Advancement of vision-based SLAM from static to dynamic environments

Glen Bright, Ardhisha Pancham
UKZN, Durban, South Africa
CSIR, Pretoria, South Africa

Nkgatho Tlale
Anglo American, Johannesburg, South Africa

Abstract—**Recent years, have seen the advancement of vision-based Simultaneous Localization And Mapping (SLAM) from static to dynamic environments, where SLAM combined with Detection And Tracking of Moving Objects (DATMO) algorithms is able to handle moving objects. This paper provides a literature review of the advancement of vision-based SLAM from stationary to dynamic environments. It looks at the algorithms and sensor configurations that have enabled mobile robots to produce a reliable map of the environment for localization and task completion in the presence of moving objects. It determines the algorithms that will be used to allow a mobile robot to perform SLAM and DATMO with two RGB-D cameras in an indoor dynamic environment and concludes where a contribution can be made for SLAM and DATMO.**

Keywords – SLAM; DATMO; SLAMIDE; dynamic; tracking; vision-based SLAM

I. INTRODUCTION

Simultaneous Localization And Mapping (SLAM) allows a mobile robot to construct a map of an unknown, static environment and simultaneously localize itself. Real world environments, however, have dynamic objects such as people, opening and closing doors, movable furniture, cars, and other robots. In order for a mobile robot to be completely autonomous in such an environment, the SLAM algorithm must be complemented by Detection And Tracking of Moving Objects (DATMO) algorithms that can manage the dynamic objects. Moving objects must be not be included in the SLAM map as they may lead to localization errors and reduce map quality.

The SLAM problem has been researched extensively in static environments. Applications have evolved from different environments such as indoor to outdoor, airborne, and underwater. Most of these applications, however, use laser scanners for their accuracy, and are undertaken in stationary environments that do not account for dynamic objects. [1-3]

Cameras have certain advantages compared to laser scanners. They supply information of a higher quality for environment perception, are generally smaller, less expensive, and have lower power consumption. [4-6]

Recent work has seen the use of cameras in singular and stereo arrangements to perform SLAM in dynamic environments owing to their advantages. Open questions, however, still exist. These questions are:

- How to distinguish between static and dynamic objects,
- How to represent static and dynamic objects, and
- How to track dynamic objects and predict their positions over time.

Dynamic objects fall into two different categories, those that are always moving and those that are temporary static and change their position over time. Dynamic objects can be incorrectly classified as stationary and lead to data association errors that decrease map accuracy. [2, 7, 8]

This paper explores the algorithms that have allowed vision-based SLAM to advance from static to dynamic environments. The paper is organized as follows: Section II explains SLAM and DATMO. Section III defines vision-based SLAM. Section IV reviews the advancement of algorithms. Section V compares the different techniques. It comes to a decision of the algorithms that will be used for the mobile robot application and indicates where a contribution will be made. Section VI concludes the paper, and Section VII describes future work.

II. SLAM AND DATMO

As illustrated in Fig. 1, the inputs of SLAM and DATMO are measurements from perception sensors, *Z*, and motion sensors, *U*. The outputs are the map of the environment, *M*, the robot pose, *x*, the positions of dynamic objects, *O*, and the trajectories of dynamic objects, *S*. [6]

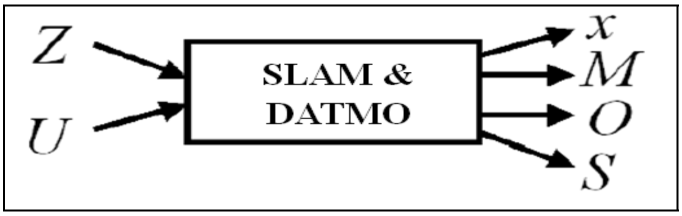

Figure 1. The SLAM and DATMO process. Where *Z* represents perception measurements, *U*, motion measurements, *x*, the true robot state, *M*, static object locations, *O*, dynamic object states and *S*, trajectories of dynamic objects. Diagram adapted from [6].

III. VISION-BASED SLAM

Vision-based SLAM involves methods where the external environment is observed only through the use of vision sensors, i.e. cameras. This also includes methods where motion sensors are used to supply the scale estimate which is not provided by Bearing-Only (BO) sensors. Geometric features such as points, lines, and arcs are extracted from the images to construct models of the environment. [9]

IV. REVIEW OF ADVANCEMENT OF VISION-BASED SLAM ALGORITHMS

The proceeding work first presents the pioneer work of Wang in a laser-based solution for SLAMIDE. It then describes the progress of vision-based algorithms from static to dynamic environments, and from single camera to multiple camera systems. Applications in dynamic environments are described in detail. The SLAM technique employed and the moving object detection and tracking methods to achieve SLAMIDE have been examined. The sensors utilized, the environment, the map representation, and objects detected have also been named.

Wang [6] was the first author to integrate SLAM and DATMO and demonstrated that the two processes were mutually beneficial, through the use of laser scanners. A Bayesian formula based on Extended Kalman Filters (EKF) was presented to solve SLAM and DATMO simultaneously. Separate posteriors for the static and dynamic objects were maintained.

Iterative Closest Point (ICP) was utilized to represent the data in grid maps. SLAM was solved locally by Maximum Likelihood (ML) of the occupancy grid maps, and globally by EKF with feature-based maps, where each local grid map denoted a feature. Moving objects were modeled and tracked using the Interacting Multiple Model (IMM) algorithm. Data association was achieved with the Multiple Hypothesis Tracking (MHT) algorithm.

A consistency-based or motion-based detection method was used to differentiate between moving and stationary objects in occupancy grids. If an object was observed in a previously defined free space, then it was considered a dynamic object. If an object was detected in a previously defined occupied space then it was regarded as stationary. If an object appeared in a previously defined unknown space, then its mobility state remained undefined. In addition, any object found in a space with several moving objects, was classified as a potential moving object. A sufficient time period between successive observations was needed to detect temporary static objects. If the time period was not long enough, motions of such dynamic objects were too minor to be noticed.

Tests were performed on the Navlab11 vehicle at high speeds in crowed urban environments. Three laser scanners were mounted in different positions on the Navlab11 vehicle, to perform horizontal or vertical profiling, and produce 3D (2.5D) maps. Odometry was also used. Pedestrians, cars, bikes and buses were detected and tracked. [5, 6]

While Wang's approach was effective it involved the use of expensive laser sensor systems. Cameras offer an affordable solution and overcome some of the disadvantages of lasers in perceiving complex environments. [6]

Davison [10] laid the foundation equations for EKF SLAM using a single camera in a stationary environment. Most works are based on his work or offer extensions.

Feature initialization has evolved from delayed initialization [10], to undelayed initialization of partial feature observations with Federated Information Sharing (FIS) [11], to undelayed initialization of features at any distance through the use of Inverse Depth Parameterization (IDP) [12], to its successor, Inverse Scaling Parametrization (ISP) [13]. [3, 14]

Algorithm advances have allowed higher level applications involving multiple camera configurations in dynamic environments.

Lowe [15] demonstrated a stereo-based SLAM algorithm based on Scale-Invariant Feature Transform (SIFT) features in a dynamic environment. SIFT features remain unaltered regardless of image translation, scaling and rotation. [15] These attributes make them ideal for SLAM and tracking applications because features are measured over time from different viewpoints.

When features were detected, a new track was initialized with a miss count parameter set to zero. If this parameter reached a predetermined threshold, it meant that the feature was not detected in its initial location for a certain number of consecutive frames. This miss count assisted in detecting moving objects. The moving feature's old location was discarded and the new location recorded. Changing viewpoints and occlusions of features were managed by view vectors that recorded the view direction from which the feature was detected.

A Real World Interface (RWI) B-14 mobile robot, equipped with a trinocular stereo system (and odometry) acquired 3D coordinates of the features for localization. Least-squares minimization was performed for the confirmed features to estimate the robot ego-motion. A 3D map served as a database of SIFT features that indicated uncertainty of feature locations. Features were tracked with Kalman Filters (KF). Dynamic objects included a chair in a laboratory. [15]

Sola [16] introduced a probabilistic and geometric technique known as BiCamSLAM, where a MonoSLAM algorithm was applied to a stereo camera system. The benefits of mono-vision (Bearing Only (BO) measurements, with endless range, but without immediate 3D data) and stereo-vision (3D data within a restricted range) were combined to yield the following advantages in BiCamSLAM:

1) Key features for reactive navigation, which are near the robot, are quickly included in maps, using stereo-like triangulation.

2) Adequate orientation localization is obtained with BO observations of distant features.

3) Updates can be executed on any feature even if it is observed from only one camera.

4) Precise prior calibration of the stereo rig cameras' extrinsic parameters is not needed because, *5)* Dynamic self-calibration of the stereo rig extrinsic parameters can be performed, to provide a sensor that is more reliable. [9]

Sola divided the SLAM and tracking algorithms. FIS SLAM [11] was used for undelayed feature initialization. Feature detection and matching methods were combined to generate a map and camera pose. 2D point features were detected using Harris point detection. A patch surrounding the point was retained as the landmark's appearance descriptor, and the current robot pose was recorded. The Active Feature Search (AFS) method [10] was used for matching.

Similar to the AFS method, Sola used strategies to guide the detection of moving and static features. This increased detection rates and reduced computational costs. Strategies were based on system knowledge of where dynamic objects were likely to occur.

Depending on system conditions, new features were initialized as either dynamic or stationary. If a new feature was initialized as dynamic, it was given a constant velocity model with an initial estimate computed from some a-priori information. If the point was in fact stationary, its velocity would converge to zero, it would be classified as static and included in the SLAM map. Alternatively, if a new feature was initialized as stationary, and it was actually dynamic, the system would quickly reject it and, a search would be conducted for a dynamic object in that area.

A two-step algorithm was used to detect and classify moving and stationary objects. A Detection Mechanism (DM) was executed first with the aim of searching a specific region in the image where a feature was likely to be found (similar to a voting strategy). A Priority-based Detection Selector (PbDS) was then applied to choose: which areas would be searched, whether a dynamic or stationary point type would be searched for, and the parameters that would be used for the feature search.

A different EKF was used to track each detected dynamic object. A robot with a stereo head and odometry was tested in a laboratory. Movable objects in the lab included a person, table, bin, small box, trunk, fence and, white-board. A 3D feature-based map was built from the sensor data. [16]

Based on Sola's work [16], Migliore [17] demonstrated that a single camera can be used to solve the SLAMIDE problem and obtain the estimated trajectories of the dynamic objects. A MonoSLAM algorithm using ISP [18], in parallel to a BO Tracker exchanged information about the camera pose and the object motion. This increased the robustness of the SLAM algorithm and produced consistent estimation of the pose, map, and dynamic object trajectories.

The SLAM filter used only stationary point features to estimate the map and camera pose, while the BO Tracker handled dynamic features in the environment. The tracker tracked new features detected by the camera with independent EKFs, and classified them as static or dynamic. Dynamic features continued to be tracked and static features were passed to the SLAM filter. Inconsistencies were therefore reduced in the SLAM process.

Moving features were detected by a geometric analysis that was similar to epipolar constraint. The intersections between three viewing rays of the same point feature, observed from three different camera poses, were compared. The feature was classified as moving if, the intersections of the viewing rays were dissimilar during camera motion or were not present. Uncertainty of BO measurements and estimates were handled by Uncertain Projective Geometry (UPG).

Simulations with a dynamic object, and experiments in real indoor environments with a moving person demonstrated that the method maintained consistent estimations appropriate for a real-time application. Data association was performed manually. [17]

Similar to Migliore [17], Perera [3] implemented EKF MonoSLAM [10, 19] with ISP [13], in parallel to an EKF tracking system. Dynamic and static points were tracked separately by independent EKF trackers. Perera maintained that detected outliers should not be rejected as they may have information about the position of dynamic points which may assist in determining if a stationary landmark is occluded.

Like Wang's method [6], the SLAM posterior was estimated first and passed to the tracker so that the dynamic point posteriors could be estimated. The fundamental matrix, containing information pertaining to moving points (and/or rejected outliers) was estimated after the SLAM observation stage and passed to the tracker.

The tracker combined the epipolar constraint from the fundamental matrix and its observation stage in order to perform point classification. Newly detected features were assumed dynamic and if the mean distance of the epipolar constraint was below a manually adjusted threshold, the object was classified as stationary and included in the SLAM map.

A handheld camera was tested both in a simulated video sequence with a moving cube, and a real video sequence with a moving box in an indoor environment. The linear camera trajectory was correctly estimated through the use of tracking unlike traditional SLAM.

Perera's method was different from Migliore's [17] in the way in which moving points were detected. Perera made use of occluded information. [3]

In monocular Simultaneous Localization, Mapping and Moving Object Tracking (SLAMMOT), camera motions and dynamic object trajectories may result in dynamic objects being unobservable. This is referred to as the observability problem of BO tracking.

Wang [20], proposed the augmented state monocular SLAMMOT approach using IDP [21]. Dynamic object states were added into the SLAM state vector, and by tracking these dynamic objects the accuracy of SLAM was increased. The camera followed circular motions to avoid the observability problem, and attain converged tracking results.

Lin [22] extended this single camera SLAMMOT approach to a stereo-based method to prevent observability problems and hence improve the performance of SLAMMOT. Each camera acted as an observer and the measurements from both were updated by EKF.

Moving objects were detected based on the observation that incorrectly identifying a dynamic object as static and including it in SLAM would decrease the performance of SLAM. This would have a negative effect on camera and static feature estimates. Dynamic objects could therefore be detected by analyzing the outcome of monocular SLAM under two hypotheses.

When a new point feature was detected, two local monocular SLAMs were initialized, each with a hypothesis: One where SLAM excluded the new feature and the other where the new feature was included and assumed stationary. A new feature was regarded as stationary if the log odds ratio of the difference in the two hypotheses was greater than a specified threshold; else the feature was regarded as dynamic. The addition of misclassified dynamic objects in the SLAM state vector resulted in negative inverse depth estimates that also assisted in indicating dynamic objects.

Monte Carlo (MC) simulations demonstrated that stereo SLAMMOT provided better accuracy under unobservable circumstances than monocular SLAMMOT. Experiments were conducted indoors on a NTU-PAL7 robot, equipped with a Point Grey Bumblebee X3 stereo camera and a laser scanner (for ground truth). Features on a pedestrian were tracked under different scenarios.

The good feature approach [23] was utilized to extract features, and non-maximum suppression was used to produce a sparse feature set from images. The Kanade-Lucas-Tomasi (KLT) tracker [24] was supplemented by the AFS method [19] to track features. After the image feature trajectory on the first camera was acquired, the image features on the second camera were determined directly utilizing known correspondences from the stereo camera. Complicated camera motions were not required to execute tracking. [22]

In Lin's stereo SLAMMOT approach [22], modified IDP was used but feature initialization remained delayed. New feature measurements that were unclassified could not be directly used in state estimation. Hsiao [25] introduced a velocity classification method that enabled new feature measurements to be directly used for state estimation using a single camera.

Hsiao built on the work of Wang [26] and accomplished feature initialization with no delay in MonoSLAM with Generalized Objects (GO). In SLAM with GO the motion modes of the GO are deduced and a joint posterior including the robot pose, static objects and dynamic objects is estimated.

MonoSLAM with undelayed initialization via IDP has been shown to be practical. Undelayed initialization in MonoSLAM with GO, however, remains difficult as there is a delay in the classification decision of stationary and dynamic objects. Unclassified objects may not be used to estimate the state.

Hsiao achieved classification of stationary and dynamic objects by using the velocity estimates obtained from SLAM with GO. This allowed all measurements to be used with no delay in complete state estimation of SLAM with GO.

New features observed in the first image were initialized with modified IDP [27]. Features were processed as GO that were initially assumed unknown. With the next images, the estimated velocity distribution of the GO was used to compute static and dynamic object score functions. The score functions were compared to predetermined static and dynamic object thresholds to classify objects accordingly without delay.

MC simulations were carried out with static and moving points. Real experiments were performed in a basement with a moving person that came into the camera's view three times. A NTU-PAL7 robot was equipped with a Point Grey Dragonfly2 wide-angle camera and a SICK LMS-100 laser scanner (for ground truth).

The SLAM with GO algorithm estimated camera poses correctly and built a 3D feature-based map. Velocity estimate-based classification allowed for correct identification of features. [25]

V. DISCUSSION

Table I summarizes the techniques reviewed in section IV. Similarities and differences of these techniques are discussed. Limitations of the techniques are also mentioned. Based on the research a decision is made on the algorithms that will be used to allow a mobile robot equipped with two RGB-D cameras to perform SLAM and DATMO in an indoor dynamic environment. (Moving objects include people.)

As indicated in Table I, the majority of approaches used an EKF-based SLAM algorithm indoors, and built 3D feature-based maps. All the approaches used point features except for Lowe [15] who used SIFT features.

The EKF-based SLAM algorithm in [10, 19] will be applied owing to its reliability, low computational complexity, and suitability for camera applications. Point features will be detected as they can be regarded as SIFT features in a 3D environment. [16] A 3D feature-based map will be constructed.

As indicated in Table I, KF-based techniques were generally used to track features. Lin [22] used the KLT tracker supplemented with AFS. EKFs will be used to track features for their computational efficiency.

In the monocular SLAMMOT methods of Migliore [17] and Hsiao [25] unobservable conditions resulted in tracking inaccuracy and object misclassification, respectively. In Wang [20] the camera was made to follow complex trajectories to cope with unobservable conditions. As in Sola [16] and Lin [22] unobservable conditions will be avoided by using two cameras instead of one to estimate the complete state of features. This will also prevent complex camera trajectories.

Perera [3], Migliore [17] and Hsiao [25] used a single camera to perform MonoSLAM. Lowe [15] used a stereo-based rig and odometry with a stereo-based SLAM algorithm. Sola [16] offered an innovative solution where he used a stereo-based system and executed MonoSLAM on each camera to attain certain benefits (mentioned in section IV).

TABLE I. TABLE OF RESEARCHED SLAMIDE TECHNIQUES

Author Year	**Criteria**								
	SLAM Method	***Tracking Method***	***Sensors***	***Features***	***Moving objects***	***Test environment***	***Experimental platform***	***Map type***	***Contribution***
Wang 2004 (PhD Thesis)	Locally: ML-SLAM Globally: EKF SLAM	IMM	Laser scanners and odometry	Points	People, cars, bikes, buses	Outdoor	Navlab11 vehicle	3D (2.5D) Grid map	Pioneer of SLAM & DATMO.
Lowe 2002	Stereo-based with least-squares minimization	KF	Trinocular stereo system and odometry	SIFT	Chair	Indoor	RWI B-14 mobile robot	3D SIFT map	Stereo-based SLAM algorithm based on SIFT features in a dynamic environment. Feature viewpoint changes and occlusion were handled by view vector.
Sola 2007 (PhD Thesis)	EKF MonoSLAM with FIS	EKF	Stereo rig of 2 nominally equal cameras	Points	Person, boxes, table, bin, fence	Indoor	Robot	3D feature-based map	BiCamSLAM - A MonoSLAM algorithm was used on each camera in a stereo head.
Migliore 2009	EKF MonoSLAM with ISP	EKF	Wide - angle lens camera	Points	Person	Indoor	Handheld	3D feature-based map	MonoSLAM with Bearing Only Tracker. Used UPG to detect dynamic objects.
Lin 2010	EKF MonoSLAM with modified IDP	KLT tracker and AFS method	Point Grey Bumblebee X3 stereo camera	Points	Person	Indoor	NTU-PAL7 robot	3D feature-based map	Extended Wang's augmented state vector single camera approach with IDP to a stereo-based system.
Perera 2011	EKF MonoSLAM with ISP	EKF	IEEE 1394 firewire camera	Points	Box	Indoor	Handheld	3D feature-based map	Perera maintained that detected outliers should not be rejected as they may have information about the position of dynamic points which may assist in determining if a stationary landmark is occluded. They were instead tracked by EKF.
Hsiao 2011	EKF MonoSLAM with modified IDP	Not named	Point Grey Dragonfly2 wide-angle camera	Points	Person	Indoor	NTU-PAL7 robot	3D feature-based map	Velocity estimate-based classification method that allowed all measurements to be used without delay for state estimation.

Lin [22] also applied MonoSLAM on each camera in a stereo-head but without the use of odometry. Similar to Sola and Lin, a MonoSLAM algorithm will be applied to each RGB-D camera. Odometry will not be used.

Wang [6], Sola [16], Migliore [17] divided the SLAM and tracking posteriors to reduce computational complexity. Lin [22], Wang [20] and Hsiao [25] included all features in the SLAM posterior and then classified them as stationary or moving.

The first approach will be taken so that moving objects are identified by the tracker before they are included in the SLAM posterior. This will allow for lower computational complexity and may also prevent temporary stationary objects from being included in the SLAM posterior.

In Lowe's approach [15], a large part of processing time was dedicated to SIFT feature extraction. Parallel processing methods will be explored to allow for faster computation and better performance of the algorithms involved.

Hähnel [28] used a laser-based method that rejected outliers, thus information was lost. Perera [3] passed the outlier information to the tracker, to allow detection and tracking of dynamic objects and assist in the detection of occluded stationary objects. An approach similar to Perera's will be used. Features will be assumed unknown and passed to the tracker for classification. This way no information will be lost and occluded objects may be accounted for.

Sola's self-calibration solution [16] experienced low observability and inconsistencies. Drifts in the convergence angle of the two cameras (which was associated with feature depth) led to drifts in the map and vice-versa. Later, Sola [14] used IDP [27] to enable constant, real-time calibration. [16]

IDP has recently been outperformed by ISP. ISP increases the observation equation linearity allowing for faster convergence of the SLAM algorithm as compared to IDP. [3] It is preferred for feature initialization as in [3, 17].

Wang [26], Lin [22], Lowe [15] used a fixed number of frames to classify stationary and moving features. Hsiao's data-driven approach [25] (explained in section IV) allowed for object classification with a flexible number of frames. It prevented incorrect classification due to an inadequate number of updates and the computational effort required was lower when objects were simple to identify. A similar classification approach (explained later in this section) will be adopted to allow these advantages. [25]

Migliore [17], Perera [3] and Lin [22] experienced difficulty in tracking dynamic objects between images over time. This is because features on a dynamic object may change during acquisition. This hindered convergence of the tracker, hence, a correct estimate of the dynamic objects in the environment could not be attained. To overcome this problem Migliore proposed scene structure analysis. Perera and Lin recommended applying techniques with dense image features. [3, 17, 22]

The reliable tracking of moving objects is needed for moving object feature estimation and prediction, and will assist in enhancing the performance of SLAM. As can be ascertained from the work of [3, 17, 22] tracking of moving features in images over time is difficult and may lead to inaccurate results. It therefore remains an open question and further research needs to be conducted.

Based on [17, 29] a contribution will be made to reduce the difficulty in tracking dynamic points in images over time. This will be endeavored by grouping similar points and tracking these groups of points.

Points will be assumed unknown and passed to the tracker. The tracker will track and classify points by taking their velocity into account (similar to Hsiao's data-driven approach [25]). If a point's velocity converges to zero it will be classified as static and passed to the SLAM filter. Else it will remain in the tracker and continue to be tracked. The pose of the point may also be monitored as a further check.

Points will be grouped based on their behavior in relation to each other to determine if they belong to the same moving object. If points meet requirements to form a group, and the group remains consistent after a defined number of images, the group will be tracked.

Groups of points will be managed. They will be merged if they exhibit similar behavior, separated if they differ from each other and deleted if they no longer display similar behavior. [29]

Research will be conducted on techniques for grouping similar points and dense image features. Techniques will be compared and assessed and a suitable technique will be chosen to optimize the application.

VI. CONCLUSION

This paper researched the algorithms that have enabled vision-based SLAM to evolve from static to dynamic environments. The advantages of cameras over lasers have seen cameras as the preferred sensor for certain SLAM applications.

The development of reliable algorithms to perform SLAM in static environments with a single camera has allowed for mobile robot applications to move from static environments to dynamic environments with moving objects. The consistency of these SLAM algorithms has allowed the extension from one to multiple cameras in dynamic environments.

Stereo-based methods have prevented some of the disadvantages of using one camera alone, such as scale estimation and unobservable conditions, but offer a limited range as compared to monocular methods. The implementation of MonoSLAM algorithms on each camera in a stereo-head has eliminated the disadvantages of single camera and stereo-based methods, to produce increased observability and flexibility that allows feasible applications in dynamic environments. [16]

The improvement of feature initialization algorithms from delayed to undelayed initialization enables faster feature initialization. This reduced computational effort allows for more emphasis to be placed on complex applications such as SLAM and DATMO in dynamic environments.

The use of image features to detect and track moving objects has allowed correct feature classification of static and dynamic objects. This has allowed for increased performance of the SLAM algorithm and a reliable map for localization and task completion.

The tracking of moving features over time, however, presents certain difficulties that prevent accurate tracking results. Further research will be conducted on techniques that may reduce these difficulties by tracking groups of points over time and using dense image features.

VII. FUTURE WORK

Drawing from the research conducted on the algorithms for SLAM and DATMO, a decision was made as to which algorithms will be used for a mobile robot equipped with two RGB-D cameras, to perform SLAM and DATMO in an indoor dynamic environment. An EKF MonoSLAM algorithm will be applied to each of the RGB-D cameras. EKFs will be used to track features. A contribution will be made in the research and development of techniques that enable more effective and reliable tracking of image features over time. The feasibility of tracking groups of points to obtain accurate tracking estimates will be validated by simulation and real experiments. Parallel processing methods will be researched to allow for optimal performance.

REFERENCES

[1] C. C. Wang and C. Thorpe. Simultaneous localization and mapping with detection and tracking of moving objects. Presented at Proceedings of IEEE International Conference on Robotics and Automation (ICRA). 2002.

[2] D. Wolf and G. S. Sukhatme. Online simultaneous localization and mapping in dynamic environments. Presented at Proceedings of IEEE International Conference on Robotics and Automation (ICRA). 2004.

[3] S. Perera and A. Pasqual. Towards realtime handheld MonoSLAM in dynamic environments. *Advances in Visual Computing* pp. 313-324. 2011.

[4] L. Montesano. Detection and tracking of moving objects from a mobile platform. Application to navigation and multi-robot localization. PhD Thesis. University of Zaragoza. 2006.

[5] T. D. Vu. Vehicle perception: Localization, mapping with detection, classification and tracking of moving objects. PhD Thesis. Grenoble Institute of Technology. 2009.

[6] C. C. Wang. Simultaneous Localisation, Mapping And Moving Object Tracking. PhD Thesis. Carnegie Mellon University. 2004.

[7] G. Lidoris, D. Wollherr and M. Buss. Bayesian state estimation and behavior selection for autonomous robotic exploration in dynamic environments. Presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2008.

[8] T. D. Vu, J. Burlet and O. Aycard. Grid-based localization and local mapping with moving object detection and tracking. *Information Fusion 12,* pp. 58-69. 2011.

[9] J. Sola, A. Monin and M. Devy. BiCamSLAM: Two times mono is more than stereo. Presented at Robotics and Automation, 2007 IEEE International Conference on. 2007.

[10] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. Presented at Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. 2003.

[11] J. Sola, A. Monin, M. Devy and T. Lemaire. Undelayed initialization in bearing only SLAM. Presented at Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on. 2005.

[12] J. Civera, A. J. Davison and J. Montiel. Inverse depth parametrization for monocular SLAM. *Robotics, IEEE Transactions on 24(5),* pp. 932-945. 2008.

[13] D. Marzorati, M. Matteucci, D. Migliore and D. G. Sorrenti. On the use of inverse scaling in monocular SLAM. Presented at Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. 2009.

[14] J. Sola, A. Monin, M. Devy and T. Vidal-Calleja. Fusing monocular information in multicamera SLAM. *Robotics, IEEE Transactions on 24(5),* pp. 958-968. 2008.

[15] S. Se, D. Lowe and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The International Journal of Robotics Research 21(8),* pp. 735-758. 2002.

[16] J. Sola. Towards visual localization, mapping and moving objects tracking by a mobile robot: A geometric and probabilistic approach. PhD Thesis. National Polytechnic Institute of Toulouse. 2007.

[17] D. Migliore, R. Rigamonti, D. Marzorati, M. Matteucci and D. G. Sorrenti. Use a single camera for simultaneous localization and mapping with mobile object tracking in dynamic environments. Presented at ICRA Workshop on Safe Navigation in Open and Dynamic Environments Application to Autonomous Vehicles. 2009.

[18] D. Marzorati, M. Matteucci, D. Migliore and D. G. Sorrenti. Monocular SLAM with inverse scaling parametrization. Presented at Proc. of the British Machine Vision Conference, Leeds. 2008.

[19] A. J. Davison, I. D. Reid, N. D. Molton and O. Stasse. MonoSLAM: Real-time single camera SLAM. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(6),* pp. 1052-1067. 2007.

[20] C. -. Wang, K. -. Wang, C. -. Hsiao, K. -. Lin and Y. -. Chao. Monocular vision-based simultaneous localization, mapping and moving object tracking. Unpublished.

[21] L. M. Paz, P. Piniés, J. D. Tardós and J. Neira. Large-scale 6-DOF SLAM with stereo-in-hand. *Robotics, IEEE Transactions on 24(5),* pp. 946-957. 2008.

[22] K. H. Lin and C. C. Wang. Stereo-based simultaneous localization, mapping and moving object tracking. Presented at Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. 2010.

[23] J. Shi and C. Tomasi. Good features to track. Presented at Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94, 1994 IEEE Computer Society Conference on. 1994.

[24] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence (IJCAI),* 1981, pp. 674–679.

[25] C. H. Hsiao and C. C. Wang. Achieving undelayed initialization in monocular slam with generalized objects using velocity estimate-based classification. Presented at Robotics and Automation (ICRA), 2011 IEEE International Conference on. 2011.

[26] C. C. Wang, C. Thorpe, S. Thrun, M. Hebert and H. Durrant-Whyte. Simultaneous localization, mapping and moving object tracking. *The International Journal of Robotics Research 26(9),* pp. 889-936. 2007.

[27] J. Montiel, J. Civera and A. J. Davison. Unified inverse depth parametrization for monocular SLAM. *Analysis 9,* pp. 1. 2006.

[28] W. Burgard, C. Stachniss and D. Hähnel. Mobile robot map learning from range data in dynamic environments. *Autonomous Navigation in Dynamic Environments,* pp. 3-28. 2007.

[29] K. E. Ozden, K. Schindler and L. Van Gool. Simultaneous segmentation and 3d reconstruction of monocular image sequences. Presented at Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. 2007.