Real-Time Lidar-Based Place Recognition Using Distinctive Shape Descriptors

Jack Collier[1], Stephen Se[2], Vinay Kotamraju[2] and Piotr Jasiobedzki[3]

[1]Defence Research and Development Canada – Suffield, Medicine Hat, AB, Canada
[2]MDA Systems Ltd., 13800 Commerce Parkway, Richmond, B.C., V6V 2J3, Canada
[3]MDA, Space Missions, 9445 Airport Road, Brampton, Ontario, L6S 4J3, Canada

ABSTRACT

A key component in the emerging localization and mapping paradigm is an appearance-based place recognition algorithm that detects when a place has been revisited. This algorithm can run in the background at a low frame rate and be used to signal a global geometric mapping algorithm when a loop is detected. An optimization technique can then be used to correct the map by 'closing the loop'. This allows an autonomous unmanned ground vehicle to improve localization and map accuracy and successfully navigate large environments. Image-based place recognition techniques lack robustness to sensor orientation and varying lighting conditions. Additionally, the quality of range estimates from monocular or stereo imagery can decrease the loop closure accuracy. Here, we present a lidar-based place recognition system that is robust to these challenges. This probabilistic framework learns a generative model of place appearance and determines whether a new observation comes from a new or previously seen place. Highly descriptive features called the Variable Dimensional Local Shape Descriptors are extracted from lidar range data to encode environment features. The range data processing has been implemented on a graphics processing unit to optimize performance. The system runs in real-time on a military research vehicle equipped with a highly accurate, 360 degree field of view lidar and can detect loops regardless of the sensor orientation. Promising experimental results are presented for both rural and urban scenes in large outdoor environments.

Keywords: Appearance-Based SLAM, place recognition, lidar, local shape descriptors, GPU

1. INTRODUCTION

An enabling technology for future autonomous military vehicles is the production of an accurate environmental map that allows an Unmanned Ground Vehicle (UGV) to model its environment so as to interact with it effectively. Past systems relied heavily on the availability of Global Positioning System (GPS) and sophisticated inertial navigation systems to provide an accurate pose estimate of the vehicle to register range data from lidar or stereo imagery. However, GPS dropouts often occur due to low signal strength and GPS is generally not available indoors. The likelihood of GPS dropout is particularly high in the urban battle-space where buildings and infrastructure occlude the receiver from the signal. The availability of low-cost GPS jamming technologies adds to the problem. An alternative technique is therefore necessary to allow accurate mapping in the absence of GPS.

Over the past decade, a family of algorithms called Simultaneous Localization and Mapping (SLAM) emerged as the dominant technique for dealing with this problem. SLAM attempts to estimate the UGVs pose by detecting and re-observing salient features in the environment. The position of these landmarks act as geometric constraints to improve the robot pose estimate. The improved robot pose estimate is used to improve the pose estimate of the landmarks themselves. On-line SLAM techniques are subject to failure, especially over large areas, due to linearization errors and improper data associations (Castellanos et al.)[1]

To overcome these problems, a family of techniques called Appearance-Based SLAM (ASLAM) can be used that decouple the loop closing process from the geometric SLAM algorithm and runs it in parallel (Cummings and Newman),[2](Magnusson et al.),[3](Ho and Newman)[4]. Here, the loop closing process is essentially a scene

Further author information:
E-mail: jack.collier@drdc-rddc.gc.ca

Unmanned Systems Technology XIV, edited by Robert E. Karlsen, Douglas W. Gage,
Charles M. Shoemaker, Grant R. Gerhart, Proc. of SPIE Vol. 8387, 83870P
 · doi: 10.1117/12.918760

recognition computer vision task. In this way, loop recognition is not bound to the pose estimates of the SLAM procedure and hence not prone to failure resulting from gross errors in the robot pose estimate. For practical tasks, such as robot navigation, loop closing data must be used to update the geometric SLAM formulation and techniques have been developed to compute the transform necessary to close the loop. Though most ASLAM algorithms were developed for use with monocular or stereo imagery, the techniques have recently been extended to range sensors such as lidar. Lidar-based ASLAM benefits from increased accuracy of range data, and is robust to large changes in lighting and orientation making it an ideal candidate for place recognition. The introduction of several new high-speed, high-definition range sensors on the market has enabled this technology.

This paper presents a range-based place recognition system that attempts to address the shortcomings of video-based ASLAM. In particular, a shape descriptor called the Variable-Dimensional Local Shape Descriptor (VD-LSD) is computed for 360 degree range scans and used to learn a generative model of place appearance and determines if an observation comes from a new or previously seen place. Section 2 provides a review of appearance-based systems, Sections 3 and 4 detail the VD-LSD based ASLAM system used here and results from initial field trials. The results are discussed in Section 5. Finally, Section 6 presents conclusions and future work to improve the system.

2. BACKGROUND

ASLAM algorithms attempt to determine when a place has been revisited by computing similarity between different image scenes. State-of-the-art ASLAM techniques have their roots in text retrieval systems used by popular search engines. Typically, documents are parsed and categorized and a vector containing the frequency of occurrence of words in the document is used to categorize it. ASLAM techniques are a visual analog of these systems where the frequency of occurrence or co-occurrence of visual features is used to match a scene against a database (Sivic and Zisserman)[5].

In this approach, an image is treated as a Bag-of-Words (BoW) much like a text document, where a "word" corresponds to a region in the space of invariant descriptors. Features are extracted from training images and BoW clustering is performed to partition the feature space. Words are formed from the centroid of each feature cluster to create a vocabulary to which new images will be categorized. A weight is associated with each cluster and is inversely proportional to the frequency of the word occurrence. A new image is characterized, by appearance, as the set of words that it contains extracted from the vocabulary (the appearance vector). Various techniques are used to compare these appearance vectors and find loop closures.

Ho and Newman[4] identified a repeated visit by checking whether a sequence of image appearances has previously occurred. The similarity between two images is defined as the normalized inner product (cosine of the angle) between the appearance vectors of the two images. A similarity matrix was constructed and used to search for repeated sequences. Noise was greatly reduced in the similarity matrix by noticing that the largest eigenvalues are associated to the noise, and should be removed from the spectral decomposition of the matrix. An entropy argument is provided for deciding how many eigenvalues to remove.

Eade and Drummond[6] used a similar BoW technique but were able to update their vocabulary as each new image was processed rather than through an off-line supervised learning technique. The authors mentioned in the paper that the worst failure mode of the system is spurious loop closure given extensive repeated structure and suggested the use of a probabilistic model to mitigate this issue.

FAB-MAP (Cummings and Newman)[7] 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. Doing so allows them to overcome the challenge of perceptual aliasing, wherein similar looking image frames from different places become particularly hard to differentiate during loop closure. Specifically, FAB-MAP begins by defining a probabilistic graph of latent variables (locations and objects) and the observed variables (appearance vectors from images). For a given appearance vector that is derived from an incoming image frame, the probability of revisiting a previous location is computed based on a Bayesian framework that includes large dimensional joint probability distributions that are learned from training data and efficiently computed using

Chow-Liu trees. The final detection of a loop is made by a Maximum Likelihood data association step followed by a validation step that uses epipolar geometry estimates and Random Sample Consensus (RANSAC). To achieve an efficient large scale implementation, FAB-MAP 2.0 (Cummins and Newman)[2] uses an inverted index retrieval architecture that requires modification to the way probabilities of revisits were computed and updated in FAB-MAP 1.0. FAB-MAP 2.0 has been demonstrated on a large 1000 Km dataset with few false positive results.

ASLAM has predominantly been applied to 2D imagery, but has begun to be applied to 3D range data. In (Steder et al.),[8] Normal-Aligned Radial features (Steder et al.)[9] were extracted from a range image and used to create a BoW vocabulary. A similarity for all the scans in the database was computed based on the appearance and ordered from most to least similar. Feature points were then extracted from the scan pairs and candidate transformations were computed. The transformation with the highest score above a certain threshold is used to select the final recognized place. In (Paul and Newman),[10] image-based FAB-MAP was augmented with 3D lidar data used to capture the spatial distribution of visual words in the scene. A random graph is used in which nodes are the words and edges are the distributions over distance and are viewpoint invariant. Results indicate improved precision and recall rates over FABMAP-2.0. Magnusson et al.[3] propose an approach based on extracting and matching a global appearance descriptor derived from the Normal Distribution Transform (NDT). The input to the algorithm was a three dimensional (3D) point cloud obtained from a LIDAR scan of the scene. NDT models the global shape of the scene as a collection of locally centered normally distributed probability density functions. The covariance matrices of the density functions are used to classify local shapes as one of many idealized classes and sub-classes of shapes. The appearance descriptor of the scan is then constructed based on a set of histograms of shape classes computed at a set of range intervals. Since the appearance descriptor uses the orientation of surfaces, rotational invariance is achieved by generating histograms based on multiple surface orientations. A metric is proposed that quantifies the difference between two appearance descriptors. A threshold on this difference metric, which is learned off-line based on an expectation maximization method, allows identification of revisits. The authors recognize the possibility of more robust results through the use of machine learning to disregard non-discriminative features and the use of more elaborate analysis of the similarity matrix between scene matches than just a single threshold. However, despite these limitations in the algorithm, the authors report recall rates of up to 70% at 100% precision for indoor and outdoor datasets.

3. SYSTEM DESCRIPTION

The system developed here is a multi-sensor ASLAM system shown in Figure 1 and is based on the FAB-MAP system previously described. The algorithm is run in parallel for both image-based FAB-MAP, using SIFT features (Lowe),[11] and range-based FAB-MAP. Separate place recognition results are generated from both sensor streams and integrated to provide a final loop closure result. The application of ASLAM to separate sensors running asynchronously is a novel aspect of this work and is the subject of future research. All further discussion in this paper focuses only on the range-based ASLAM system developed here.

The system consists of two phases, training and real-time loop detection. In the training phase, representative lidar data consisting of full 360 degree scans are input to the feature extraction algorithm. VD-LSD features, are output from the training imagery. Bag of words clustering is performed in the VD-LSD feature space to group features of similarity and create the representative vocabulary.

During run-time, new VD-LSD features are extracted from data. A probabilistic detector model converts the extracted features into an appearance vector which determines which words are present in the current scene. For a vocabulary of n words, the associated VD-LSD 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 scene and decision of a match is made. If a match has been determined, the current and matched frames are then fed to a 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 match, the 6-DOF module will detect the geometric inconsistency and discard the match.

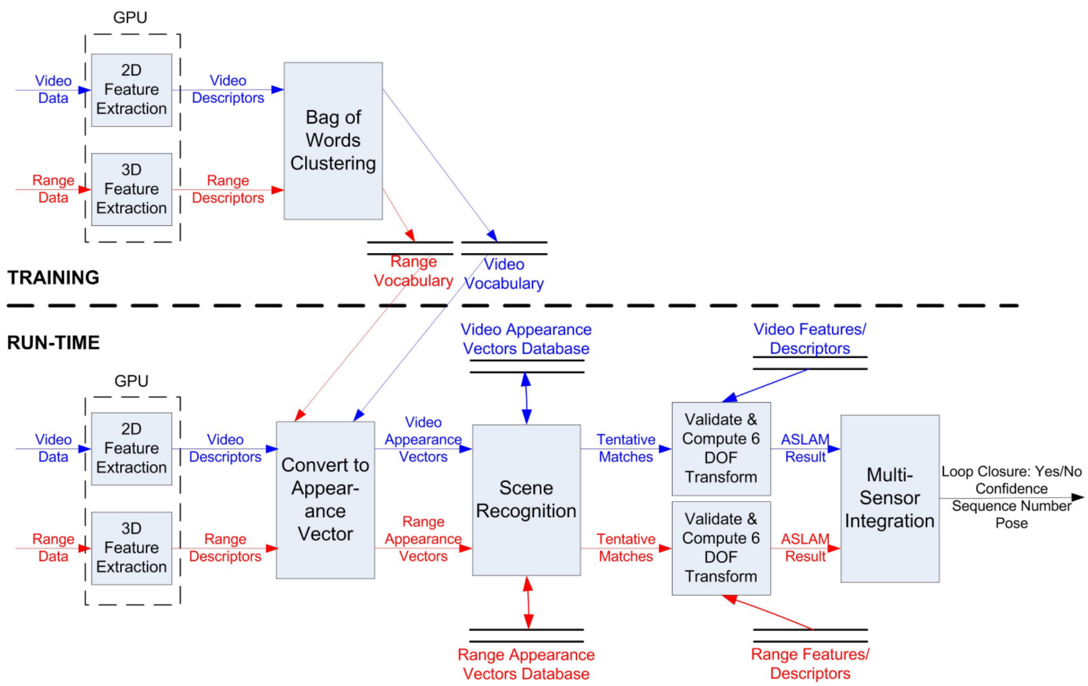

Figure 1. The lidar-based ASLAM system is based on the FAB-MAP algorithm and uses Variable Dimensional Local Shape Descriptor to create the vocabulary used for place recognition. VD-LSD features are extracted from real-time lidar data, converted to an appearance vector and input to the FAB-MAP algorithm. If a match is detected, a 6-DOF transform can be computed to determine the transformation between matched frames.

3.1 Variable-Dimensional Local Shape Descriptor

The principal input to the place recognition algorithm is the Variable-Dimensional Local Shape Descriptor, described in (Taati et al.),[12] that has previously been used for model-based object recognition and pose estimation of range data.

The VD-LSD for a salient point is generated from invariant properties extracted from the principal component space of the local neighbourhood. The 3x3 covariance matrix of the local neighbourhood 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.

The VD-LSD offers a generalized form that subsumes a large class of popular descriptors in literature, such as Spin Images (Johnson and Hebert),[13] Point Signatures (Chua and Jarvis)[14] and Tensor Correspondence (Main et al.)[15] 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.

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. A forward selection scheme (Taati and Greenspan)[16] is utilized to select the near-optimal subset for each object. The desired subset of properties is determined off-line for each object and only those properties are extracted in the on-line computation.

The VD-LSD extraction algorithm is parallel and highly suitable for a Graphics Processing Unit (GPU) implementation. Processor-intensive modules have been implemented in CUDA to run on an 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 two, thereby producing descriptors of length 64.

3.2 6-DOF Transform

The 6-DOF transform between the current frame and the recognized frame is essential for map correction in metric SLAM. Therefore, if a match is found during scene recognition, the 6-DOF transform is then computed. This process can also help to validate the FAB-MAP output as improper loop detections can be discarded during the computation of the transformation.

Two metrics are used to measure the degree of similarity between the VD-LSDs from the current frame and the VD-LSDs from the recognized frame. The first is the Euclidean distance between two VD-LSD descriptor vectors to evaluate the similarity between them. The second metric is an ambiguity measure based on the ratio of the top two ranking matches. The ratio between the Euclidean distance of the best-ranking match and the second-ranking match should be smaller than 1 to be a good match.

Once a set of tentative matches are obtained, a RANSAC approach (Fischler and Bolles)[17] is performed to remove the outliers. Three tentative matches are selected randomly to compute the 6-DOF transform using the rigid body transformation approach, to serve as a hypothesis. The number of supporting matches from the tentative set are obtained for each hypothesis. This process is performed repeatedly and the hypothesis with the highest number of supporting matches is selected. All the supporting matches are used to compute the 6-DOF transform using a least-squares minimization approach.

Various validation schemes can be performed during the 6-DOF transform computation as well. A minimum number of supporting VD-LSD matches can be specified to be a valid recognition. Moreover, the 6-DOF transform obtained can be used for verification, since both the translations and two of the three rotations should be small for a correct recognition constrained by ground vehicle motion.

Figure 2. Images of the trial environments. (**top**) The rural environment is an open prairie area with sparsely populated infrastructure, rolling grasslands and gravel roads. (**bottom**) The urban environment contains buildings, cars, and paved roads.

4. RESULTS

Field trials were conducted in September of 2011 on the DRDC Suffield Experimental Proving Ground to test the lidar-based ASLAM systems ability to perform place recognition. Two separate sets of trials were conducted to test the performance of the system in representative rural and urban environments (Figure 2). A research variant of the Multi-Agent Tactical Sentry (Figure 3), a modified Kawasaki mule developed at DRDC Suffield, was retrofitted with a Velodyne HD lidar and computing hardware to carry out the experiments. Data from the

Velodyne, and GPS were collected using DRDC's Architecture for Autonomy logging capabilities. This allowed the researchers to playback the sensor data at a later time and exercise the ASLAM system as though it were operating on live data. Additional tests were performed by routing the data directly into the ASLAM system to ensure real-time use. All software, off-line and in real-time, was tested on a high performance server housed on the MATS vehicle. The server contains 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.

The Velodyne Lidar rotates 360 degrees in the azimuth at 15Hz and produces approximately one million range reading per second. 360 degree scans were captured twice per second, converted to 3D point clouds (X,Y,Z values), and sent to the ASLAM system.

Tests were performed in the urban environment shown in Figure 2 (bottom). The vehicle was manually driven and data was logged for later processing. Figure 4 (left) shows the path that the vehicle was driven during testing. Blue dots indicate the GPS position during non-loop closure while red dots indicate the GPS position when a loop closure is detected. GPS was used for display and validation purposes and is not input to the ASLAM algorithm. No 6-DOF validation was performed in this analysis. The average run-time is less than one second per frame. To prevent loop detections between consecutive scans, the prior probability for the previous 20 scans was set to zero. This equates to roughly 10 seconds of driving and was chosen based on user experience.

Figure 3. Research platform used to conduct the tests. The Velodyne lidar housed on top of the vehicle was used for all tests.

The trajectory included several loops around paths in either the same or the reverse direction. Several confidence thresholds were tested and used to make the final determination if the detection represented a true loop closure. Table 1 summarizes the results. Predictably, more loops could be detected using a lower confidence threshold, but at the expense of more false positives. Even with a confidence threshold of 99.99%, 92 loops were detected with only 7 false loops. Since false positives can lead to catastrophic failures in a geometric SLAM formulation, a high threshold ($> 99\%$) should be used. The loop validation scheme outlined in Section 3.2 can be used to further eliminate false positives, but it was not implemented at the time of testing. Figure 5 shows typical scan results for true and false loop detections. Note that the algorithm is able to detect loops when passing in opposite directions. True matches were verified by comparing GPS location of the resulting scans. A match was deemed a true loop closure if the distance between the 2 scans was less than 10m. Despite the high threshold numerous positive loops were detected that could be used to correct a geometric SLAM formulation.

Table 1. Urban dataset test results. Total number of loops, true loops, and false loops detected for a given confidence threshold. False positive were those loops where the GPS position between matches was greater than 10 meters.

Threshold %	**loops detected**	**positive (%)**	**false (%)**	**Total Scans**
60	237	134 (57)	103 (43)	596
80	207	129 (62)	78 (38)	596
90	189	127 (67)	62 (33)	596
97	135	112 (83)	23 (17)	596
99	127	108 (85)	19 (15)	596
99.99	92	85 (92)	7 (8)	596

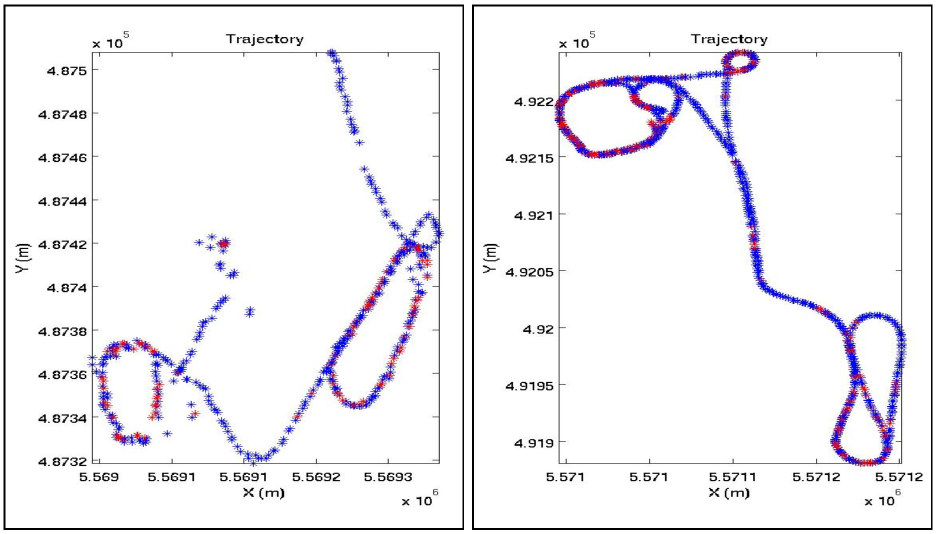

Figure 4. GPS trajectories for the urban (**left**) and rural (**right**) tests. Red dots indicate the nearest recorded waypoint from where a loop closure was detected. A confidence threshold of 99% was used to determine true loop closures in this case.

Similar results are reported for the rural environment (Table 2). The percentage of positive loop closures is much lower than in the urban dataset while the percentage of false positives is higher especially when using the high confidence thresholds. This can most likely be attributed to the urban environment having more features and more distinct features than the rural dataset. The average number of features detected per image was 1567 for the urban dataset as opposed to 1314 for the rural dataset.

Table 2. Rural dataset test results. Total number of loops, true loops, and false loops detected for a given confidence threshold. False positive were those loops where the GPS position between matches was greater than 10 Meters

Threshold %	**loops detected**	**positive (%)**	**false (%)**	**Total Scans**
60	528	292 (55)	236 (45)	1343
80	490	287 (59)	203 (41)	1343
90	454	278 (61)	176 (39)	1343
99	309	225 (73)	84 (27)	1343
99.99	179	148 (83)	31 (17)	1343

The validation and 6-DOF transformation scheme described in Section 3.2 was not available at the time that the initial tests were performed, but has since been added to the system. An ambiguity ratio threshold of 0.8 was used to reject ambiguous candidate features. If a minimum of six tentative matches could not be found after applying the RANSAC rigid body transformation approach, the loop closure was rejected. Figures 5 show the results of VD-LSD alignment for several true loop closures while Figure 6 shows false positive loops that were rejected by the 6-DOF validation algorithm.

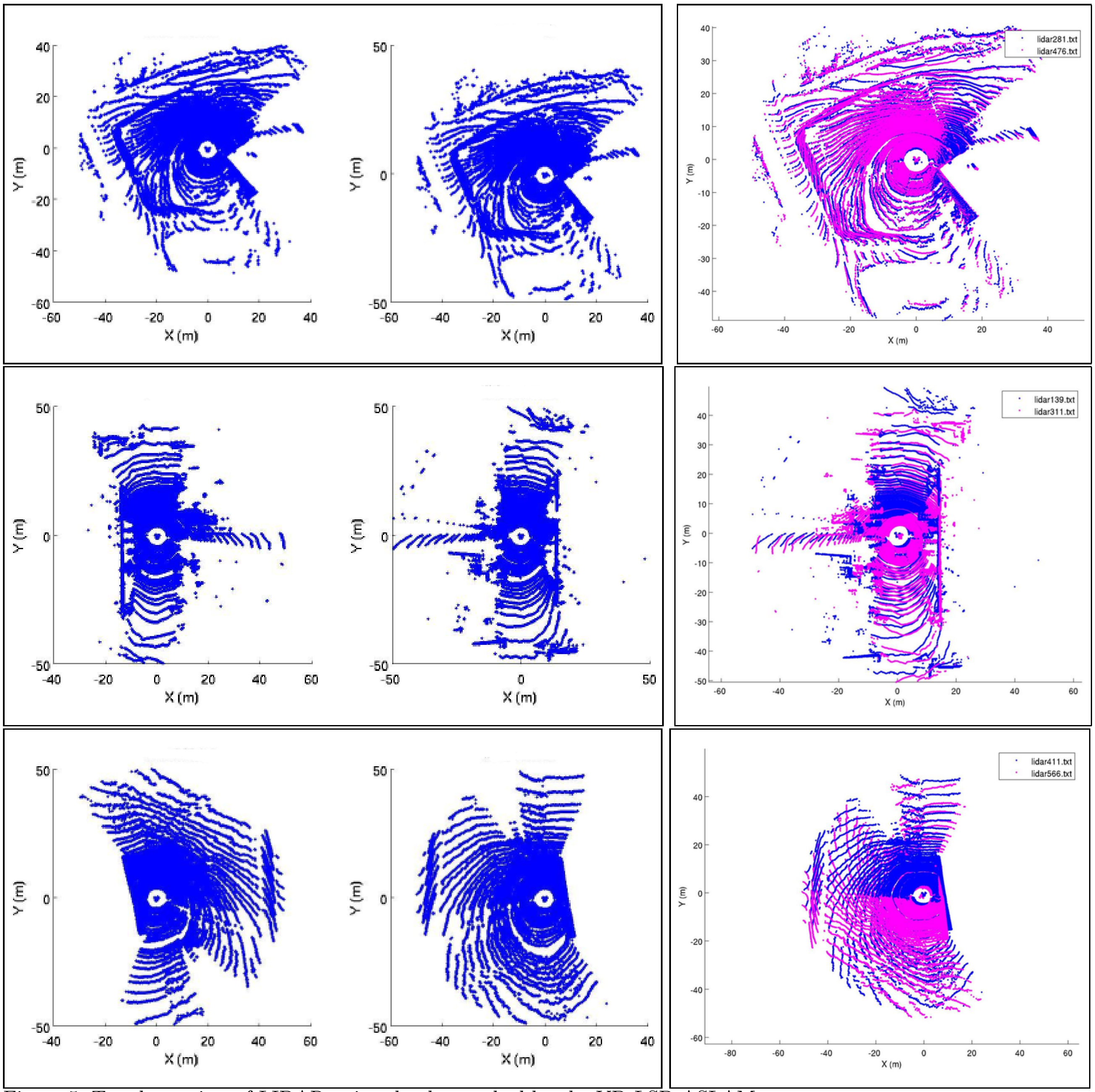

Figure 5. Top-down view of LIDAR point clouds matched by the VD-LSD ASLAM system.
(top) True loop closure (same direction) **6-DOF transformation:** $[0.1m, -0.2m, -0.7m, 1.2^\circ, 0.03^\circ, -1.5^\circ]$.
(middle) True loop closure (opposite direction) **6-DOF transformation:** $[0.4m, -0.1m, -1.2m, -0.2^\circ, -7.7^\circ, 178.3^\circ]$.
(bottom) True loop closure (opposite direction) **6-DOF transformation:** $[-1.1m, 0.3m, 1.9m, -0.6^\circ, -4.3^\circ, 173.1^\circ]$.

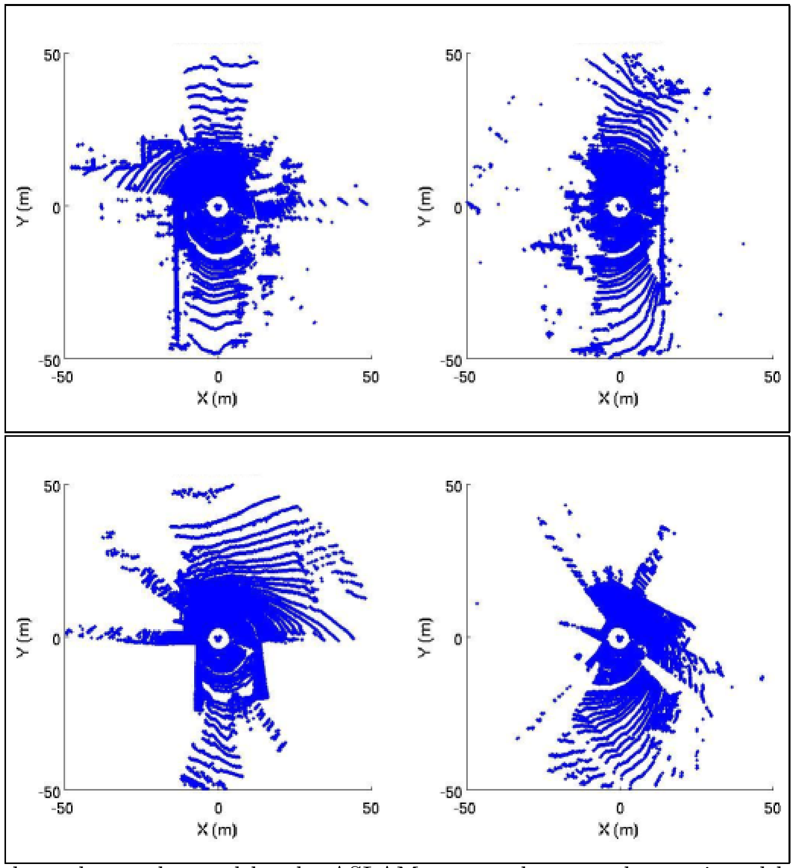

Figure 6. False positive loop closure detected by the ASLAM system that were later rejected by the 6-DOF validation and transformation algorithm. Note the similarity in the overall look of the scan. The top scans appears to be similar, but rotated 180°.

5. DISCUSSION

Even in the urban environment with the highest confidence level, the system successfully detected 92 loops with many successive detections around a single loop. This is more than enough detections to augment a geometric map and 'close the loop'. State-of-the-art systems perform loop closure at a low rate since it is an intensive procedure. In this case, it may be advisable to ignore loop detections so that real-time mapping can be maintained.

The multi-sensor approach presented at the beginning of Section 3 should improve the recognition results, especially in cases where their is an abundance of scene texture, but a lack of 3D lidar-based features. In this case, the image-based ASLAM should still be capable of detecting loops. Conversely, in low light or low texture areas, the lidar-based place recognition would still function.

An automated keyframe detection method, similar to (Muhammad and Lacroix),[18] could be used that measures similarity between scans and only input scans that are sufficiently different for loop detection. This may increase the efficiency of the system as unnecessary information is not added to the database. Under such a mechanism, scans would be added more often when the vehicle is travelling at a faster rate or turning and wouldn't include new scans when the vehicle is stationary. The Euclidean similarity between consecutive VD-LSDs, that is used in validation and 6-DOF calculation (Section 3.2), may have merit as a measure of scan distinctiveness.

6. CONCLUSIONS AND FUTURE WORK

It was previously noted that monocular and stereo-based ASLAM systems suffer from an inability to cope with large variations in lighting and approach angle (especially in the case of forward looking cameras). This paper presented a novel approach to place recognition that sought to alleviate these problems by using a range sensor and a feature descriptor called the Variable Dimensional Local Shape Descriptor that was computed from high definition lidar data and fed to the well known FAB-MAP ASLAM system. Preliminary results show, for the datasets tested, the system is able to accurately detect previously seen environments even when approached from 180 degrees out of phase and in different lighting conditions. Furthermore, it is able to perform this operation in real-time due in large part to the GPU implementation of the VD-LSD algorithm.

The authors believe a multi-modal approach to sensing is important, especially in the context of the battlefield where environmental conditions vary so widely. Integration of lidar-based ASLAM into the larger multi-sensor approach presented here will allow the system to operate in a large variety of environmental conditons and will be the main focus of future efforts. To date, all tests have been performed using a VD-LSD with 6 dimensions and a bin size of 2. Alternate descriptor sizes may allow better detection of loop closures but at the expense of a longer processing time. Study of the optimal number of dimensions and bin size along each dimensions will be the subject of future work. Integration of alternative 3D range sensors such as the well known Kinect sensor and the Autonosys lidar will also be explored. The output of the system will also be integrated into a full SLAM solution with the 6-DOF transform being used to perform geometric loop closure. We have presented an architecture that computes video and lidar-based ASLAM results in parallel. Successful integration of these parallel streams to form a seamless ASLAM output will be the subject of future research.

7. ACKNOWLEDGEMENTS

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

REFERENCES

[1] Castellanos, J. A., Neira, J., and Tardós, J. D., "Limits to the consistency of ekf-based slam," in [*5th IFAC Symp on Intelligent Autonomous Vehicles IAV04*], (2004).

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

[3] Magnusson, M., Andreasson, H., Nüchter, A., and Lilienthal, A. J., "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*], *ICRA'09*, 3364–3369, IEEE Press, Piscataway, NJ, USA (2009).

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

[5] Sivic, J. and Zisserman, A., "Video google: A text retrieval approach to object matching in videos," in [*Proceedings of the Ninth IEEE International Conference on Computer Vision*], (2003).

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

[7] Cummins, M. and Newman, P., "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," in [*The International Journal of Robotics Research*], **27**, 647–665, SAGE Publications (June 2008).

[8] Steder, B., Ruhnke, M., Grzonka, S., and Burgard, W., "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).

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

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

[11] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**, 91–110 (2004). 10.1023/B:VISI.0000029664.99615.94.

[12] Taati, B., Bondy, M., Jasiobedzki, P., and Greenspan, M., "Variable dimensional local shape descriptors for object recognition in range data," in [*3dRR-07: ICCV '07Workshop on 3D Representation for Recognition*], 8 pages (October 2007).

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

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

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

[16] Taati, B. and Greenspan, M., "Satellite pose acquisition and tracking with variable dimensional local shape descriptors," in [*IEEE/RSJ International Conference on Intelligent Robots and Systems - Workshop on Robot Vision for Space Applications*], (2005).

[17] Fischler, M. A. and Bolles, R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM* **24**, 381–395 (June 1981).

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