manager: Stephen Se

Stephen Se Mar 20, 2013

MacDonald, Dettwiler and Associates Ltd
13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

CHANGE RECORD

ISSUE	DATE	PAGE(S)	DESCRIPTION
1/0	Mar. 22, 2013	All	First Issue

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACRONYMS AND ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
API	Application Programming Interface
ASLAM	Appearance-based Simultaneous Localization And Mapping
CFB	Canadian Forces Base
DOF	Degrees Of Freedom
DRDC	Defence Research & Development Canada
EPG	Experimental Proving Ground
FAB-MAP	Fast Appearance Based Mapping
GHz	GigaHertz
GPS	Global Positioning System
GPU	Graphics Processing Unit
HD	High Definition
Hz	Hertz
IMU	Inertial Measurement Unit
LIDAR	Light Detection and Ranging
MATS	Multi-Agent Tactical Sentry
MB	MegaByte
MDA	MDA Systems Ltd.
RAM	Random Access Memory
RFP	Request For Proposal
ROC	Receiver Operating Characteristics
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous Localization And Mapping
UGV	Unmanned Ground Vehicle
VD-LSD	Variable Dimensional Local Shape Descriptor

THIS PAGE INTENTIONALLY LEFT BLANK

1 INTRODUCTION

1.1 Project Objectives

The objective of the Appearance-based Simultaneous Localization And Mapping (ASLAM) project is the research and development of an Appearance-Based Simultaneous Localization and Mapping system for day/night operations in indoor and outdoor environments. These algorithms would perform place recognition based on sensor data gathered from an Unmanned Ground Vehicle (UGV) as it travels through the environment. When the vehicle returns to a previously visited scene, the ASLAM algorithm would recognize the scene, update its internal representation, report this to the UGV, and finally provide information to aid in closing the loop with geometric Simultaneous Localization And Mapping (SLAM).

1.2 Project Deliverables

The project consists of four phases and includes the following milestone deliverables:

- Phase 1 (October – December 2010)

    o Task 1.2: System Design Plan [R-2] (December 17, 2010)

- Phase 2 (January – September 2011)

    o Task 2.2: Initial ASLAM Software (version 1.0 on March 31, 2011 and version 1.1 on August 31, 2011)

    o Task 2.4: Field Trial Report [R-3] (October 7, 2011)

- Phase 3 (October 2011 – September 2012)

    o Task 3.2: Multi-sensor ASLAM Software (version 1.2 on July 4, 2012 and version 1.2.1 on July 27, 2012)

    o Task 3.4: Multi-Sensor Field Trial Report [R-4] (September 18, 2012)

- Phase 4 (October 2012 – March 2013)

    o Task 4.3: Final ASLAM Software (version 1.3 on March 22, 2013) and Final Report (this report)

    o Task 4.4: ROC Analysis Report [R-7] (March 28, 2013)

1.3 Scope

This final report document is to fulfil Task 4.3 milestone of this contract [R-1] and describes the work done in this project, including the components of the developed system and the results from the various field trials.

The design and interface of the ASLAM Application Programming Interface (API) are described in [R-5]. Details on the vocabulary generation procedure and the visualization of results are described in [R-6]. Doxygen documentation for modules developed in this project is included with the source code delivery to Defence Research & Development Canada (DRDC).

2   SYSTEM DEVELOPMENT

This section summarizes the work done in the various project phases.

2.1   Phase 1 Summary

A full literature review of the state-of-the-art ASLAM systems and a system design plan [R-2] was performed, proposing the key algorithm components to be developed:

- Fast Appearance Based Mapping (FAB-MAP) algorithm using Bag-of-Words, which offers a unified approach for video and range sensors

- Offline training to generate vocabulary to be used during run-time

- Scale Invariant Feature Transform (SIFT) extraction for video data and Variable Dimensional Local Shape Descriptor (VD-LSD) extraction for range data, and the use of Graphics Processing Unit (GPU) to speed up performance

- 6 Degrees Of Freedom (DOF) pose estimation and validation

- Multi-sensor integration

2.2   Phase 2 Summary

The initial system was implemented in Phase 2 according to the system design plan. FAB-MAP software was licensed from Oxford. SIFT and VD-LSD feature extraction on the GPU was implemented by our Brampton team. The initial ASLAM system worked with video sensor or Light Detection and Ranging (LIDAR) sensor individually. It is assumed that stereo camera would be used as the video sensor, in order to allow pose computation in the subsequent phase.

The ASLAM API was implemented as well as a test harness. Moreover, MATLAB scripts were implemented to help with visualizing the results. Two user guide documents were prepared: [R-6] described how to generate the vocabulary and visualize the results, while [R-5] described how to use the ASLAM API and the test harness.

Field trials were performed successfully at the end of Phase 2 and the field trial results were documented in [R-3].

2.3   Phase 3 Summary

Phase 3 continued with the development of the complete ASLAM system where the sensors can work also in combination, with 6 DOF transform estimation. The focus was on the synchronous mode, where the stereo imagery and LIDAR data are collected simultaneously at each frame.

The ASLAM API, the test harness as well as the MATLAB visualization scripts were updated. The two user guide documents [R-5, R-6] were updated accordingly.

Field trials were performed successfully at the end of Phase 3 and the field trial results were documented in [R-4].

2.4   Phase 4 Summary

Based on the multi-sensor field trial results, MDA Systems Ltd. (MDA) and DRDC agreed on a list of ASLAM improvements that were completed in Phase 4, including:

- Output visualization enhancements: Indicate whether the loop detection comes from video/LIDAR, or both, also display the 6 DOF transform when a loop is detected.

- ASLAM API enhancements: Fix continuous mode indexing, move multi-sensor integration to API, use 6 DOF to check pose validity, update AND/OR mode logic.

- Addressing probability splitting:

  o   Merge probabilities of high-likelihood scenes by checking whether they are similar, based on the SIFT/VD-LSD feature matching.

  o   Implement keyframe detection based on how well the SIFT/VD-LSD features from consecutive frames match, which was found to be the best metric for keyframe detection according to a study in [R-9].

- Ease of use: Implement scripts to create folder and populate configuration files automatically

The ASLAM API, the test harness as well as the MATLAB visualization scripts were updated. Multi-threading for multi-sensor synchronous mode has been implemented in the test harness to speed up performance, so that the video and LIDAR processing can run in parallel. The two user guide documents [R-5, R-6] were updated accordingly.

Moreover, an additional task (Task 4.4) was added to the contract to perform detailed Receiver Operating Characteristics (ROC) analysis, which is described in a separate report [R-7]. The ROC analysis was useful to characterize the system performance thoroughly and help fine-tune system parameters.

3   SYSTEM OVERVIEW

The ASLAM system consists of a training phase and a run-time phase. The training phase learns a vocabulary which is used during scene recognition at run-time.

3.1   Training

The training modules are shown in Figure 3-1. Blue corresponds to the video data flow while red corresponds to the range data flow. A video vocabulary and a range vocabulary are generated from the video training data and range training data respectively.

There are four steps to generate the vocabulary:

1. Extract descriptors from the training data. This is done using WordMakerMDA executable. Depending on the type of data:

    a) Compute SIFT descriptors on the GPU from video data

    b) Compute VD-LSD descriptors on the GPU from range data

2. Perform K-Means clustering on the extracted descriptors. This is done using pKMeans executable.

3. Convert the descriptors into appearance vectors based on the K-Means clusters. This is done using WordMakerMDA executable.

4. Learn the probability distribution for the clusters. This is done using pAcceleratedChowLiuFast executable.
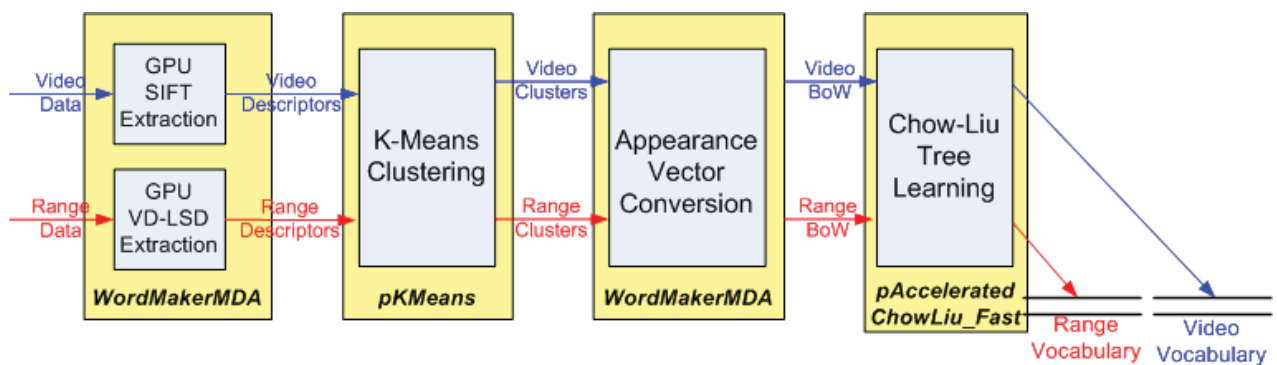

Figure 3-1   ASLAM Training Architecture

3.2 Run-time

Video ASLAM and range ASLAM can run individually, or in combination for multi-sensor ASLAM. Figure 3-2 shows the run-time architecture when keyframe detection is disabled while Figure 3-3 shows the run-time architecture when keyframe detection is enabled.

The key steps are as follows:

1. SIFT/VD-LSD Extraction: Use the GPU to compute SIFT descriptors for video data or VD-LSD descriptors for range data. If keyframe detection is disabled, this is done by WordMakerMDA, otherwise, this is done by the ASLAM Test Harness.

2. Keyframe detection: If enabled, SIFT/VD-LSD features from consecutive frames would be matched to determine whether the new frame has changed much, so that only significantly different frames would be sent to the ASLAM API. The extracted features for the keyframes would also be passed to the ASLAM API, to avoid re-computing the SIFT/VD-LSD features. This is done by the ASLAM Test Harness.

3. Appearance vector conversion: Convert the descriptors into appearance vectors based on the vocabulary from the training phase. This is done by WordMakerMDA executable.

4. Scene recognition: Perform probabilistic scene recognition using the FAB-MAP 2.0 approach. This is done by FabMapV2MDA executable.

5. Merge Probability for Similar Scenes: Check whether the high-likelihood scenes are similar using the SIFT/VD-LSD features already computed. If so, merge probabilities for those scenes. This is to address the probability splitting issue. This is done by the ASLAM API.

6. Validate & Compute 6 DOF Transform:

   a) For video data:

      1. Match SIFT features between the current left and right images to obtain 3D SIFT features.

      2. Match SIFT features from the previous left and right images (according to the loop hypothesis) to obtain 3D SIFT features.

3. Match the current 3D SIFT features with the previous 3D SIFT features.

4. Use RANSAC to select the best matches to discard outliers, by repeatedly selecting 3 matches randomly, computing the number of support and selecting the one with the most support.

5. Compute the 6 DOF rigid body transform from the inlier matches.

b) For range data:

1. Match the current 3D VD-LSD features with the previous 3D VD-LSD features (according to the loop hypothesis).

2. Use RANSAC to select the best matches to discard outliers, by repeatedly selecting 3 matches randomly, computing the number of support and selecting the one with the most support.

3. Compute the 6 DOF rigid body transform from the inlier matches.

If pose computation is not successful, the validation fails. This is done by the ASLAM API.

7. Multi-Sensor Integration: Integrate the video ASLAM result with the range ASLAM result to produce an integrated output. The user can choose OR/AND integration mode. For the OR mode, it would report recognition if either sensor reports recognition, whereas for the AND mode, it would only report recognition if both sensors report recognition. This is done by the ASLAM API.

3.3   Performance

Table 3-1 shows the average processing time per frame for a rural dataset where the vehicle drove around a loop twice, with 200 frames in each loop. The performance is 1 Hz for multi-sensor ASLAM, and higher than 1 Hz for single-sensor ASLAM. Loop detection and pose computation were successful for most of the frames during the second loop. The run-time was measured on the ASLAM desktop computer provided by DRDC with:

- 3.33GHz Hexa-Core Intel Core i7 Extreme Processor

- 12GB DDR3 RAM

- GeForce GTX 580 1536MB graphics card

Table 3-1   System Run-time Performance

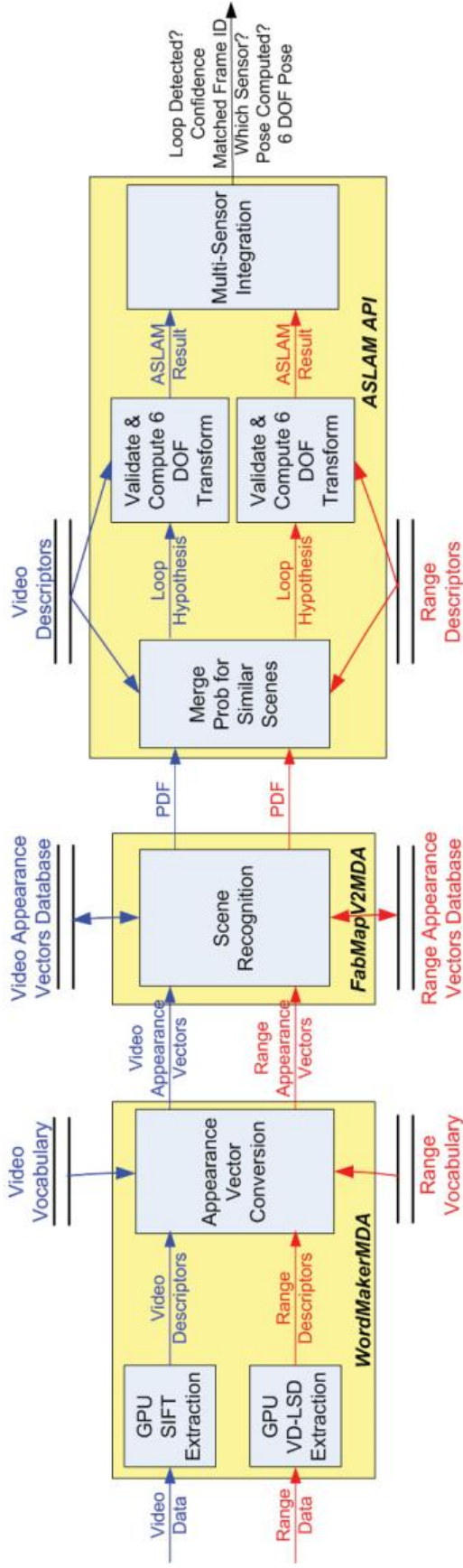| Processing Mode	Average Processing Time Per Frame
Video only with pose computation	0.9 second
Range only with pose computation	0.6 second
Multi-sensor integration with pose computation	1.0 second

Figure 3-2   ASLAM Run-time Architecture without Keyframe Detection
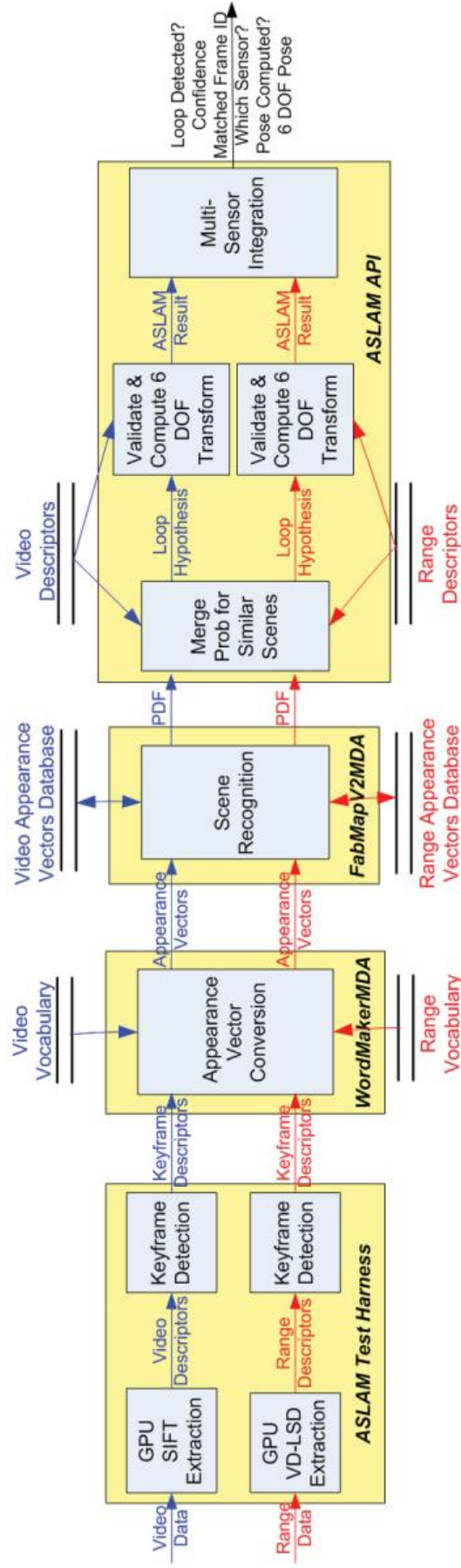

Figure 3-3   ASLAM Run-time Architecture with Keyframe Detection

4   FIELD TRIALS

Field trials consisted of manually driving the Multi-Agent Tactical Sentry (MATS) vehicle for several kilometres while collecting LIDAR, stereo and Global Positioning System (GPS) data. The MATS vehicle is equipped with a Velodyne High Definition LIDAR (360 degrees field-of-view) and Point Grey Bumblebee XB3 stereo camera (45 degrees field-of-view), as shown in Figure 4-1. Differential GPS data was collected for verification and display purposes only, but was not used by the ASLAM system.

The field trial data was collected at the DRDC Experimental Proving Ground (EPG) at CFB Suffield, Alberta. Two environments were considered:

- Urban environment with many buildings, vehicles, pavements, etc.

- Rural environment with prairie grass, gravel roads, sparse buildings

Figure 4-1   DRDC MATS Vehicle with Velodyne HD LIDAR & Point Grey Stereo Camera

4.1   Phase 2 Field Trial Summary

The initial field trial for Phase 2 included using logged data and live trial on September 28-29, 2011, when MDA personnel were at Suffield. Please refer to [R-3] for details.

The initial field trial was completed successfully, meeting all the applicable evaluation criteria listed in DRDC's field trial plan [R-8]:

- Real-time operation: Processing time within 1 second per frame for both video and LIDAR ASLAM.

- Single-sensor ASLAM: Both video and LIDAR ASLAM were working individually.

- Indoor/Outdoor: Only outdoor environment was tested in Phase 2. Both urban and rural environments were tested. Some false positives in video ASLAM were due to similarity of the cloud in the images.

- Day/Night: Only day-time operation was tested in Phase 2.

- Rotation Invariant: Image ASLAM could recognize scenes from a number of different viewing angles. LIDAR ASLAM could recognize scenes with 180 degree rotation.

- MATLAB Interface: MATLAB tools were delivered and used for evaluation and displaying results.

- Geometric Loop Closing: Not tested in Phase 2.

- Continuous Operations: Tested for running over 1 hour after the field trial.

- Database Storage: ASLAM results using a database were identical to results obtained from running continuously.

- Area of Operation: Both video and LIDAR ASLAM were working in previously unseen areas.

4.2   Phase 3 Multi-sensor Field Trial Summary

The multi-sensor field trial for Phase 3 included using logged data over a 2-week period in August 2012. MDA personnel were at Suffield on August 24-25, 2012. Please refer to [R-4] for details.

The multi-sensor field trial was completed successfully. The trial results showed good recall with zero false alarms. In addition to the evaluation criteria met in Phase 2, the following criteria were also tested:

- Multi-sensor ASLAM: Both video and LIDAR ASLAM worked in combination and the results were integrated.

- Day/Night: System was tested using hourly datasets spanning from 9am to 9pm, covering day and night time. The results showed that the video and LIDAR sensors complement each other well.

- Geometric Loop Closing: The 6 DOF transform was estimated when a loop detection occurred. The computed transform compared well with the GPS data. This proved to be useful in validating the FAB-MAP results, thereby resulting in zero false alarms.

4.3   Phase 4 Final Field Trial

The final field trial for Phase 4 included re-processing the field trial datasets from 2012 with the final software. The focus of the final field trial is on multi-sensor ASLAM and the Phase 4 improvements over the Phase 3 field trial results. Mainly synchronous datasets are considered as they allow for multi-sensor integration.

Since the frames were collected relatively sparsely for the synchronous datasets, keyframe detection does not find many similar consecutive frames and most of the frames are sent to the ASLAM API, therefore keyframe detection is disabled for the synchronous datasets. One asynchronous dataset is included to evaluate the keyframe detection functionality.

4.3.1   Data Description

The following datasets from 2012 are used for the final field trial:

- August 9, 2012:
    - Urb1: Asynchronous dataset (large loop course)
- August 12, 2012:
    - Urb1: Synchronous dataset (random loops and backtracking)
    - Urb2: Synchronous dataset (large loop course)
- August 15, 2012:
    - Rur1: Synchronous dataset
    - Rur2: Synchronous dataset
- August 20, 2012:
    - 9am, 10am, 11am, 12pm, 1pm, 2pm, 3pm, 4pm, 5pm, 6pm, 7pm, 8pm, 9pm: Synchronous variable lighting rural datasets (repeat same loop twice every hour)

Each of the datasets includes the stereo imagery, the LIDAR data and the GPS and Inertial Measurement Unit (IMU) heading ground truth information.

Separate urban and rural vocabularies are used for the following trials, using default parameters for vocabulary generation, i.e., SIFT scale threshold of 2.0 for video vocabulary, min_abs_E3_to_E1 of 0.1 for LIDAR vocabulary.

- Urban vocabulary: Subset of the Aug8_Urb4 dataset
- Rural vocabulary: Subset of the Aug16_Rur1 dataset

The FAB-MAP loop detection probability threshold is set to 99% in the following tests, to minimize the number of false positives. Detailed ROC analysis for various thresholds is described in [R-7].

4.3.2   Urban Trials

4.3.2.1   Aug12_Urb1

For this dataset, the vehicle traversed around 5.9km of a large outdoor urban environment, including several loops for portions of the trajectory. Figure 4-2 shows the distance between the vehicle positions computed from the GPS information, to serve as the ground truth. Red indicates locations far away from each other, while blue indicates close by.

Figure 4-3 shows the computed scene recognition probabilities for the OR integration scheme. The ith row and j[th] column indicate whether frames i and j are similar. The diagonal elements indicate whether the current frame is a new scene. Red indicates a high probability while blue indicates a low probability.

Figure 4-4 shows the trajectory of this trial, where the positional information was provided by the GPS. The blue dot indicates no loop detection, the red dot indicates successful loop detection, while the green dot indicates a FAB-MAP loop detection that was rejected by 6 DOF validation.

Table 4-2 shows a comparison of the true positives and false positives between Phase 3 (2012) and Phase 4 (2013) results. It can be seen that the number of true positives have improved substantially in 2013, with very few false positives.
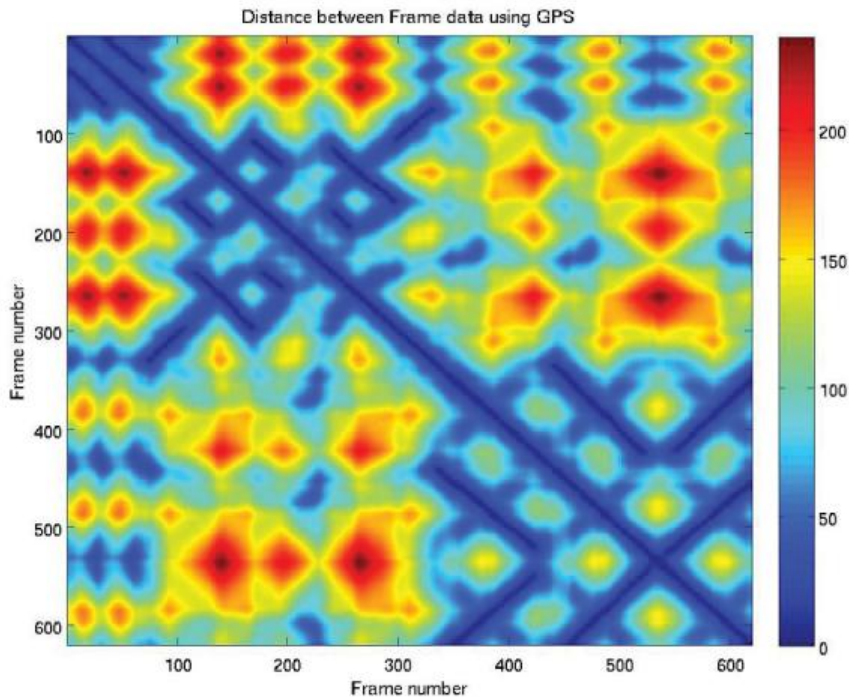
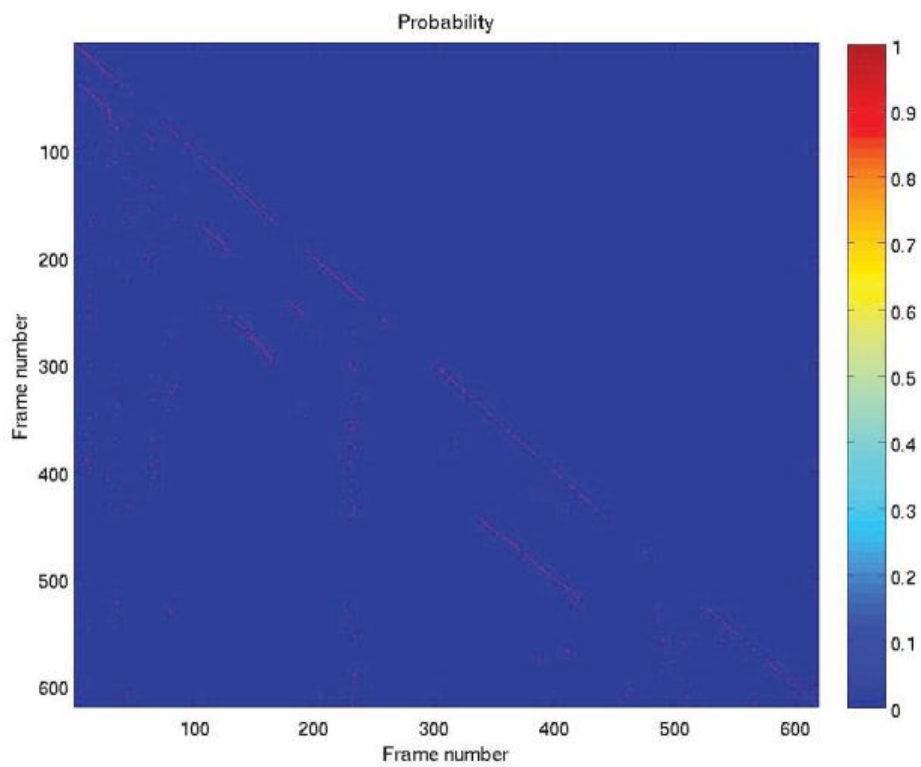**Figure 4-2   Ground Truth for Aug12_Urb1 Dataset**

Figure 4-3   Probability Output for Aug12_Urb1 Dataset
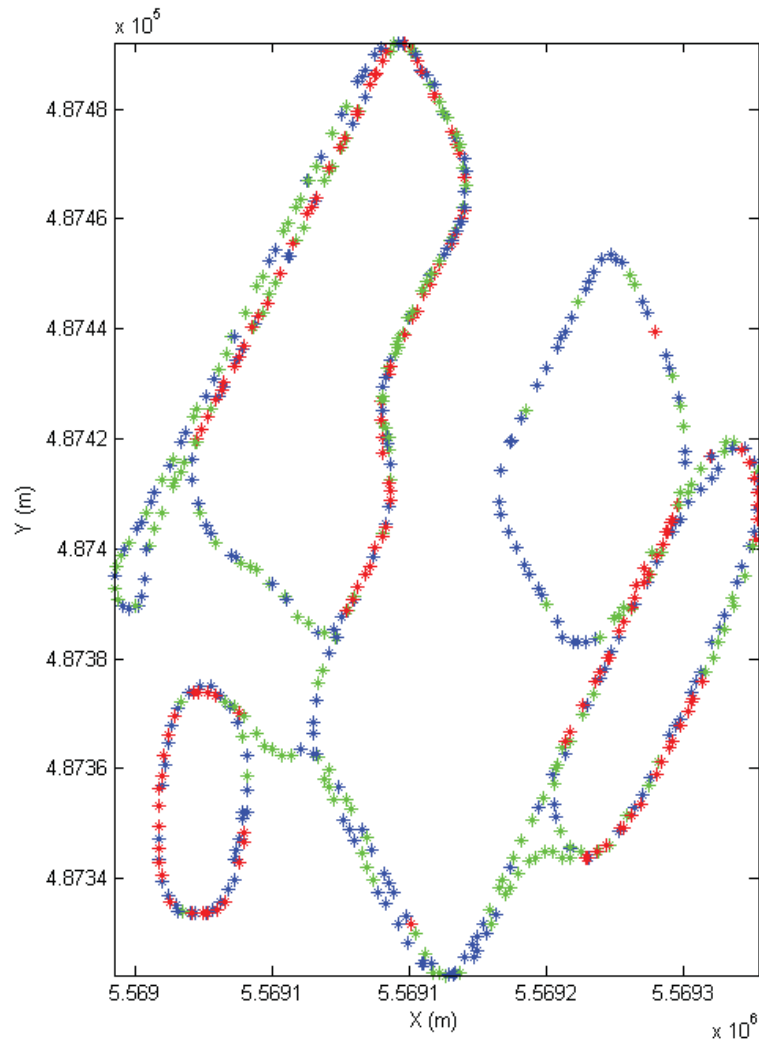
4-5

Figure 4-4   Trajectory and Loop Closure for Aug12_Urb1 Dataset (99% Threshold)

Table 4-2   True Positives and False Positives Comparison for Aug12_Urb1 Trial

	2012 Results		2013 Results	
	True Positives	**False Positives**	**True Positives**	**False Positives**
Imagery only	107	0	110	0
LIDAR only	30	0	93	5
OR Integration Scheme	112	0	142	5
AND Integration Scheme	25	0	64	0

4.3.2.2   Aug12_Urb2

For this dataset, the vehicle traversed around 6.5km of a large outdoor urban environment, including two large loops. Figure 4-5 shows the distance between the vehicle positions computed from the GPS information, to serve as the ground truth. Figure 4-6 shows the computed scene recognition probabilities for the OR integration scheme. Figure 4-7 shows the trajectory of this trial, where the positional information was provided by the GPS. The blue dot indicates no loop detection, the red dot indicates successful loop detection, while the green dot indicates a FAB-MAP loop detection that was rejected by 6 DOF validation.

Table 4-3 shows a comparison of the true positives and false positives between Phase 3 (2012) and Phase 4 (2013) results. It can be seen that the number of true positives have improved substantially in 2013, with very few false positives.
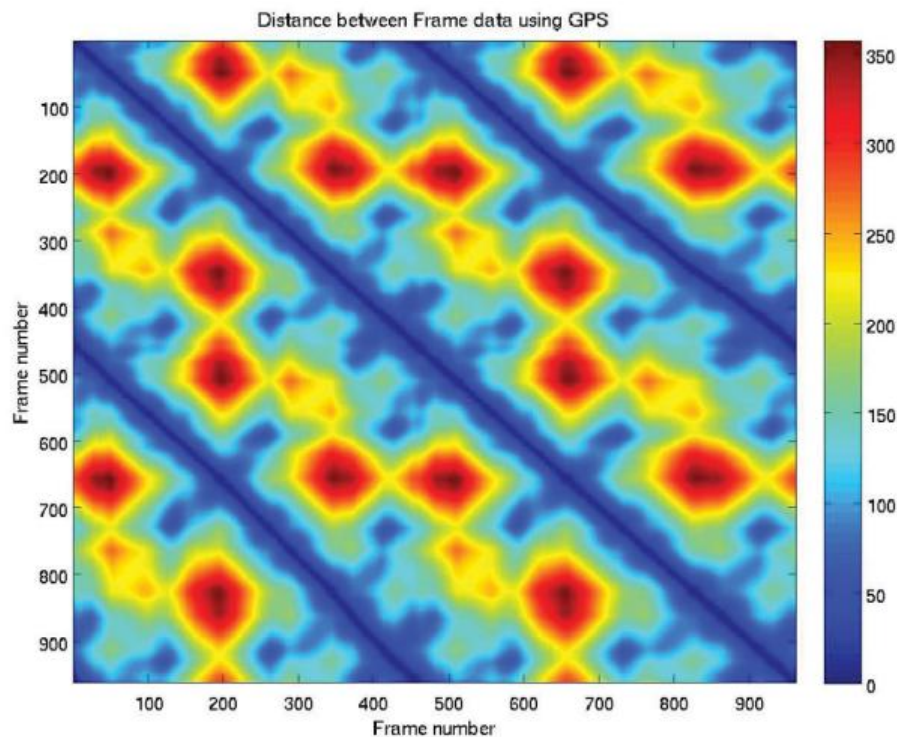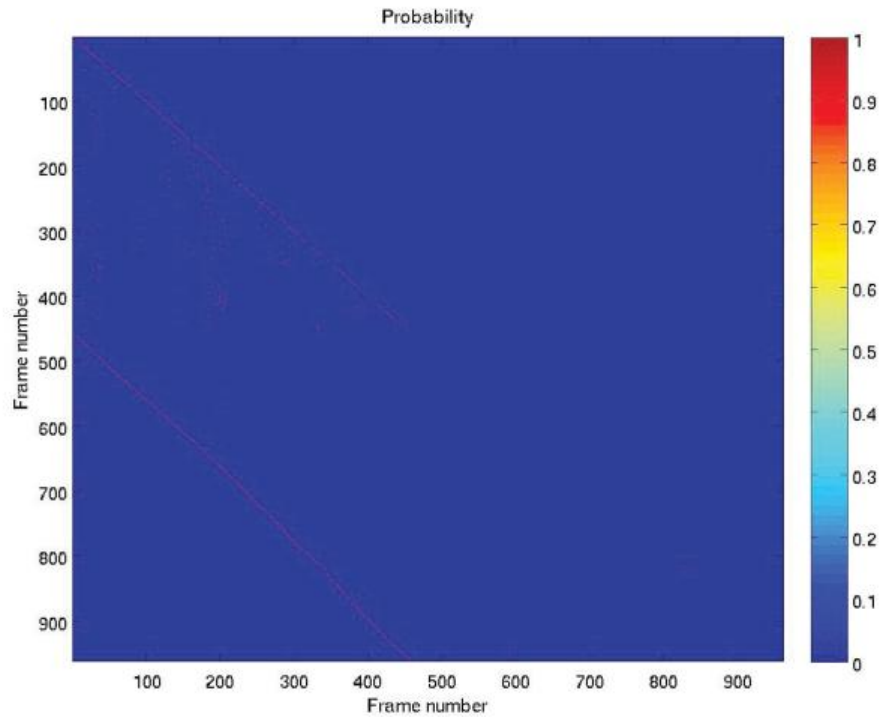

Figure 4-5   Ground Truth for Aug12_Urb2 Dataset

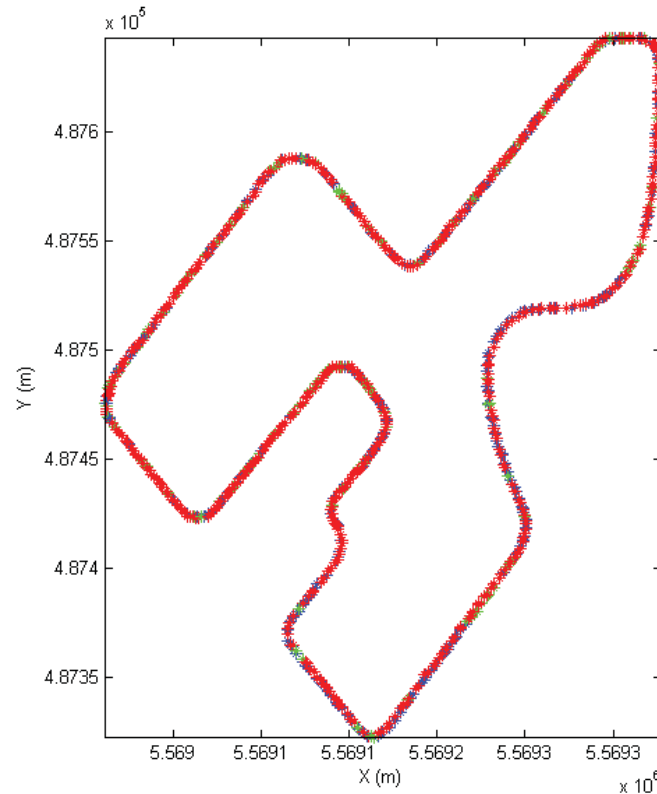**Figure 4-6   Probability Output for Aug12_Urb2 Dataset**

Figure 4-7   Trajectory and Loop Closure for Aug12_Urb2 Dataset (99% Threshold)

Table 4-3 True Positives and False Positives Comparison for Aug12_Urb2 Trial

	2012 Results		2013 Results	
	True Positives	False Positives	True Positives	False Positives
Imagery only	370	0	451	0
LIDAR only	210	0	408	2
OR Integration Scheme	412	0	490	2
AND Integration Scheme	167	0	369	0

4.3.3　Rural Trials

4.3.3.1　Aug15_Rur1

For this dataset, the vehicle traversed around 3.9km of a large outdoor rural environment, including several loops for portions of the trajectory. Figure 4-8 shows the distance between the vehicle positions computed from the GPS information, to serve as the ground truth. Figure 4-9 shows the computed scene recognition probabilities for the OR integration scheme. Figure 4-10 shows the trajectory of this trial, where the positional information was provided by the GPS. The blue dot indicates no loop detection, the red dot indicates successful loop detection, while the green dot indicates a FAB-MAP loop detection that was rejected.

Table 4-4 shows a comparison of the true positives and false positives between Phase 3 (2012) and Phase 4 (2013) results. It can be seen that the number of true positives have improved substantially in 2013, with very few false positives.
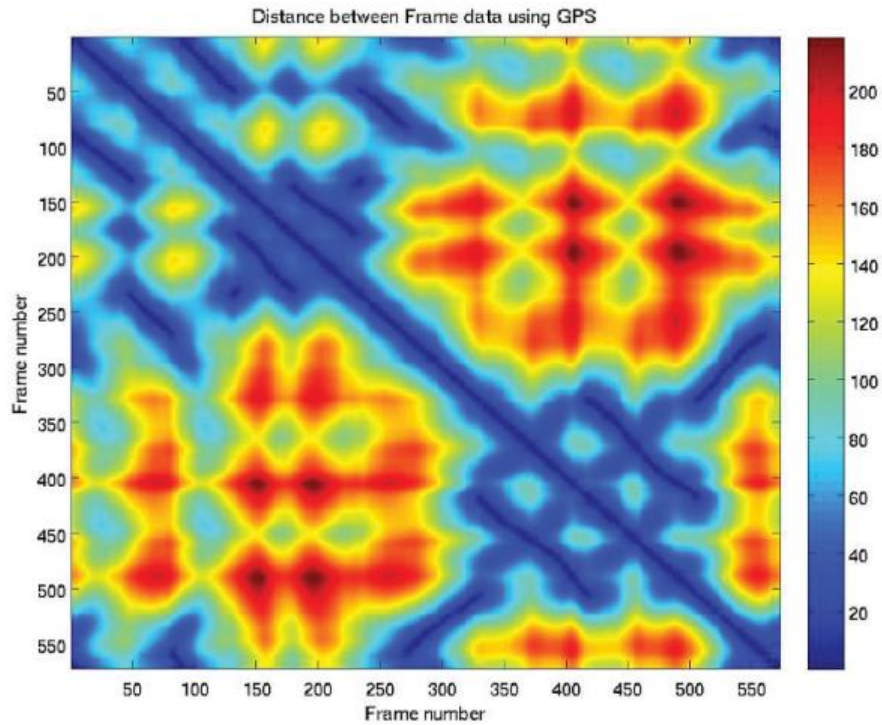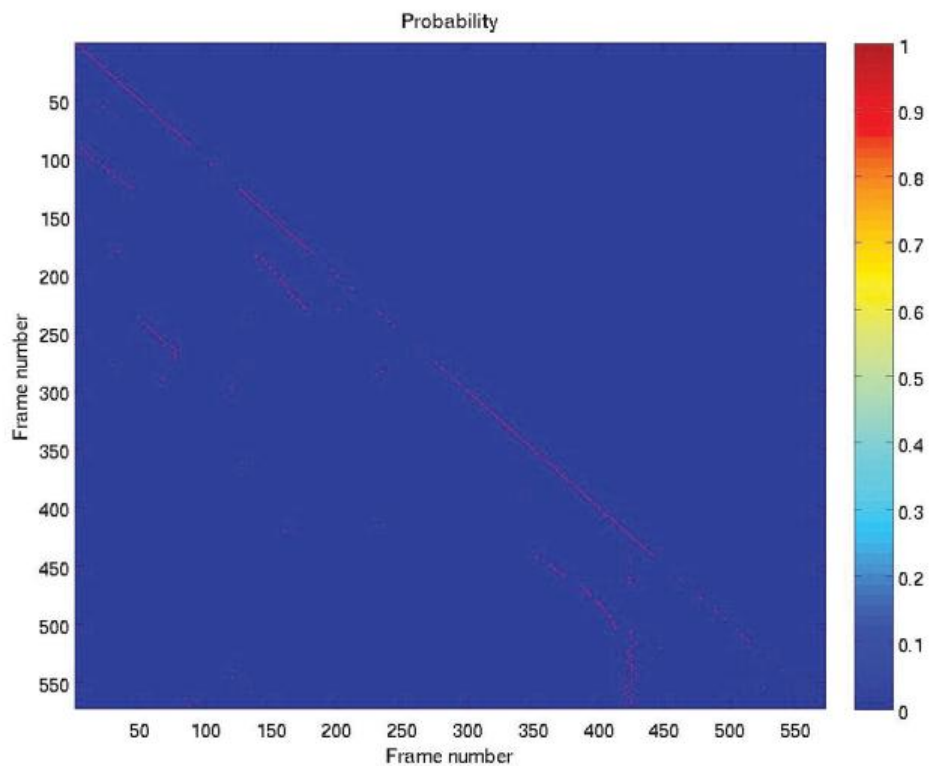
Figure 4-8   Ground Truth for Aug15_Rur1 Dataset

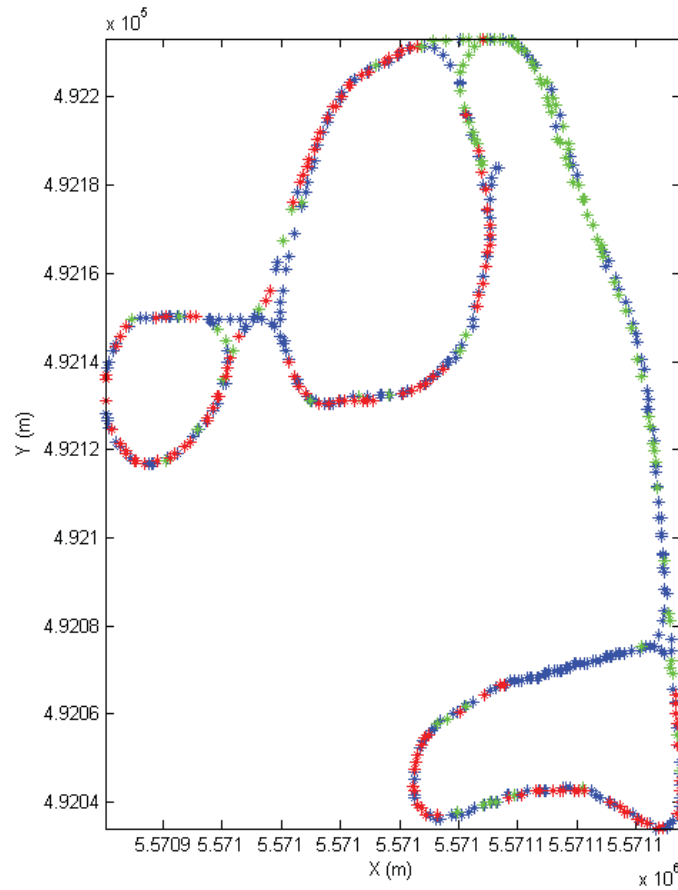**Figure 4-9   Probability Output for Aug15_Rur1 Dataset**

4-10

Figure 4-10   Trajectory and Loop Closure for Aug15_Rur1 Dataset (99% Threshold)

Table 4-4   True Positives and False Positives Comparison for Aug15_Rur1 Trial

	2012 Results		2013 Results	
	True Positives	**False Positives**	**True Positives**	**False Positives**
Imagery only	70	0	87	0
LIDAR only	21	0	68	2
OR Integration Scheme	84	0	131	2
AND Integration Scheme	8	0	23	0

4-11

4.3.3.2 Aug15_Rur2

For this dataset, the vehicle traversed around 5.1km of a large outdoor rural environment, including several loops for portions of the trajectory. Figure 4-11 shows the distance between the vehicle positions computed from the GPS information, to serve as the ground truth. Figure 4-12 shows the computed scene recognition probabilities for the OR integration scheme. Figure 4-13 shows the trajectory of this trial, where the positional information was provided by the GPS. The blue dot indicates no loop detection, the red dot indicates successful loop detection, while the green dot indicates a FAB-MAP loop detection that was rejected.

Table 4-5 shows a comparison of the true positives and false positives between Phase 3 (2012) and Phase 4 (2013) results. It can be seen that the number of true positives have improved substantially in 2013, with very few false positives.
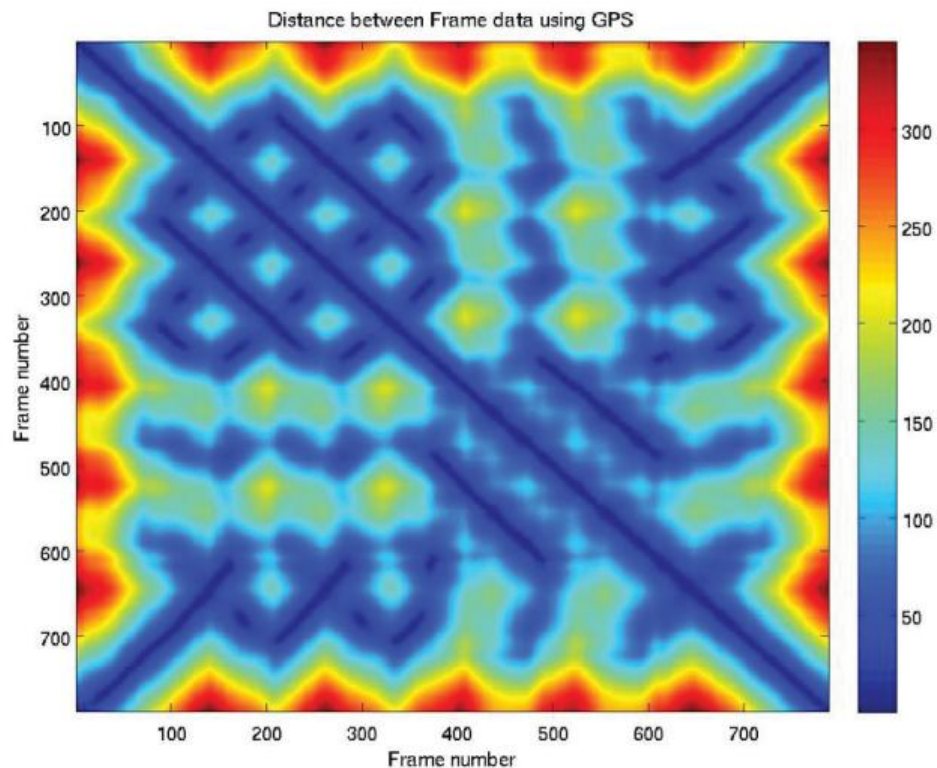

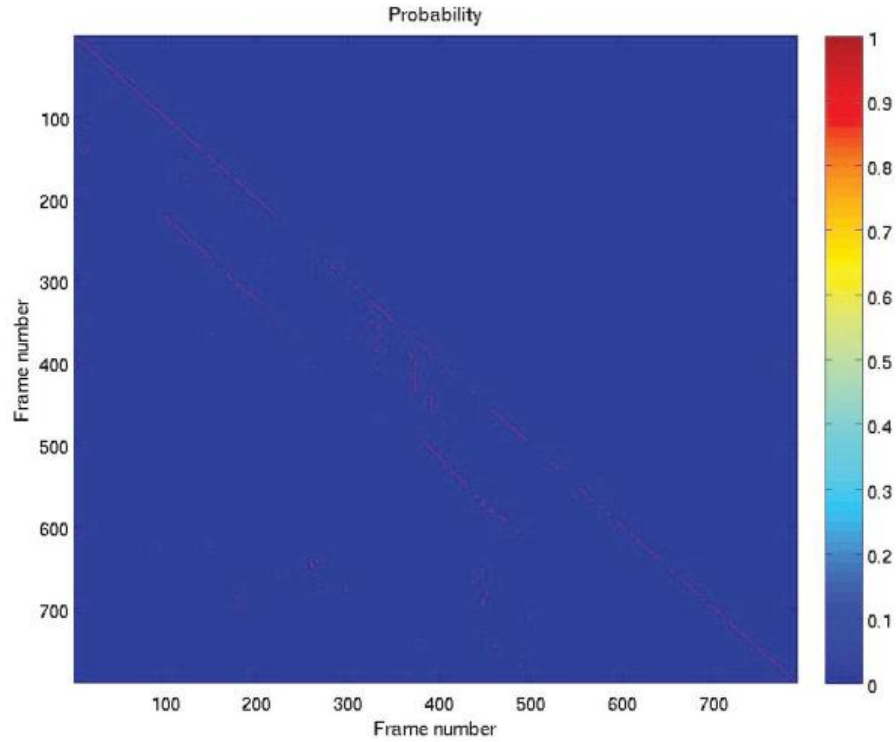**Figure 4-11   Ground Truth for Aug15_Rur2 Dataset**
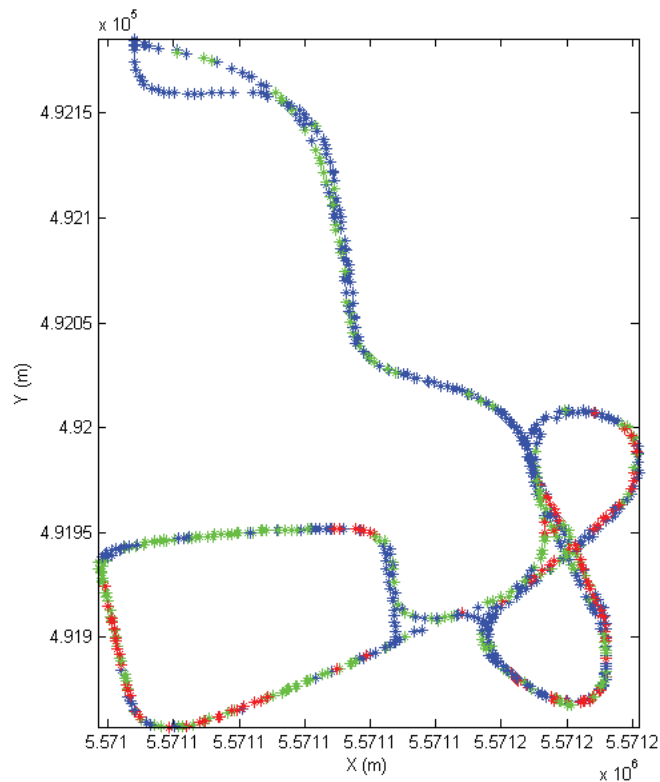
Figure 4-12   Probability Output for Aug15_Rur2 Dataset

Figure 4-13   Trajectory and Loop Closure for Aug15_Rur2 Dataset (99% Threshold)

4-13

Table 4-5   True Positives and False Positives Comparison for Aug15_Rur2 Trial

	2012 Results		2013 Results	
	True Positives	**False Positives**	**True Positives**	**False Positives**
Imagery only	60	0	72	0
LIDAR only	33	0	58	1
OR Integration Scheme	86	0	115	1
AND Integration Scheme	7	0	15	0

4.3.4   Variable Lighting Trials

The multi-time rural dataset Aug20 is used for this trial. We evaluate the effectiveness of the multi-sensor approach over a 12-hour period (9am – 9pm) where the vehicle traversed the same loop of a rural environment (around 3km for each loop) as the illumination varies. The imagery varies considerably due to the illumination changes throughout the 12-hour period while the LIDAR data is stable and not affected by the ambient illumination, as expected.

Using the 12pm dataset as the initial loop, we saved the results to the database, and then processed each additional hourly dataset as the second loop in the continuous mode. By traversing the same loop and knowing each loop has 200 frames, we can calculate the recall rate easily.

Figure 4-14 shows the ground truth GPS distance for the 11am test run, i.e. using the 12pm as the first loop (200 frames) and using the 11am as the second loop (200 frames). Figure 4-15 shows the probability output for the 11am test run while Figure 4-16 shows the trajectory and loop detection results. It can be seen that the recognition is quite consistent for most of the loop. The right hand side of the trajectory does not have much detection as that region does not have any distinctive 3D structures or 2D features.
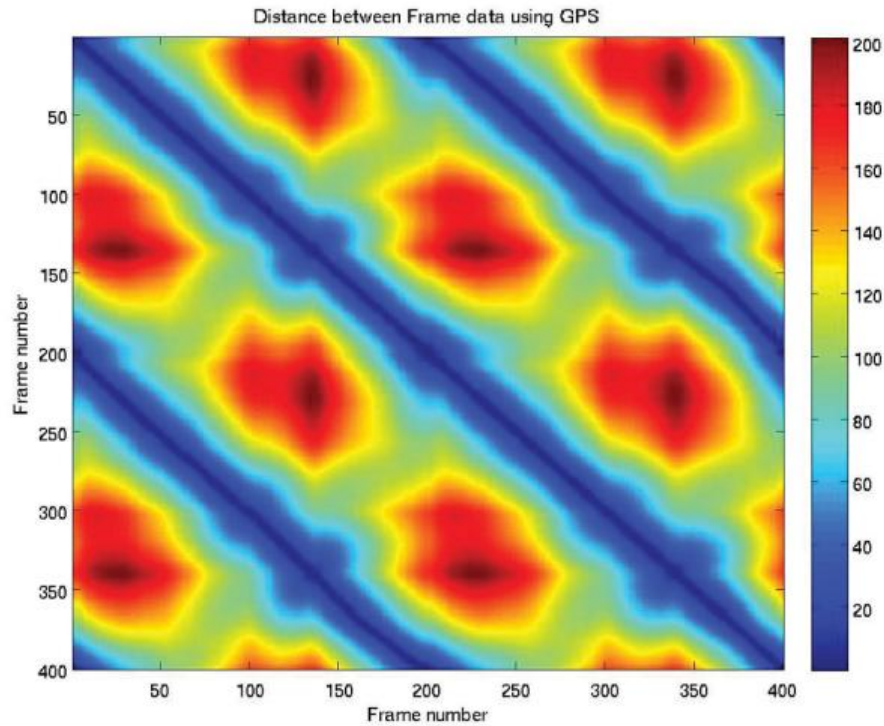
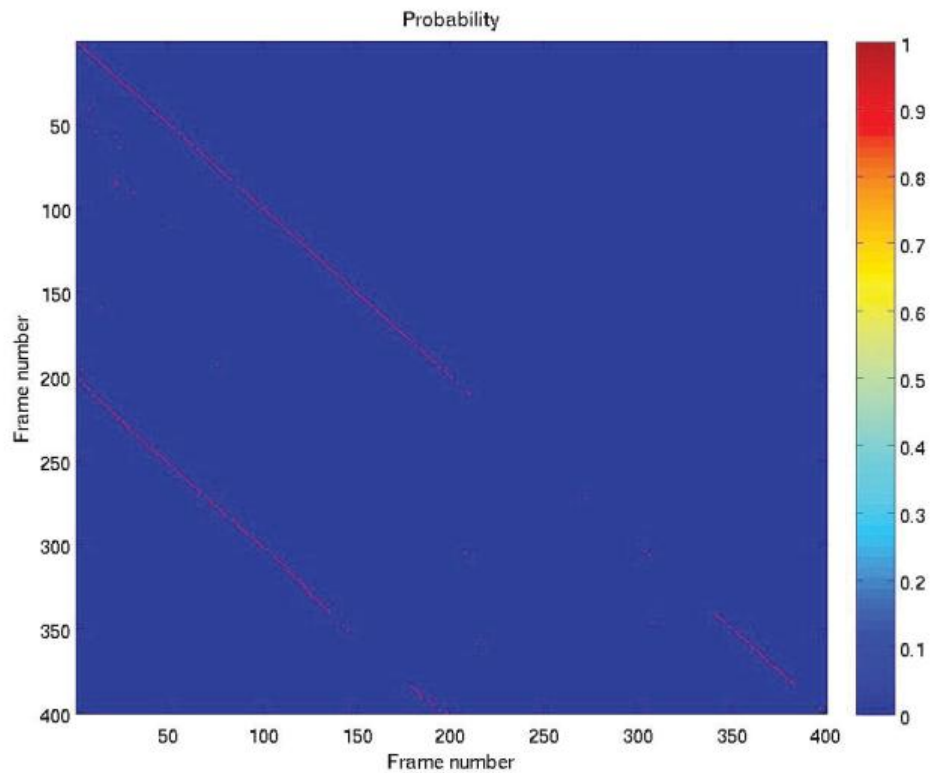**Figure 4-14   Ground Truth for Aug20 11am Dataset**

Figure 4-15   Probability Output for Aug20 11am Dataset
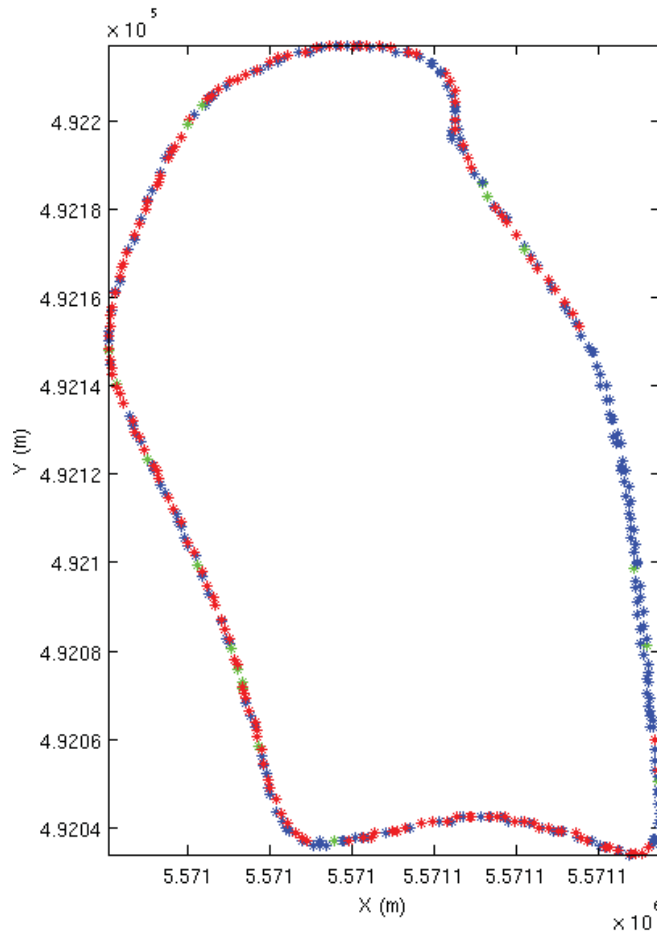
4-15

Figure 4-16   Trajectory and Loop Closure for Aug20 11am Dataset (99% Threshold)

Figure 4-17 shows the hourly place recognition recall rate percentages for imagery only, LIDAR only and both imagery and LIDAR. As the 3D structure is sparse in the rural environment, the recall rate for LIDAR is not that high. The imagery recall rate is much higher due to the feature-rich environment, but varies significantly as the illumination changes.

Table 4-6 shows the number of true positives for imagery only, LIDAR only and both with the OR integration scheme throughout the 12-hour period, comparing Phase 3 (2012) results with Phase 4 (2013) results. Figure 4-18 compares the multi-sensor recall rates between 2012 and 2013. It can be seen that the results have improved considerably, thanks to the Phase 4 enhancements.

When both the imagery and LIDAR data are used under the OR integration scheme, it offers the best of both worlds. Under favourable illumination, imagery place recognition provides very high recall rate (over 80%). Under adverse lighting conditions, the system still provides an adequate recall rate from the LIDAR place recognition (over 20%).
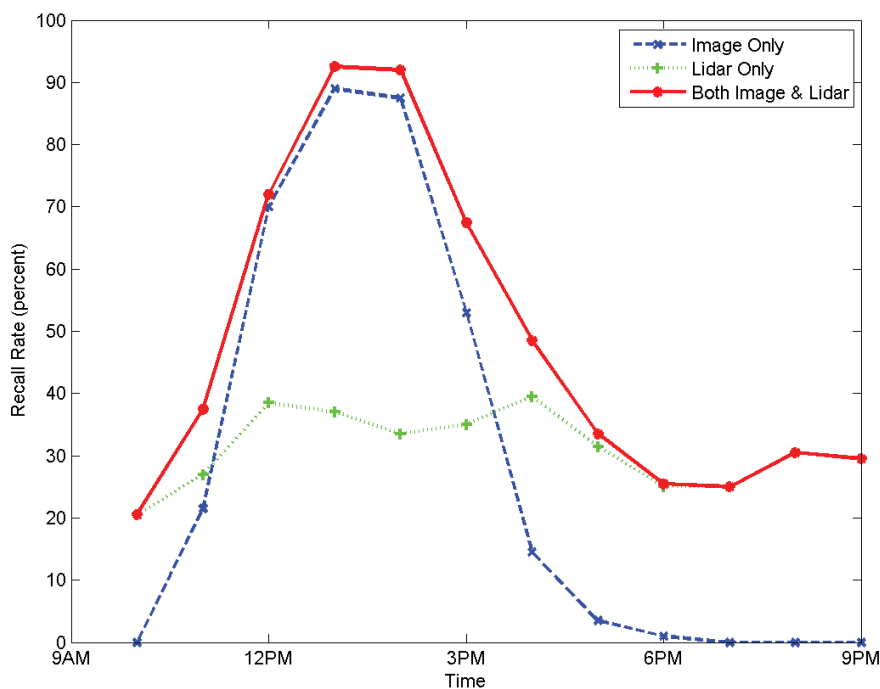
Figure 4-17   Recall Rates During 12-hour Period for Imagery only, LIDAR only and for Both

Table 4-6   Comparison for Variable Lighting Trials Results Between 2012 and 2013

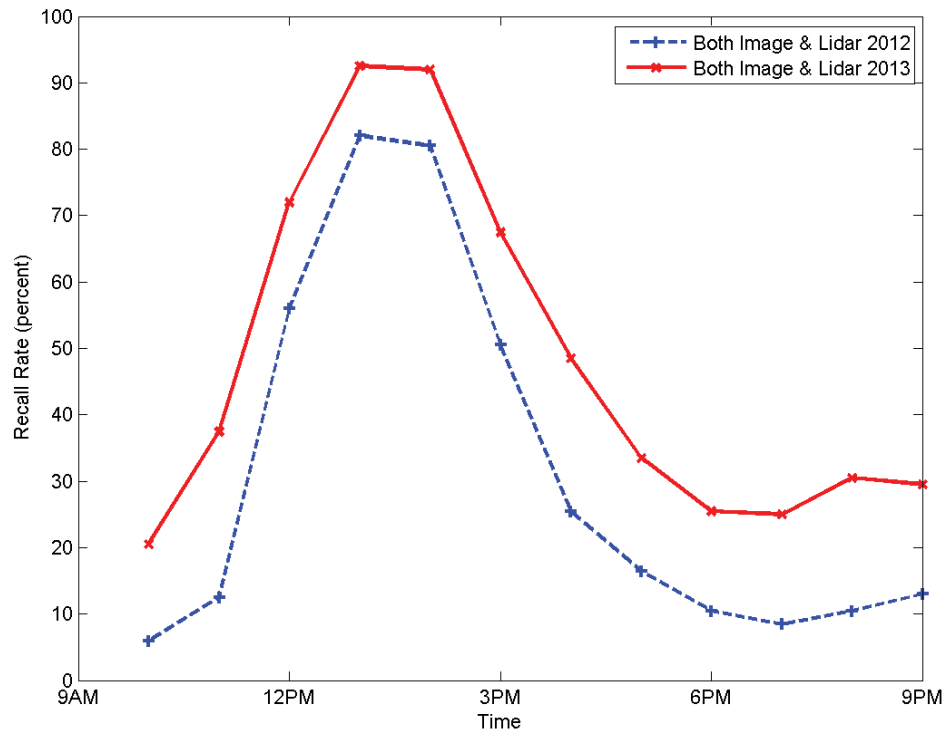| Time	2012 Results			2013 Results		
	Imagery: True Positives (False Alarms)	LIDAR: True Positives (False Alarms)	Imagery & LIDAR with OR Integration: True Positives (False Alarms)	Imagery: True Positives (False Alarms)	LIDAR: True Positives (False Alarms)	Imagery & LIDAR with OR Integration: True Positives (False Alarms)
9am	0 (0)	12 (0)	12 (0)	0 (0)	41 (0)	41 (0)
10am	20 (0)	14 (0)	25 (0)	43 (0)	54 (0)	75 (0)
11am	107 (0)	38 (0)	112 (0)	140 (0)	77 (1)	144 (0)
12pm	159 (0)	34 (0)	164 (0)	178 (0)	74 (0)	185 (0)
1pm	155 (0)	31 (0)	161 (0)	175 (0)	67 (1)	184 (0)
2pm	84 (0)	33 (0)	101 (0)	106 (0)	70 (0)	135 (0)
3pm	36 (0)	23 (0)	51 (0)	29 (0)	79 (0)	97 (0)
4pm	5 (0)	28 (0)	33 (0)	7 (0)	63 (1)	67 (1)
5pm	0 (0)	21 (0)	21 (0)	2 (0)	50 (0)	51 (0)
6pm	1 (0)	16 (0)	17 (0)	0 (0)	50 (0)	50 (0)
7pm	0 (0)	21 (0)	21 (0)	0 (0)	61 (0)	61 (0)
8pm	0 (0)	26 (0)	26 (0)	0 (0)	59 (2)	59 (2)
9pm	0 (0)	23 (0)	23 (0)	0 (0)	57 (0)	57 (0)

4-17

Figure 4-18   Comparison of Recall Rates During 12-hour Period for Multi-Sensor Integration

4.3.5   Asynchronous Dataset

For the asynchronous dataset, the video ASLAM and LIDAR ASLAM are run independently, where no multi-sensor integration is performed. As the sensor acquisition rate is higher than in synchronous datasets, there could be substantial overlap between consecutive frames. The processing is done first with keyframe detection disabled, and then done again with keyframe detection enabled.

4.3.5.1   Without Keyframe Detection

Aug9_Urb1 dataset traverses two large loops and consists of 2762 video frames and 2730 LIDAR frames. When keyframe detection is disabled, video ASLAM processes all the 2762 video frames while LIDAR ASLAM processes all the 2730 LIDAR frames.

The ground truth, probability and trajectory results for video ASLAM are shown in Figure 4-19, Figure 4-20 and Figure 4-21 respectively. The ground truth, probability and trajectory results for LIDAR ASLAM are shown in Figure 4-22, Figure 4-23 and Figure 4-24 respectively. Scenes are recognized successfully during the second loop.
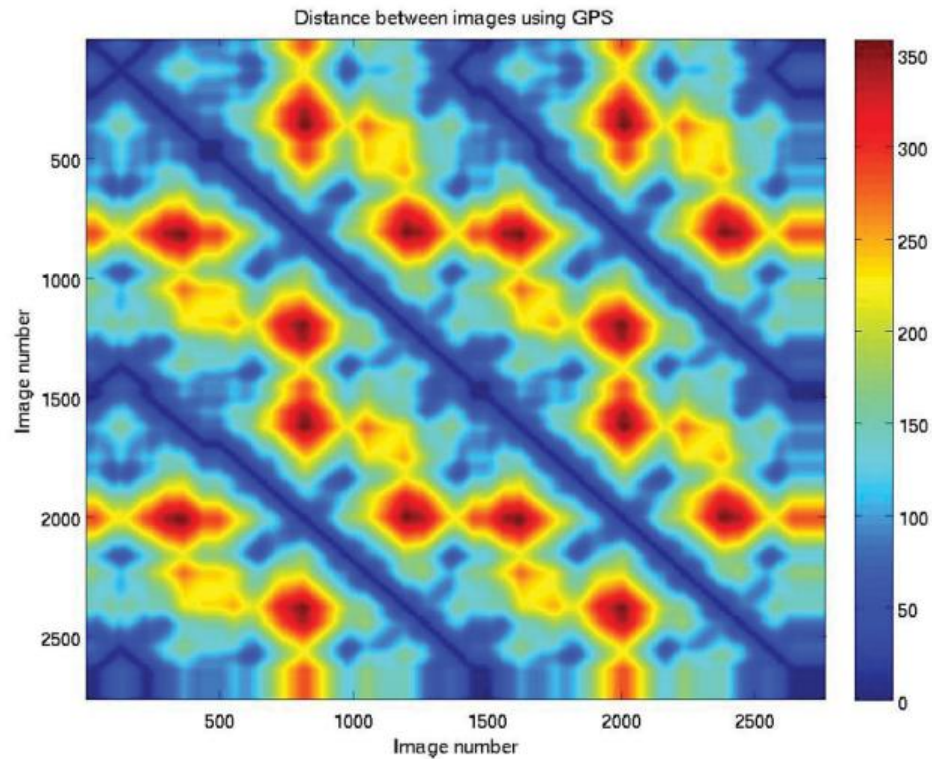
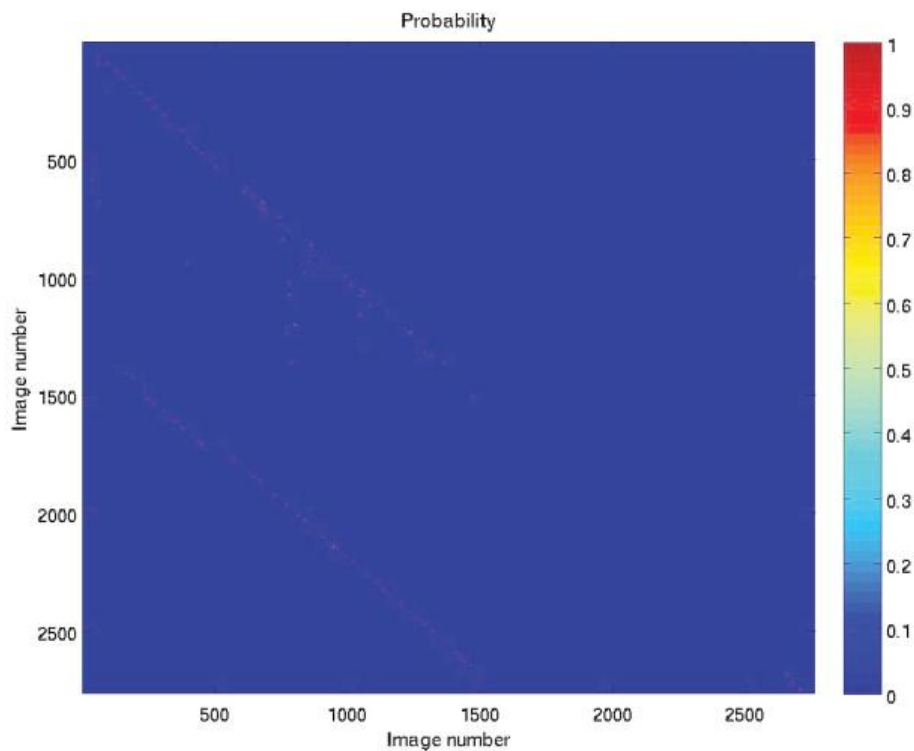**Figure 4-19   Ground Truth for Aug9_Urb1 Dataset with Video ASLAM (All Frames)**

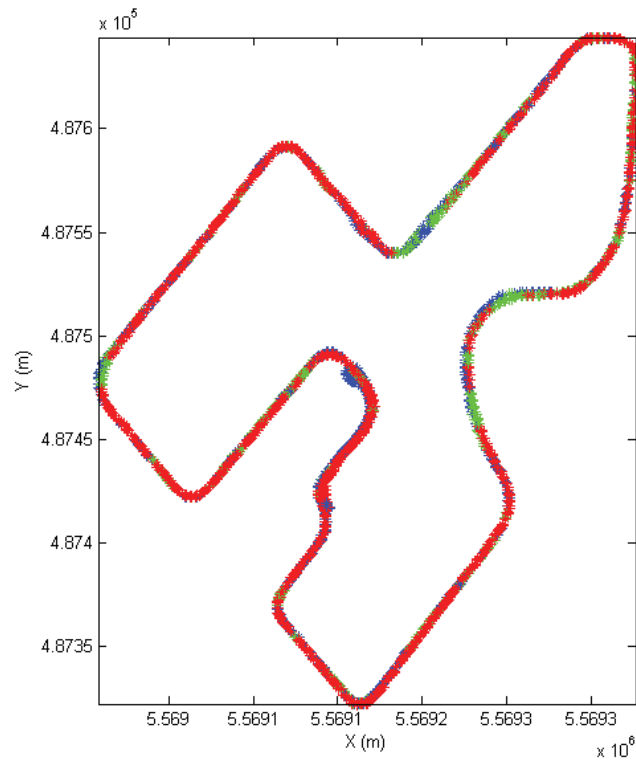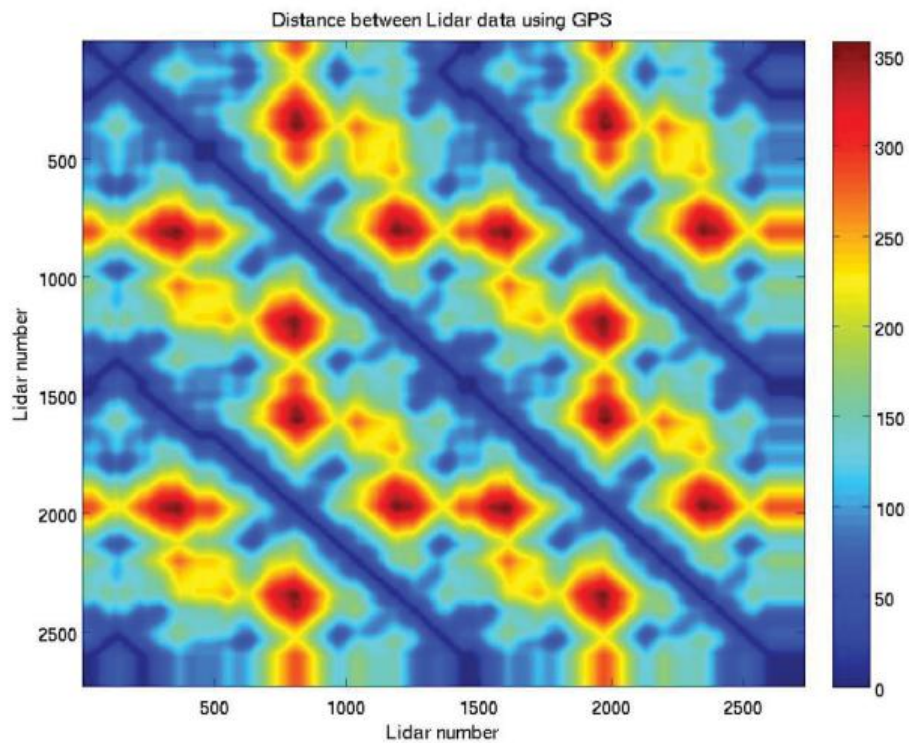**Figure 4-20   Probability Output for Aug9_Urb1 Dataset with Video ASLAM (All Frames)**

4-19

UNCLASSIFIED

Figure 4-21   Trajectory and Loop Closure for Aug9_Urb1 Dataset with Video ASLAM (All Frames)

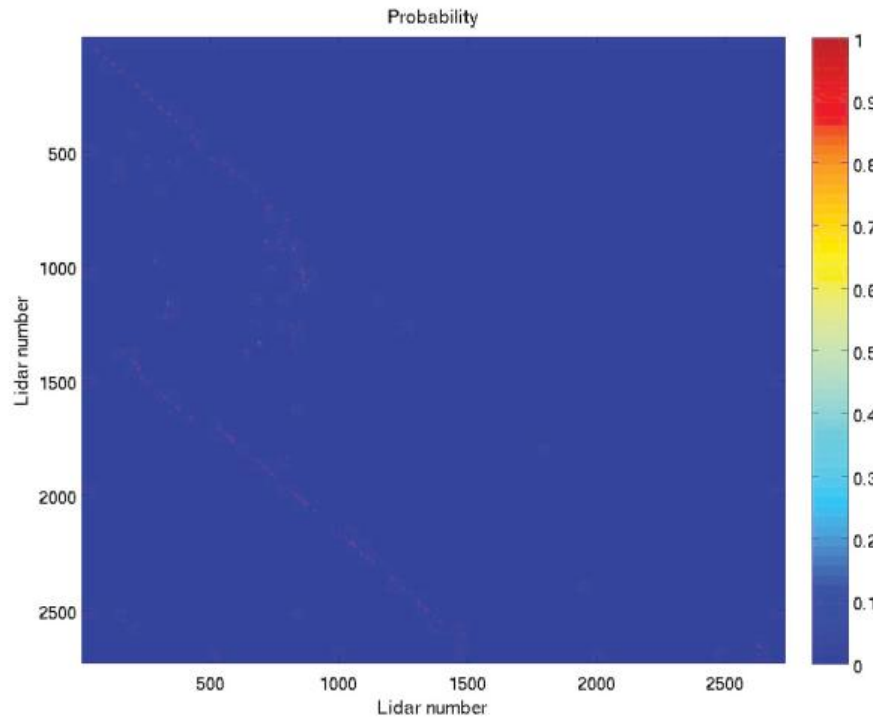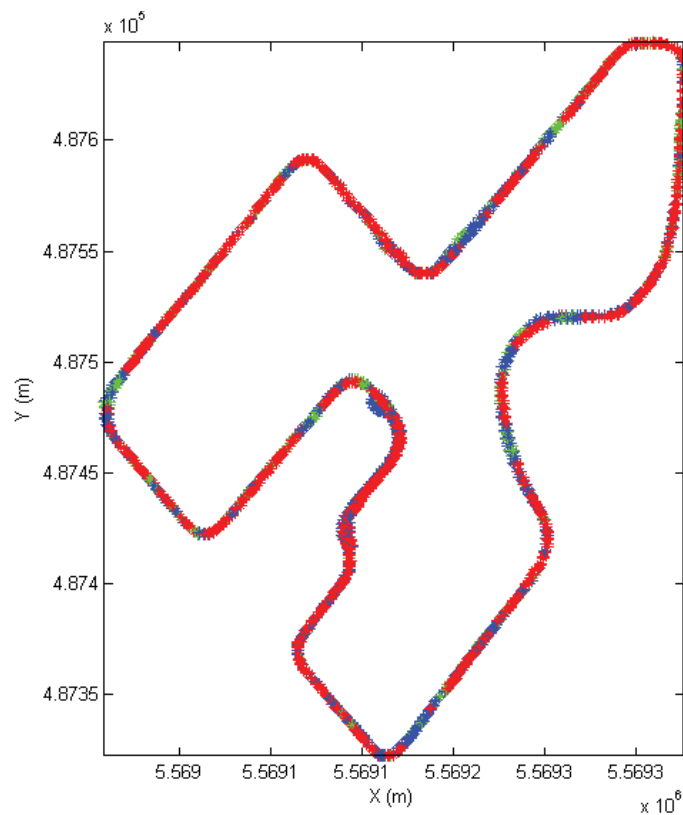**Figure 4-22   Ground Truth for Aug9_Urb1 Dataset with LIDAR ASLAM (All Frames)**

4-20

Figure 4-23   Probability Output for Aug9_Urb1 Dataset with LIDAR ASLAM (All Frames)

Figure 4-24   Trajectory and Loop Closure for Aug9_Urb1 Dataset with LIDAR ASLAM (All Frames)

4-21

4.3.5.2   With Keyframe Detection

At the default settings, the keyframe detection selects the current frame as a keyframe whenever there are fewer than 25% feature matches between the current frame and the last keyframe. Therefore, if the vehicle is stationary or has moved very little, it would not send the similar frames to the ASLAM API.

When keyframe detection is enabled, video ASLAM only processed 697 frames while LIDAR ASLAM only processed 1832 frames.

The ground truth, probability and trajectory results for video ASLAM are shown in Figure 4-25, Figure 4-26 and Figure 4-27 respectively. The ground truth, probability and trajectory results for LIDAR ASLAM are shown in Figure 4-28, Figure 4-29 and Figure 4-30 respectively.

With keyframe detection, only a quarter of video frames and two-thirds of LIDAR frames are processed, which helps reduce the processing time. As expected, the loop detection is sparser now, but consistent loop detection is still obtained throughout the second loop. The true positives and false positives comparison is shown in Table 4-7.
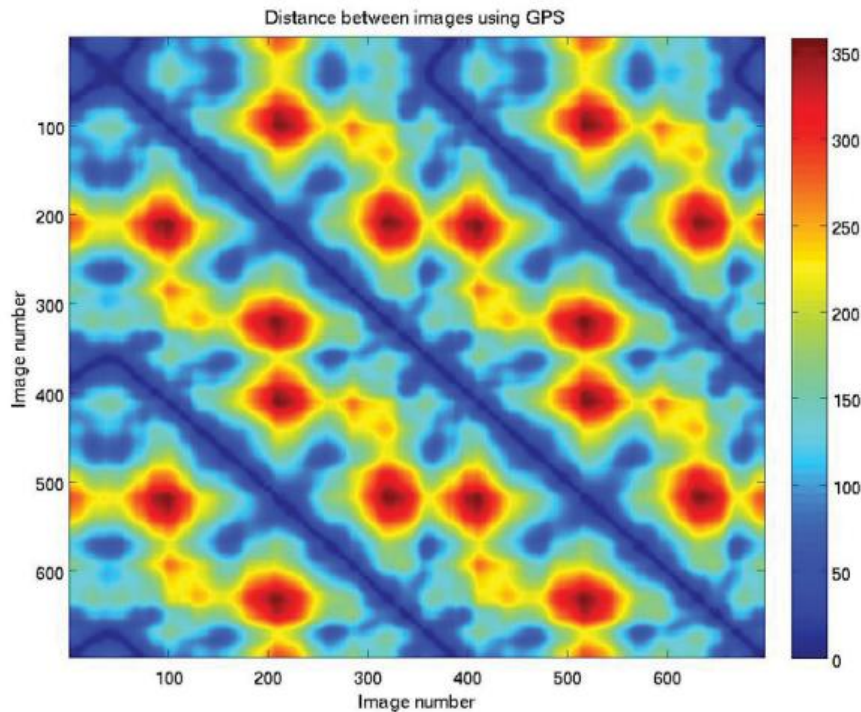

Figure 4-25   Ground Truth for Aug9_Urb1 Dataset with Video ASLAM (Keyframes)
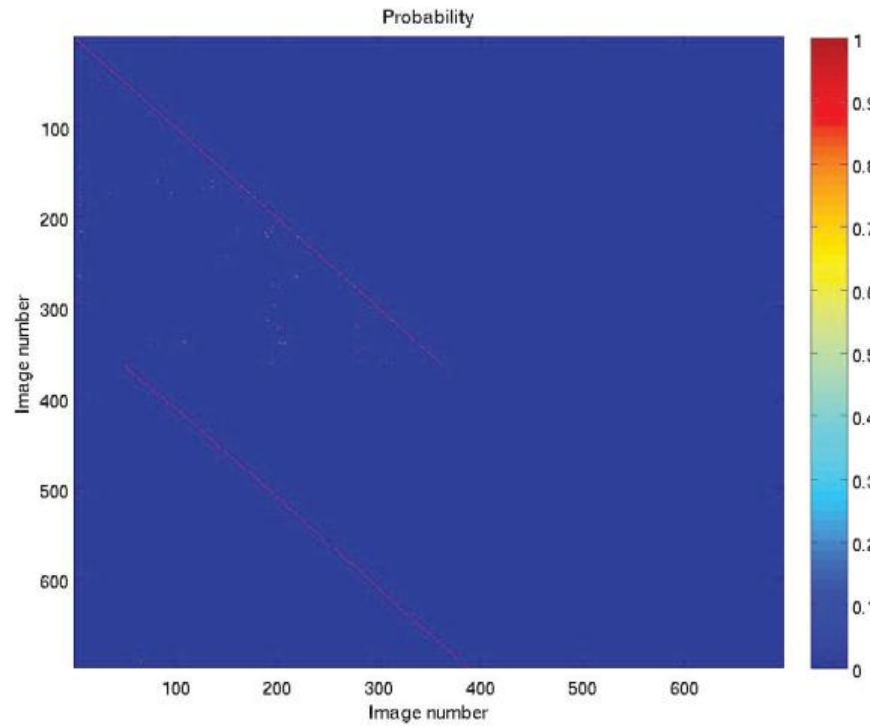
Figure 4-26   Probability Output for Aug9_Urb1 Dataset with Video ASLAM (Keyframes)
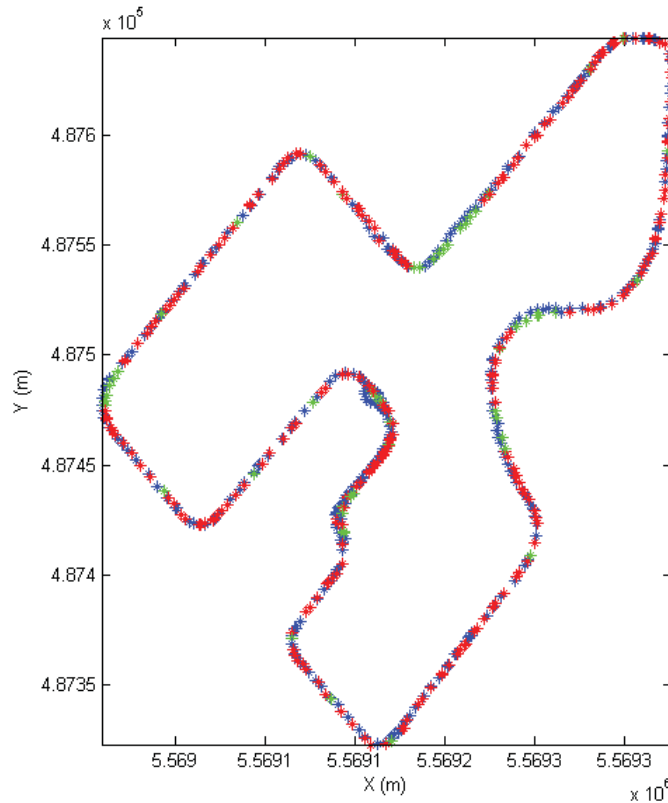

Figure 4-27   Trajectory and Loop Closure for Aug9_Urb1 Dataset with Video ASLAM (Keyframes)
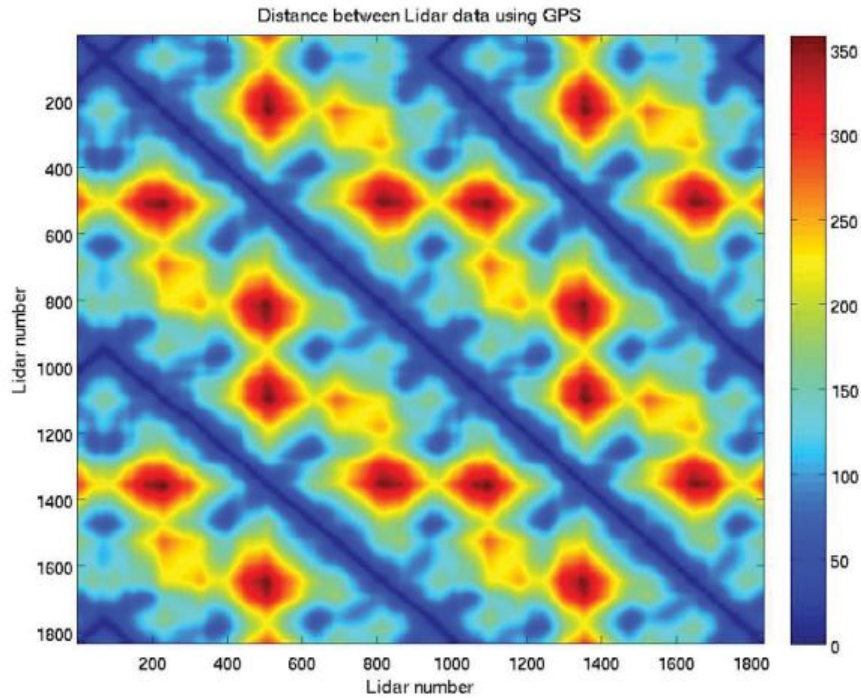
4-23

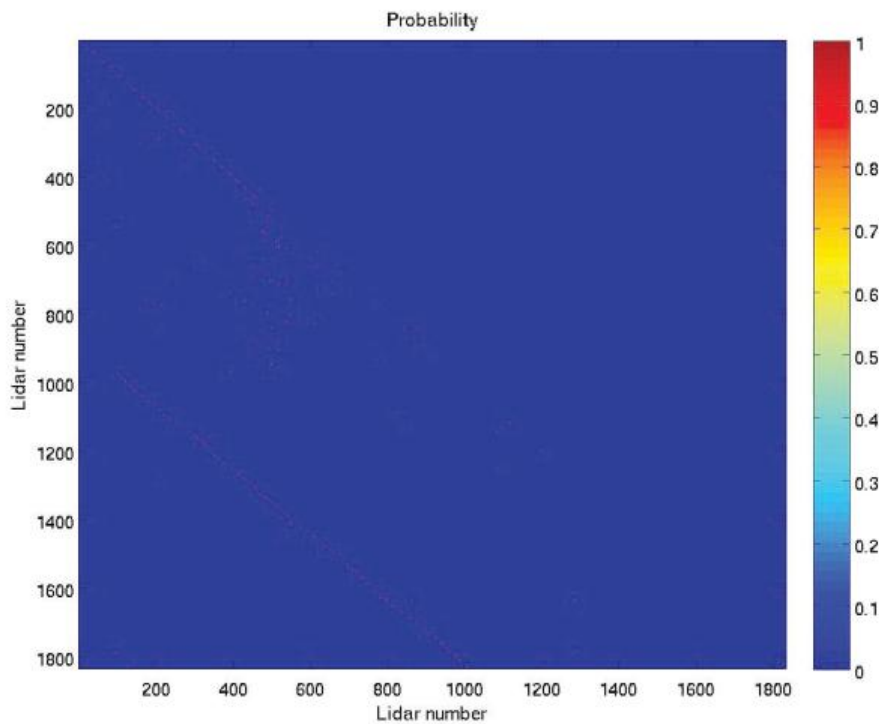**Figure 4-28   Ground Truth for Aug9_Urb1 Dataset with LIDAR ASLAM (Keyframes)**

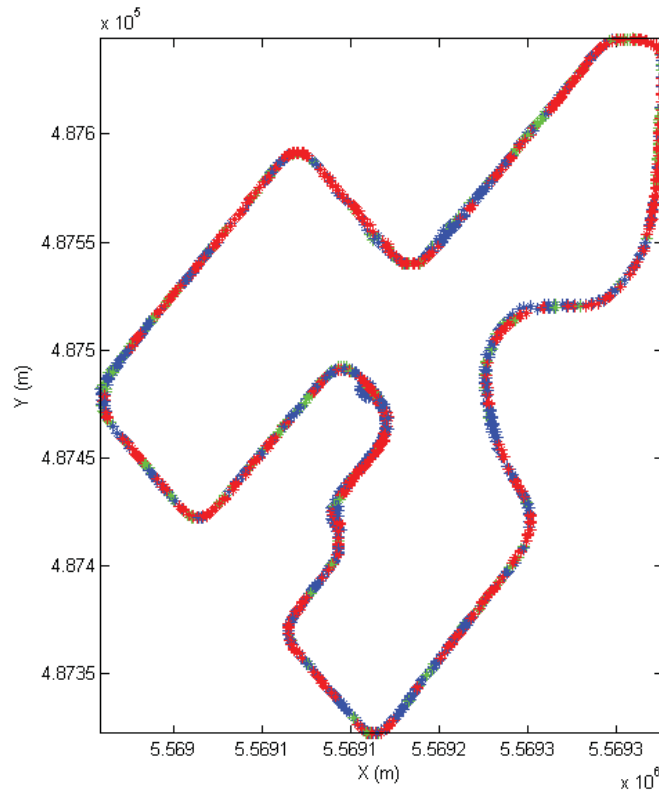**Figure 4-29   Probability Output for Aug9_Urb1 Dataset with LIDAR ASLAM (Keyframes)**

Figure 4-30   Trajectory and Loop Closure for Aug9_Urb1 Dataset with LIDAR ASLAM (Keyframes)

Table 4-7   Comparison for Aug9_Urb1 Trial Without and With Keyframe Detection

	Without Keyframe Detection		With Keyframe Detection	
	True Positives	**False Positives**	**True Positives**	**False Positives**
Video ASLAM	1250	0	268	0
LIDAR ASLAM	1138	5	589	3

4-25

THIS PAGE INTENTIONALLY LEFT BLANK

5 CONCLUSIONS

During the course of the project, we started with the initial video ASLAM and LIDAR ASLAM development in Phase 2, followed by the multi-sensor ASLAM with 6 DOF transform in Phase 3, and various improvements in Phase 4. The various field trials show that the developed system works well on various rural and urban datasets, with many true positives and very low false positives.

We demonstrated that the system works properly during the various field trials, meeting all the technical requirements in the RFP [R-1]:

1. Real-time Operation: Multi-sensor ASLAM runs at 1 Hz, while single-sensor ASLAM runs at higher than 1 Hz, thanks to the GPU implementation of SIFT and VD-LSD extraction and multi-threading.

2. Multi-sensor ASLAM: Multi-sensor ASLAM works properly with video and LIDAR sensors, showing that they complement each other well. DRDC agreed that RADAR needs not be considered.

3. Indoor/Outdoor: Extensive outdoor testing was performed for both urban and rural environments. DRDC agreed that indoor environment needs not be tested.

4. Day/Night: The system was tested with datasets from 9am to 9pm, including day and night scenarios. Video ASLAM works well only under optimal lighting conditions, while LIDAR ASLAM works irrespective of the ambient lighting.

5. Rotation Invariant: LIDAR ASLAM has recognized 180 degrees rotation, i.e. traversing in opposite directions, thanks to its 360 degrees field-of-view.

6. Operating System: All the software development and testing were done on an Ubuntu Linux computer.

7. Software API: A C++ ASLAM API has been developed, which allows integration into DRDC's architecture. A C++ test harness has been implemented to illustrate how to use the ASLAM API.

8. Matlab Interface: Matlab scripts have been developed to help with visualization, vocabulary generation, setting up folder and parameters automatically.

9. Geometric Loop Closing: A 6 DOF transform is estimated upon detecting a loop closure for both video and LIDAR ASLAM. The 6 DOF transform can be used by DRDC to correct the SLAM metric map in the future. This proves to be useful for validation, as it correctly rejects many FAB-MAP false alarms.

10. Continuous Operation: System has been tested to run for over 1 hour and has processed many large datasets, each of which traverses over several kilometres.

11. Database Storage: System can save the current database to file and then can re-start in continuous mode later by reading the database saved.

12. Area of Operation: The system has successfully processed a wide variety of urban and rural datasets, traversing previously unseen areas.

6 REFERENCES

R-1   Appearance Based SLAM Contract, W7702-115043/001/EDM, PWGSC, October 2010.

R-2   S. Se, ASLAM System Design Plan (Task 1.2), ASLAM-PL-53-1039, December 2010.

R-3   S. Se, ASLAM Field Trial Report (Task 2.4), ASLAM-RP-53-2411, October 2011.

R-4   S. Se, ASLAM Multi-Sensor Field Trial Report (Task 3.4), ASLAM-RP-53-4017, September 2012.

R-5   V. Kotamraju and S. Se, User Guide for ASLAM Recognition API, ASLAM-TN-53-2412 v1/4, March 2013.

R-6   A. Valizadeh and S. Se, User Guide for ASLAM Training and Visualization, ASLAM-TN-53-2413 v1/2, March 2013.

R-7   M. Ghazel and S. Se, ASLAM ROC Analysis Report (Task 4.4), ASLAM-RP-53-4754, March 2013.

R-8   J. Collier, Phase II ASLAM Field Trial Plan, August 2011.

R-9   H. Zhang, B. Li and D. Yang, Keyframe Detection for Appearance-Based Visual SLAM, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taiwan, October 2010.

THIS PAGE INTENTIONALLY LEFT BLANK

A   APPENDIX: ASLAM DELIVERY LISTINGS

The directory listings for the ASLAM delivery version 1.3 (March 22, 2013) are as follows:

```
code/
   bin/                     Pre-compiled binary executables
   lib/
      Release/             Pre-compiled release libraries
      Debug/               Pre-compiled debug libraries
   src/
      WordMakerMDA/        Modified WordMaker source code
      FabMapV2MDA/         Modified FabMapV2 source code
      GPU-LSD/             GPU LSD testing source code
      GPU-SIFT/            GPU SIFT testing source code
      ASLAM_TEST_HARNESS/  ASLAM test harness source code
      ASLAM/               ASLAM API source code
scripts/
   6DOF/                    Scripts related to 6 dof pose computation
   SetupFolderConfig/       Scripts to setup folder & config files
   VisualizationScripts/    Scripts for visualization of the results
   VocabularyGenerator/     Scripts to help generate vocabulary
doc/                        Various users guide documents pdf
config/
   Recognition/
      Image/               Sample recognition config files for image mode
      Range/               Sample recognition config files for range mode
   Training/
      Image/               Sample training config files for image mode
      Range/               Sample training config files for range mode
```

THIS PAGE INTENTIONALLY LEFT BLANK

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The objective of the Appearance-based Simultaneous Localization And Mapping (ASLAM) project is the research and development of an Appearance-Based Simultaneous Localization and Mapping system for day/night operations in indoor and outdoor environments. These algorithms would perform place recognition based on sensor data gathered from an Unmanned Ground Vehicle (UGV) as it travels through the environment. When the vehicle returns to a previously visited scene, the ASLAM algorithm would recognize the scene, update its internal representation, report this to the UGV, and finally provide information to aid in closing the loop with geometric Simultaneous Localization And Mapping (SLAM).

This final report describes the work done in this project, including the components of the developed system and the results from the various field trials.

L'objectif du présent contrat est la recherche et le développement d'un système de localisation et de cartographie en temps réel basé sur l'apparence pour les opérations menées de jour et de nuit, à l'intérieur comme à l'extérieur. Ces algorithmes doivent effectuer une reconnaissance de l'endroit basée sur les données recueillies par le capteur de l'UGV alors que celui-ci se déplace dans un environnement donné. Lorsque le véhicule revient sur une scène déjà visitée, l'algorithme ASLAM reconnaît la scène, met à jour sa représentation interne, la communique au UGV et, enfin, dispose d'un mécanisme pour fermer la boucle à l'aide du SLAM géométrique.

Le présent rapport final présente les travaux effectués, y compris les éléments du système conçu et les résultats de divers essais sur le terrain.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

ASLAM; UGV; SLAM; Appearance-based Simultaneous Localization And Mapping; Unmanned Ground Vehicle; Mapping system