Multi-Sensor Appearance-Based Place Recognition

Jack Collier
DRDC Suffield
Box 4000, Station Main
Medicine Hat, Alberta, T1A 8K6, Canada
Email: jack.collier@drdc-rddc.gc.ca

Stephen Se
MDA Systems Ltd.
13800 Commerce Parkway,
Richmond, B.C., V6V 2J3, Canada
Email: sse@mdacorporation.com

Vinay Kotamraju
MDA Systems Ltd.
13800 Commerce Parkway
Richmond, B.C., V6V 2J3, Canada
Email: vinay.kotamraju@gmail.com

***Abstract*—This paper describes a multi-sensor appearance-based place recognition system suitable for robotic mapping. Unlike systems that extract features from visual imagery only, here we apply the well known Bag-of-Words approach to features extracted from both visual and range sensors. By applying this technique to both sensor streams simultaneously we can overcome the deficiencies of each individual sensor. We show that lidar-based place recognition using a generative model learnt from Variable Dimensional Local Shape Descriptors can be used to perform place recognition regardless of lighting conditions or large changes in orientation, including traversing loops backward. Likewise, we are still able to exploit the feature rich place recognition that visual systems provide. Using a pose verification system we are able to effectively discard false positive loop detections. We present experimental results that highlight the strength of our approach and investigate alternative techniques for combining the results from the individual sensor streams. The multi-sensor approach enables the two sensors to complement each other well in large urban and rural environments under variable lighting conditions.**

I. Introduction

In this paper we examine the problem of recognizing previously visited places based on both visual and range sensor data. This issue is essential to the large-scale Simultaneous Localisation and Mapping (SLAM) problem. SLAM techniques are subject to failure, especially over large areas, due to linearization errors and improper data associations [1]. To overcome these problems, appearance-based place recognition can be used to decouple the loop closing process from the geometric SLAM algorithm[2],[3],[4]. Here, the loop closing process is essentially a scene recognition computer vision task. In this way, loop recognition is not bound to the pose estimates of the SLAM procedure and hence is not prone to failure resulting from gross errors in the robot pose estimate.

State-of-the-art place recognition systems detect loop closures by measuring the similarity between features extracted from the sensory data. Visual sensors are the predominant sensing modality used as they exploit the feature rich nature of their output. Visual features such as SIFT [5] and SURF [6] work well under small variations in lighting and orientation, but are prone to failure under large variations. Conversely, lidar sensors are very robust to variations in lighting and orientation, but rely on the presence of 3D structures in the scene. In this work, we seek to combine the strengths of the two sensing modalities to improve the robustness of the place recognition system and create a system that can perform in both day and night conditions. This is an essential element to autonomous UGV operation when GPS availability is not guaranteed due to occlusion or jamming.

Our essential approach is to perform visual and lidar place recognition in parallel and then combine the results to achieve better recognition. Current techniques that exploit both visual and range data use the range data for verification purposes only and do not perform recognition directly on the range data. To perform recognition on range data, we extend the use of traditional Bag-of-Words (BoW) techniques by incorporating a shape descriptor called the Variable-Dimensional Local Shape Descriptor (VD-LSD) [7]. This descriptor has been shown to be highly distinctive, robust to orientation and is compact in storage [8]. We examine different methods of combining the visual and lidar-based place recognition results.

II. Related Work

The use of visual features for place recognition has received recent attention from the robotics community. Popular techniques grew out of content-based image retrieval systems used by popular search engines [9]. The basis of this approach is the Bag-Of-Words (BoW) technique that clusters similar features extracted from training imagery to form a vocabulary of words representing the environment. New imagery can be categorized by matching its features to the vocabulary. The resulting appearance vector can be used to compare the image to previously seen images to determine if the image is from a new or previously visited place.

In [4], image similarity is defined as the normalized inner product between the image appearance vectors. A similarity matrix is constructed and used to search for repeated sequences. In GraphSLAM [1], the authors used a similar BoW technique but were able to update their vocabulary as each new image was processed rather than through off-line supervised learning. FAB-MAP [3] learns a generative model for the BoW data that captures the tendency of combinations of appearance words to co-occur reflecting that they are generated by common objects in the environment. By learning a Bayesian model of these common objects in an unsupervised way, FAB-MAP improves the inference mechanism by robustly reasoning about which set of features are likely to appear or disappear together. Probability distributions are efficiently computed using Chow-Liu trees. To achieve an efficient large-scale implementation, FAB-MAP 2.0 [10] uses an inverted

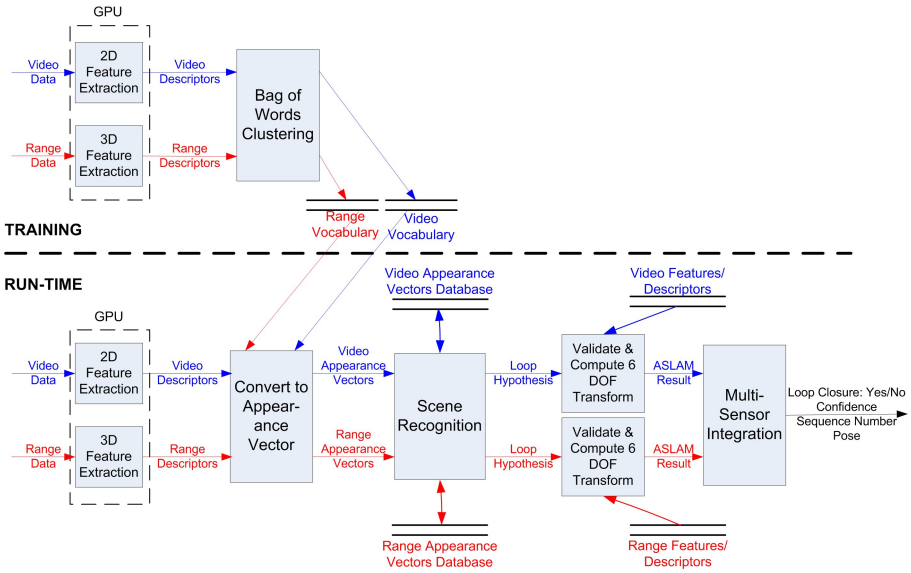

Fig. 1. The proposed system is a multi-sensor extension of the FAB-MAP algorithm. SIFT and Variable Dimensional Local Shape Descriptor features are used to train a vocabulary for video and lidar sensors respectively. During run-time, features are extracted from real-time sensor data, converted to their respective appearance vectors and used to detect loop closures. The outputs from the parallel loop detectors are then input to a validation and pose computation algorithm before being integrated to generate a final loop detection.

index retrieval architecture that requires modification to the way probabilities of revisits were computed and updated in the original algorithm. FAB-MAP 2.0 has been demonstrated on a large 1000km dataset with few false positive results.

Place recognition has also been applied to 3D range data. In [11], the authors extracted Normal-Aligned Radial features [12] from a range image and create a BoW vocabulary from which similarities can be measured between scans. In Magnusson et al. [2], the authors extract Normal Distribution Transform (NDT) features from a point cloud and match a global appearance descriptor derived from the NDT. In [13], statistical and range histogram features are used as input to an AdaBoost classifier.

Algorithms that use multiple sensors for place recognition have also been proposed. In [14], the authors augmented video-based FAB-MAP with 3D lidar data that was used to capture the spatial distribution of visual words in the scene. Lidar data was used to reject ambiguous loops but is not used to perform the recognition. In [15], authors propose a place recognition system that combines features extracted from both visible and thermal imagery. Recognition is performed using FAB-MAP. Results show an increased performance of the combined system across a variety of lighting conditions.

III. System Overview

The system presented here is a multi-sensor extension of the FAB-MAP system described in [3], [10]. The system is outlined in Figure 1. The algorithm is run in parallel for both video-based FAB-MAP (using SIFT features) and range-based FAB-MAP (using VD-LSD). Separate place recognition results are generated from both sensor streams and integrated to provide a final loop closure result.

The system consists of two phases, training and real-time loop detection. In the training phase, representative sensor data is input to the feature extraction algorithm. Bag-of-Words K-means clustering is performed in the feature space to group similar features and create the vocabulary. As the system is sensor parallel, separate vocabularies are generated from the range and video data.

During run-time new features are extracted from data, as shown in Figure 2. A probabilistic detector model converts the extracted features into an appearance vector that determines which words are present in the current scene. For a vocabulary of n words the associated appearance vector is denoted by $Z_k = z_1, \ldots, z_n$, where the binary z value 1 indicates the presence of the word in the current frame. The appearance vectors are then fed to the scene recognition module where the current scene is compared to previous scenes and a loop is determined using a Bayesian model. The image with the highest probability above a threshold is determined to be a tentative loop closure. As before, scene recognition is performed separately on each sensor stream.

If a tentative loop closure has been determined, the two frames are then input to a validation and 6-DOF transform computation module. If the detected loop is a true loop then a valid transform can be computed and used to perform geometric loop closing. In the case of a false positive loop detection, the 6-DOF module will detect the geometric inconsistency and discard the loop hypothesis.

Once the results have been validated and a transform between the two frames is calculated the outputs from the two processing streams are combined. Two different strategies discussed in section III-C are investigated to perform the multi-sensor integration.

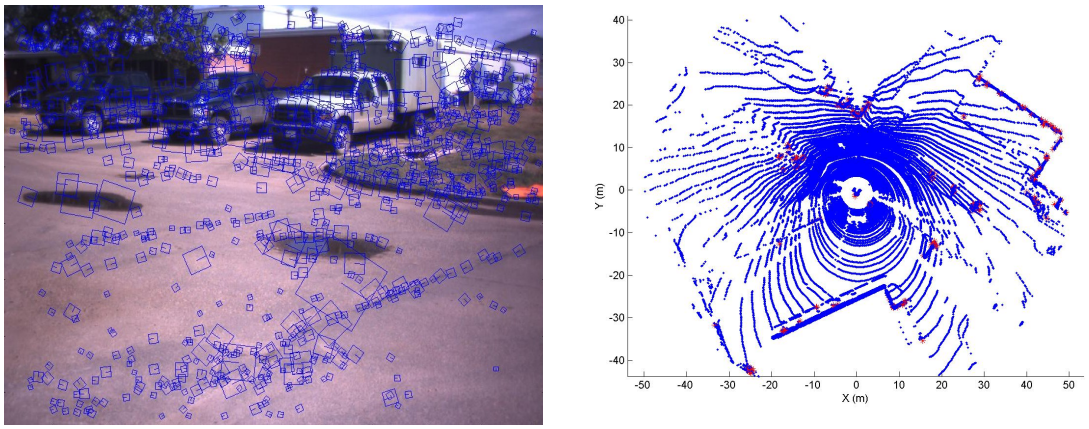

Fig. 2. Example of features extracted from a scene with buildings and parked vehicles from both video imagery (left) and range imagery (right). Note the difference in scale and field of view for both sensors as well as the sparseness of the features in the range imagery compared with the video image.

A. Variable-Dimensional Local Shape Descriptor

As previously mentioned, the principal input to the range place recognition algorithm is the Variable-Dimensional Local Shape Descriptor described in Taati et al. [7]. The VD-LSD has previously been used for model-based object recognition and pose estimation of range data. Since this is a novel aspect of this paper, we discuss it further here.

The VD-LSD for a salient point is generated from invariant properties extracted from the Principal Component Space of the local neighborhood. The 3x3 covariance matrix of the local neighborhood around each point is computed and their eigenvalue decomposition is used to associate an orthonormal frame and the three eigenvalue scalars. Using these vectors and scalars, seven positional, nine rotational and six dispersion properties for all points can be generated to form a variety of histograms that carry various levels of information about the local geometry.

One distinguishing feature of the VD-LSD is that it offers a generalized form that subsumes a large class of popular descriptors in literature, such as Spin Images [16], Point Signatures [17] and Tensor Correspondence [18]. These descriptors lie in very small dimensional subspaces spanned by the at most 22-dimensional VD-LSD. In taking such a maximalist approach, VD-LSD captures local geometry of the range data better and therefore achieves better robustness in the matching phase.

Some of these properties might perform more robustly than others. The descriptive power of these properties is not equal and depends on the local geometries. Therefore, the optimal property subset could be different for different models. In applying VD-LSD to our place recognition application consisting of a variety of real-world objects, the desired subset of properties was empirically determined based on the trade-off between the time required to compute these properties for large point clouds and their robustness to place recognition.

The VD-LSD extraction algorithm is parallel and therefore is highly suitable for Graphics Processing Unit (GPU) implementation. Several processor-intensive modules have been implemented in CUDA to run on NVIDIA GPU, to offload the CPU. For this work, a total of six dimensions have been chosen, including three positional and three dispersion properties. The number of histogram bins along each dimension is set to two, thereby producing descriptors of length 64. The number of bins and which dimensions to use can be configured in the software.

B. Validation and Transform Computation

After our system has identified a potential loop between frames, a validation scheme is used to discard false positives. For lidar data, the 3D VD-LSD for the two frames are considered, whereas for visual imagery, the SIFT 3D features are used as the system collects stereo imagery data.

A ratio test is first applied on the Euclidean distance between the feature vectors for the best and second best matches to discard ambiguous matches. Once a set of tentative feature matches are obtained, a RANSAC approach [19] is performed to remove the outliers. Three tentative feature matches between the candidate frames are selected randomly to compute the 6-DOF transform using the rigid body transformation approach, to serve as a hypothesis. The number of supporting feature matches from the tentative set are obtained for each hypothesis. This process is performed repeatedly and the hypothesis with the highest number of supporting feature matches is selected. All the supporting feature matches are used to compute the 6-DOF transform using a least-squares minimization approach.

Since the transformation between frames is constrained by vehicle geometry and ground vehicle motion, an additional level of validation can be performed to discount for tentative loops with large translations and excessive pitch and roll between the candidate frames. After validation the final 6-DOF transform and the loop hypothesis are sent to the integration module.

C. Integration Schemes

Two different integration strategies, each with their own strengths, were implemented for this work. The AND scheme

only outputs a loop closure if both sensors detect a loop closure at the same location. In this case, the resulting 6-DOF transformation will be the average of the transformations computed separately from the lidar and range data. Conversely, the OR scheme outputs a valid loop closure if either sensor detects a loop. In the case where both processing streams detect the closure the OR scheme chooses the loop hypothesis with the higher probability and uses its associated transform. Both strategies have their merits. With the AND scheme outputting a final valid loop only if both sensors agree, the potential for false positive loop closures is minimized. However, as the visual image sensor relies on ambient lighting, the AND scheme will not detect any loops in low light conditions, a key goal of the system. Conversely, the OR scheme will detect more loops, but may be more susceptible to false positives. It will also detect loops during night operations, as lidar does not require ambient lighting.

After successful detection, validation, and integration a valid loop and its associated transformation can be used to close the loop in a geometric SLAM formulation.

IV. Experimental Results

This section presents results from field trials conducted over a two week period at the DRDC Experimental Proving Ground (EPG) at CFB Suffield, Alberta, Canada. Field trials were concentrated on two separate environments on the EPG, a representative urban environment with many buildings, vehicles, pavements, etc. and a representative rural environment containing prairie grass, gravel roads and a sparse population of buildings. Trials consisted of manually driving the vehicle for several kilometres while collecting sensor data (lidar, stereo, and GPS) which was then processed offline to perform place recognition. It should be noted that the entire system can run online at around 1Hz, but we chose to process the data offline for convenience purposes.

A. Hardware and System Configuration

All field trials were performed on a research variant of the Multi-Agent Tactical Sentry (MATS) (Figure 3), a modified Kawasaki Mule developed at DRDC Suffield. The platform was retrofitted with a Velodyne HD lidar (360 degrees field-of-view) and Point Grey XB3 stereo camera (45 degrees field-of-view). Only left camera images are used as input to the FAB-MAP engine, while SIFT features from the left and right cameras are used in the pose verification and 6-DOF transformation stages. Differential GPS data was collected during trials and used for verification and display purposes only and was not used by the place recognition algorithm. All software was tested on a high performance server housed on the MATS vehicle. The server included a 3.33GHz Hexa-Core Intel Core i7 Extreme Processor, 12GB of memory and a NVIDIA Geforce GTX 580 GPU used for feature detections and descriptor calculation.

From experimental observation, visual features or lidar features from the ground may confuse the loop detection. Therefore, the VD-LSD extraction is configured to ignore features from planar surfaces, to avoid VD-LSDs arising from the ground. Similarly, the SIFT extraction is configured to ignore the smallest-scale features, to avoid SIFT arising from gravel on the ground.

Vocabularies were generated by subsampling data sets collected from similar rural and urban environments. Representatives rural data was used to generated the rural vocabulary while representative urban data was used to generate the urban vocabulary.

Fig. 3. Research platform used to conduct the tests. The Velodyne lidar and the Point Grey Research stereo camera housed on top of the vehicle were used for all tests.

B. Rural Trials

In this trial, the vehicle traversed 5.1km of a large outdoor environment, including several loops for portions of the trajectory. Using the OR integration scheme with pose verification, we obtain many true positives and no false positives. While FAB-MAP sometimes reported false positives, they were correctly rejected by the pose verification step. Figure 4 shows the trajectory of this trial, where the positional information was provided by the GPS. The blue dots 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 pose verification.

As 3D structure is sparse in the rural environment, a larger proportion of the loop detections come from the imagery data. However, Figure 5 shows a true positive loop where both visual and lidar FAB-MAP detected a loop, while Figure 6 shows an example where visual FAB-MAP does not detect a loop but lidar FAB-MAP does. This is likely due to the large lighting variation in the images. With the multi-sensor approach, the place recognition recall rate is improved over using the individual sensors.

Figure 7 shows a successful lidar loop detection where the vehicle traversed the loop in the opposite direction. Since the stereo camera has a 45 degree Field of View (FOV), it could not recognize the scene. But with the 360 degree FOV, the lidar place recognition could correctly recognize the loop and also provide a pose estimate of (X=1m, Y=-0.35m, Z=2.77m, Rx=-

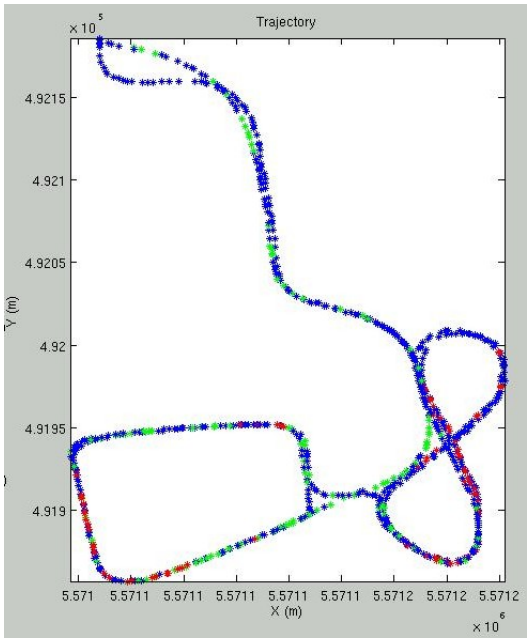

Fig. 4. Trajectory for a rural trial: blue indicates no loop detection, red indicates successful loop detection while green indicates loop detection rejected by pose verification.

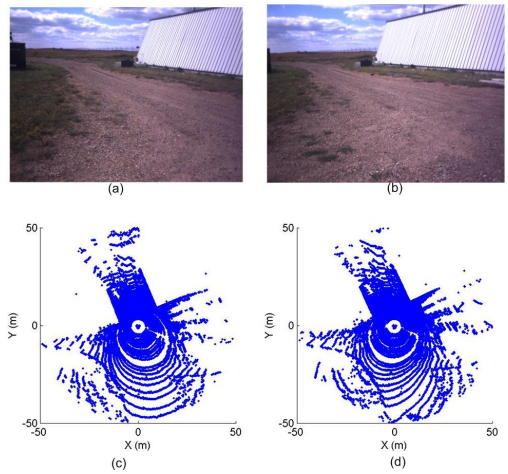

Fig. 5. A true positive in the rural environment where the place is recognized by both visual and lidar. The image pair is shown in (a) and (b) while the corresponding top-down view of the point cloud is shown in (c) and (d) respectively.

4.7deg, Ry=-161.3deg, Rz=1.2deg). This shows the benefit of using a 360 degrees FOV lidar or an omni-directional camera.

Figure 8 shows a false positive detection from video due to the far field cloud pattern similarity. The loop is rejected correctly by the pose verification process. It can be seen visually that the scale is quite different, i.e. one image is taken closer to the building than the other. The GPS distance between these two frames is around 80 metres, thereby confirming that this false positive is indeed rejected correctly. This scenario occurs frequently in this data set as there is an abundance of

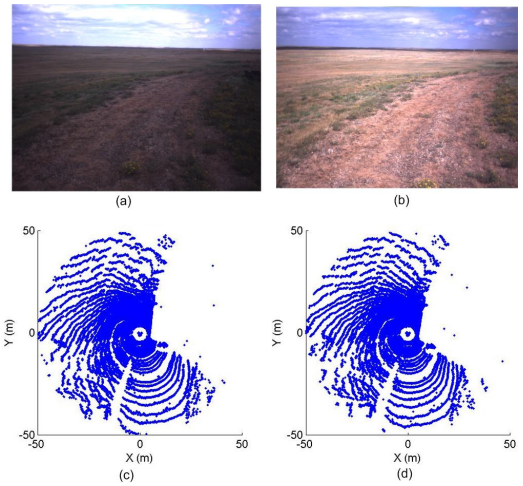

Fig. 6. A true positive in the rural environment where the loop is detected by lidar only due to the large lighting variation. The image pair is shown in (a) and (b) while the corresponding top-down view of the point cloud is shown in (c) and (d) respectively.

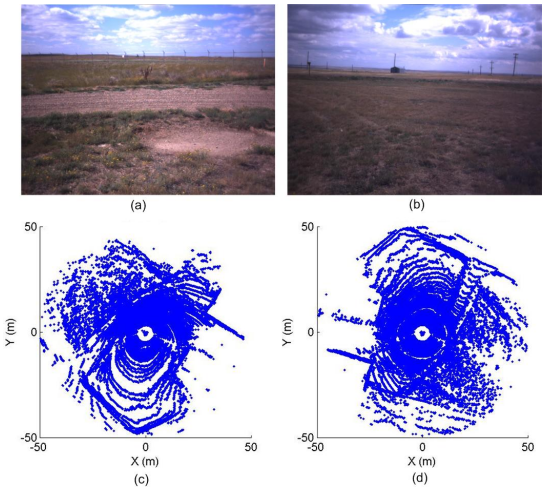

Fig. 7. A true positive from lidar data when the vehicle traverses in the opposite direction. It can be seen that the top-down view of the point cloud is rotated around 180 degrees relatively to each other but resembles each other well, while the images look quite different as they are looking at different directions.

clouds in the visual imagery. In all cases, the pose verification algorithm has discarded them correctly.

Using the AND integration scheme, the number of true positives decreases as expected, as it requires video and lidar to recognize at the same frame. Since our pose verification is very effective in discarding false positives from FAB-MAP, we already obtain zero false positives using the OR integration scheme for this experiment. More details of the results are shown in Table I.

C. Urban Trials

In this trial, the vehicle traversed a large path in an outdoor environment (5.8km), including several loops for portions of the trajectory. Using the OR integration scheme with the pose

Fig. 8. A false positive from imagery data due to the similar cloud pattern but has been rejected by the pose verification process correctly.

TABLE I
TRUE POSITIVES AND FALSE POSITIVES FROM INDIVIDUAL SENSORS AND FROM MULTI-SENSOR APPROACH USING THE TWO INTEGRATION STRATEGIES FOR THE RURAL TRIAL.

	Video	Lidar	OR	AND
True Positives	60	33	86	7
False Positives	0	0	0	0

verification, we obtain many true positives and zero false positives. While FAB-MAP sometimes reports false positives, they have been correctly rejected by our pose verification step. Figure 9 shows the trajectory of this trial, where positional information is provided by the GPS[1].

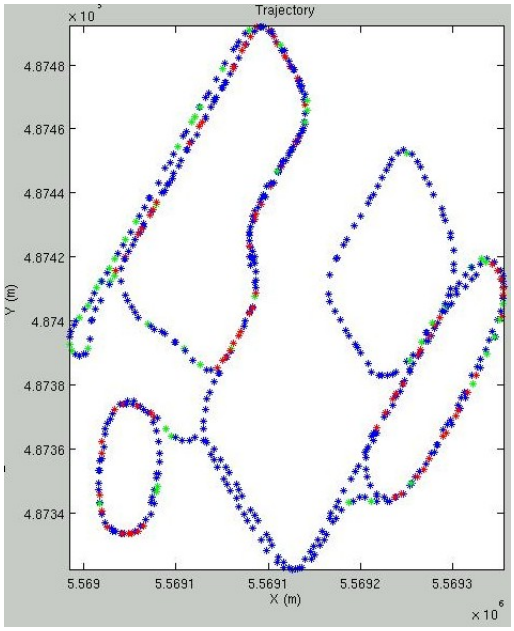

Fig. 9. Trajectory for an urban trial: blue indicates no loop detection, red indicates successful loop detection while green indicates loop detection rejected by pose verification.

Figures 10 and 11 shows some examples of true positives, in which the top-down view of the point clouds show that there are more 3D structures in the urban scene, which would facilitate loop detection using lidar data. Using the OR integration scheme, a total of 112 true positives have been detected and the average GPS distance between the two frames is 1.6m and all are within 6.5m, indicating that all the detected loops are correct. More details of the results are shown in Table II.

[1]A video showing the output from this trial has been included with the submission

TABLE II
TRUE POSITIVES AND FALSE POSITIVES FROM INDIVIDUAL SENSORS AND FROM MULTI-SENSOR APPROACH USING THE TWO INTEGRATION STRATEGIES FOR THE URBAN TRIAL.

	Video	Lidar	OR	AND
True Positives	107	30	112	25
False Positives	0	0	0	0

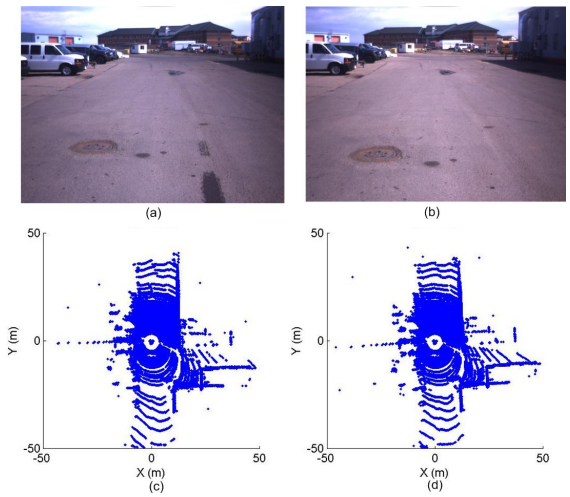

Fig. 10. A true positive in the urban environment where the place is recognized by both visual and lidar. The image pair is shown in (a) and (b) while the corresponding top-down view of the point cloud is shown in (c) and (d) respectively.

D. Variable Lighting Trials

In this trial we evaluate the effectiveness of the multi-sensor approach for 24-hour vehicle operation. Over a 12 hour period (0900-2100), the vehicle traversed the same loop of a rural environment collecting data. Using the 1200 test set as the initial loop we then processed each additional hourly dataset as the second loop. By traversing the same loop we are able to isolate the effect of illumination and easily calculate the recall rate.

Figure 12 shows images at a similar location at different times, i.e. 9am, 1pm, 5pm and 9pm, while Figure 13 shows the corresponding lidar data. It can be seen that the imagery varies considerably due to the illumination changes while the lidar data is stable throughout the day.

Figure 14 shows the hourly place recognition recall rates for video only, lidar only and both video and lidar. All recall rates are at 100% precision (no false positives). As this is a rural environment, the 3D structure is sparse, therefore, the recall rate for lidar is not that high. The visual imagery recall rate is much higher due to the feature-rich environment, however, it varies significantly as the illumination changes during the

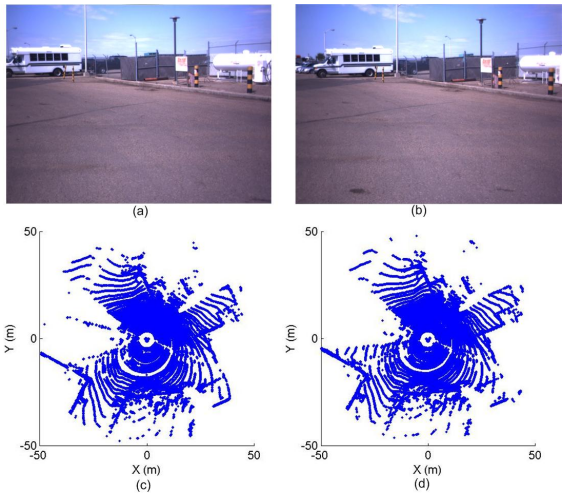

Fig. 11. Another true positive in the urban environment where the place is recognized by both visual and lidar. The image pair is shown in (a) and (b) while the corresponding top-down view of the point cloud is shown in (c) and (d) respectively.

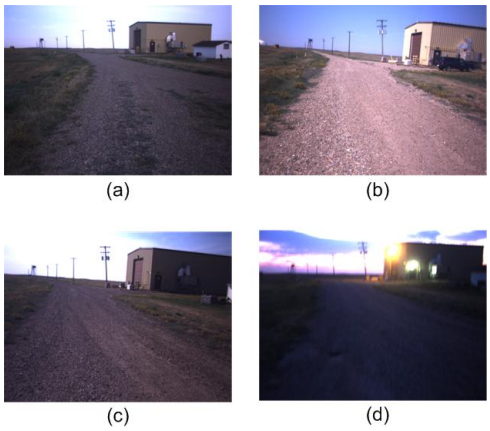

Fig. 12. Example image at a similar location at different times of the day: (a) 9am (b) 1pm (c) 5pm (d) 9pm. Note that the illumination changes significantly throughout the 12-hour period.

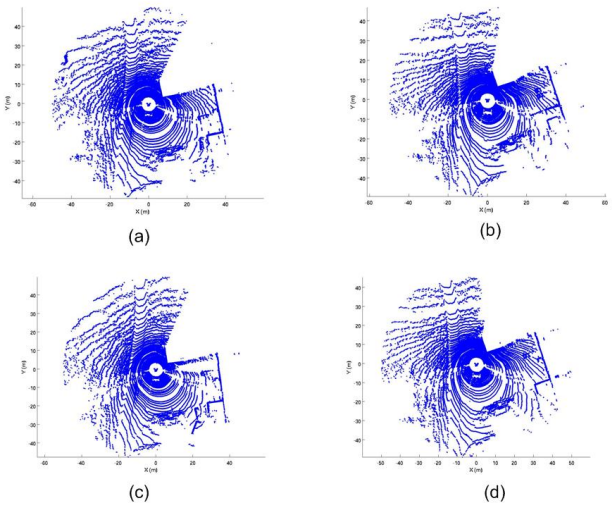

Fig. 13. Example lidar at a similar location at different times of the day: (a) 9am (b) 1pm (c) 5pm (d) 9pm. Note that the lidar data is stable throughout the 12-hour period, as it is not affected by ambient illumination.

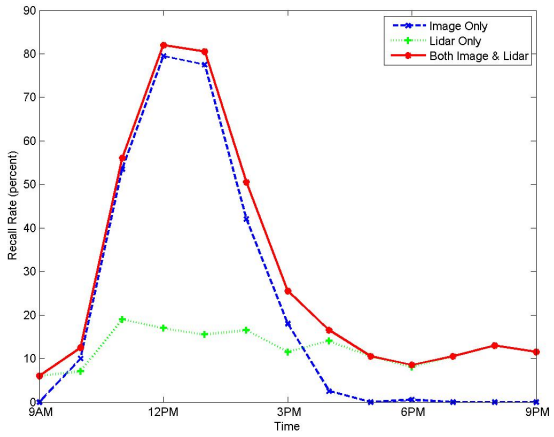

Fig. 14. Recall rates during a 12-hour period for imagery only, lidar only and for using both imagery and lidar together. It can be seen that imagery recall rate is much higher but affected significantly by illumination, while the lidar recall rate is lower but independent of ambient lighting condition.

day. When both the stereo and lidar data are used with the OR integration scheme, it offers the best of both worlds. Under favourable illumination, visual place recognition provides very high recall rate. On the other hand, under adverse lighting conditions, the system still provides an adequate recall rate thanks to the lidar data.

V. Discussion

For this work, the purpose of place recognition is to provide the loop closure information for the geometric SLAM to update its trajectory and map. False positives are highly undesirable, as they will corrupt the SLAM map. On the other hand, a low recall rate is acceptable, as there is no need to correct the SLAM map per frame. At least one correct place recognition is sufficient for each loop the vehicle traverses. The experimental results show that the proposed multi-sensor approach provides a sufficient recall rate throughout variable lighting condition at 100% precision, i.e., without any false positives, and hence is highly suitable for mobile robot SLAM applications.

The final decision of a loop detection is based on a probability threshold. For all the experimental results shown above, we use a high probability threshold (0.99) for FAB-MAP to be considered for 6-DOF validation. A lower threshold could be used which may lead to more loop detections, but may also introduce false positives into the system and would increase the 6-DOF computation of the system. Practically, a high probability threshold was chosen to reduce the possibility of false positives and reduce computation time of the system such that it can run in real-time. With this high threshold, we obtain zero false positives even with the OR integration

scheme and hence the AND integration scheme is not useful for these experiments.

Future work includes more experimentation by varying this threshold to determine the true positives and false positives rates, in order to generate the ROC (Receiver Operating Characteristics) curve to characterize its performance. The benefit of using the AND integration scheme over the OR integration scheme should be more apparent at a lower probability threshold.

Probability thresholding works well when sequential images added to the database are significantly distinct. In the case where two or more images from the same physical location are used as input, FAB-MAP will miss these loop closures as the probability distribution is essentially split between the two images. Therefore, we plan to develop an automated keyframe detection method, similar to [20], that measures similarity between the input data and only process frames that are sufficiently different for loop detection. This will not alleviate the problem if a vehicle were continuously looping through the same route. In this case, more sophisticated methods for managing the database or detecting cyclical loops need to be developed.

VI. Conclusion

This paper proposed a multi-sensor approach for appearance-based mapping that exploits the inherent strengths of both visual and lidar sensors. The purpose of this research is to develop a place recognition system that can operate effectively regardless of lighting conditions enabling 24-hour vehicle operation. Experimental results from both urban and rural environments show good recall rates with 100% precision, and that the multi-sensor system was capable of accurately detecting loops during a 12-hour period from dawn to dusk.

VII. Acknowledgements

A portion of the source code used in this work was licensed from Isis Innovation Limited. The authors would like to thank Paul Newman from the Oxford Mobile Robotics Group for his software support.

References

[1] E. Eade and T. Drummond, "Unified loop closing and recovery for real time monocular slam," in *British Machine Vision Conference (BMVC)*, 2008.

[2] M. Magnusson, H. Andreasson, A. Nüchter, and A. J. Lilienthal, "Appearance-based loop detection from 3d laser data using the normal distributions transform," in *Proceedings of the 2009 IEEE international conference on Robotics and Automation*, ser. ICRA'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 3364–3369. [Online]. Available: http://dl.acm.org/citation.cfm?id=1703775.1703991

[3] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," in *The International Journal of Robotics Research*, vol. 27, no. 6. SAGE Publications, June 2008, pp. 647–665.

[4] K. Ho and P. Newman, "Detecting loop closure with scene sequences," *International Journal of Computer Vision (IJCV)*, vol. 74(3), pp. 261–286, 2007.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004, 10.1023/B:VISI.0000029664.99615.94. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94

[6] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008, similarity Matching in Computer Vision and Multimedia. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314207001555

[7] B. Taati, M. Bondy, P. Jasiobedzki, and M. Greenspan, "Variable dimensional local shape descriptors for object recognition in range data," in *International Conference on Computer Vision - Workshop on 3D Representation for Recognition*, 2007.

[8] J. Collier, S. Se, V. Kotamraju, and P. Jasiobedzki, "Real-time lidar-based place recognition using distinctive shape descriptors," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8387, May 2012.

[9] J. Sivic and A. Zisserman, "Efficient visual search of videos cast as text retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 591 –606, april 2009.

[10] M. Cummins and P. Newman, "Highly scalable appearance-only slam - fab-map 2.0," in *Robotics: Science and Systems Conference*, 2009.

[11] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard, "Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation," in *IEEERSJ International Conference on Intelligent Robots and Systems IRSO*, 2011.

[12] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "Point feature extraction on 3d range scans taking into account object boundaries," in *Proceedings of the International Conference on Robotics and Automation*, 2011.

[13] K. Granström, T. Schön, J. I. Nieto, and F. T. Ramos, "Learning to close loops from range data," *The International Journal of Robotics Research*, vol. 30, no. 14, pp. 1728–1754, Dec. 2011.

[14] R. Paul and P. Newman, "FAB-MAP 3D: Topological mapping with spatial and visual appearance," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, may 2010, pp. 2649 –2656.

[15] W. Maddern and S. Vidas, "Towards robust night and day place recognition using visible and thermal imaging," in *Robotics Science and Systems Conference 2012*, University of Sydney, July 2012. [Online]. Available: http://eprints.qut.edu.au/52646/

[16] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 21(5), pp. 443–449, 1999.

[17] C. Chua and R. Jarvis, "Point signatures: a new representation for 3d object recognition," *International Journal of Computer Vision (IJCV)*, vol. 25(1), pp. 63–85, 1997.

[18] M. B. A.S. Main and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 28(10), pp. 1584–1601, 2006.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

[20] N. Muhammad and S. Lacroix, "Loop closure detection using small-sized signatures from 3d lidar data," in *Safety, Security, and Rescue Robotics (SSRR), 2011 IEEE International Symposium on*, nov. 2011, pp. 333 –338.