Chapter 2

Literature Review

2.1 Introduction

Mobile robot is a device that can move autonomously from place to place in order to accomplish a variety of objectives. Mobile robot works in a broad range of applications like searching and rescue [Magid et al. (2008)], security services (e.g., recognition, tracking) [Luo et al. (2006)], military operations [Peter (2015)], medical care [Wang et al. (2009)], transportation (e.g., autonomous and assisted driving) etc. Each of these applications has a number of conditions including time (real-time, off-line or on-line), occlusion, robustness and detection. Mobile robots are imparted with autonomy since these are required to perform tasks that seem difficult for humans.

The main focus of this research was on wheel mounted mobile robots, while various locomotive mechanisms such as crawling, walking, swimming and flying could be used [Lam et al. (2012), Ignatiev et al. (2016), Chhotray & Parhi (2019)]. Wheeled mobile robot can be categorized according to the number of wheels into two, three, four and six wheels, while it can be classified as difference drives, omnidirectional drive motors, synchro drives and skid steering. A detailed description and degree of freedom of mobile robot types were provided [Roland & Illah (2004)]. The mobile robot is the most common mobile robotics locomotive mechanism because of its efficiency and flexibility by the use of wheels and cheaper than its counterparts. The shortcoming of wheel is that it cannot navigate over obstacles, such as stony terrain, unsmooth surfaces [Sharma & Dwivedy (2012)]. The wheeled robots are being employed in large numbers not only in industries but also in service centric applications which include domestic and entertainment applications as well. Considering the numerous scientific publications published in this area in recent years, attempts have been made to investigate different mobile robotic navigation techniques. In this context the navigational competence of a mobile robot is one of the key problems in the development of an autonomous mobile robot. Researchers at the module and system integration levels encounter challenges because technology is extremely complex and multidisciplinary in the mobile robot system [Wätzoldt et al. (2012)]. A variety of effective frameworks have been developed for certain tasks such as localization, trajectory generation, local hazards or to safely bypass a known obstacle which is considered as a result of good navigation [Gupta et al. (2015)]. The above-mentioned common designs must be solved by each navigation system to make sure that every job can be completed.

The navigation of mobile robot tries to find the optimal path based on the data obtained from the sensors, representing a local map which can or cannot be part of a global map [Koi and Kim (2013), Gene et al. (2013)]. Till date, there is still no ideal navigation system and is difficult to compare the results of researches, since there is a huge gap between the mobile robots and the environment of each research. Robot navigation includes different interrelated activities such as perception and interpreting sensory information; exploration of the strategy that guides the robot to select the desired direction of movement; mapping, the spatial representation by perceived sensory information; localization to estimate the robot position within the spatial map; path planning to find a path towards a goal location being optimal and path execution, where motor actions are determined and adapted to environmental changes [Ramón et al. (2018)].

The mobile robot cannot correctly interpret the world around in certain dynamic environments due to the communication and sensory capabilities limitations of some conventional sensors [Pandey et al. (2019)]. Given the challenges, in recent years the scientific community has made much efforts to resolve these shortcomings which need to develop a more reliable and successful system. This is illustrated by technological advances in system architecture design, sensors and decision-making driven by advancement in related fields of artificial intelligence, communications, mechanical mechanism and networking.

This chapter addresses research and implementations of mobile robot's navigation on a number of important issues such as detection and tracking, trajectory planning, path planning and obstacle avoidance. To some extent, the process of path planning requires a-priori environmental knowledge. The first one determines the path to follow which is called as trajectory generation as usually the quickest and safest path. Trajectory planning is a rather independent process that takes place once the map has been built and the robot’s position is estimated [Dalla & Pathak (2015)]. Following the trajectory generated by a mobile robot is dependent on several factors, including the mobile robot's motion model (e.g. differential drive), its kinematic constraints and others. Actually following the trajectory generation by a mobile robot depends on several factors including the motion model (e.g. differential drive) of the mobile robot, its kinematic constraints and others. The localization must be used to allow the information of current position of mobile robot in a known or unknown environment.

The existence of the inaccuracies and the noise uncertainty in sensor measurements, which often lead to inconsistent maps, is a major problem for map building. Mobile robots can navigate in an environment where it has to construct its own map in the case of an unknown environment. Localization and map-building are interdependent processes where map is used to localize a robot when the map exists, while building a map it is required that the position should be estimated relative to the partial map learned so far.

Path planning is one of the important problems in the field of robotics. Over the years several path planning algorithm has been developed and used to fulfill the given tasks to robots, such as approaching the destination or avoiding collision with obstacles [Imen et al. (2017), Han-ye et al. (2018)]. A good path planner must be able to make a path to reach the destination efficiently and safely, and should be able to deal with both static and dynamic obstacles in the real time environment.

The mobile robots need information about the world so that robots can relate to the environment, just like animals. When compared with other on-board sensing techniques, vision based approaches to navigation continue to demand a lot of attention from the mobile robot research community, due to its ability to provide detailed information about the environment, which may not be available using combinations of other types of sensors. Vision-based indoor navigation for mobile robots is still an open research area as the rapid development of vision-based sensing for indoor navigation tasks took place in the earlier years. Mobile robots operating in an unknown and uncertain environment have to cope with dynamic changes in the environment and for a robot to navigate towards its goals while avoiding both static and dynamic obstacles [Wang at al. (2017)]. Most current techniques are based on complex mathematical equations and models of the working environment, however following a predetermined path may not require a complicated approach. To identify current state of the art and the directions a thorough literature review is presented below.

2.2 Autonomous Navigation

This section describes mobile robot navigation framework, organized as an environmental knowledge and the ability to drive the actuator and eventually generates robot motion. Path planning, localization, path execution, obstacle detection and avoidance are the common modules of mobile robotic navigation.

The navigation of mobile robot can be achieved by two systems; (i) coordinate based system, (ii) behavior based system. The coordinate based system uses the information of one's position inside a global coordinate system of the environment. It is based on models of the environment to generate paths to guide the robot. Some techniques are Mapping like topological map and a hybrid map [Bersan et al. (2018)], Occupancy Grid Navigation [Nikdel et al. (2020), Xianyu et al. (2020)] and Potential Fields [Apoorva et al. (2018)]. Work by [Trung et al. (2011)] represents the current state of the art /related work/ of grid-based map representation implementation. Grid-based map representation is easy to construct and maintain, and all measurement data is provided explicitly in the map. However, they suffer from high a significant complexity in space and time [Habib (2007)]. Grid-based map representation is easy to construct and maintain and all measuring data are explicitly included in the map. However, suffer from a significant complexity in space and time [Habib (2007)]. A topological map is a graph-like representation of the environment provides a number of benefits for mapping such as current position, target position and the trajectory between two positions [Varadarajan (2015)]. The reason for explanation is that the resolution of a grid must be sufficiently high to collect all essential world data. The quality and the model of sensors used in map update are rather dependent on performance. In addition, information is lost when grid cell measurements are assigned. The behavior based system allows the robot to understand the environmental characteristics through its sensors and use the acquired information to find its goal [Pandey and Parhi (2020)]. Since the map building is the basis of the robot navigation, it has therefore been an active field of robotic field using different sensors in the past three decades [Michaud and Nicolescu (2016)]. The mapless navigation systems are those that use no explicit representation of the space in which navigation is to take place and they employ to recognize objects found in the environment by generating motion commands based on sensor signals.

Many techniques have been studied in the process of integrating data from different sensors. The techniques are Neuro-fuzzy approach, Radio frequency techniques and Sensor network approach [Batalin et al. (2004), Chaochao & Paul (2012)]. Neuro-fuzzy approach is the combination of neural network and Fuzzy logic [Phinni et al. (2008)]. Radio frequency (RF) technique controls the mobile robots using mobile and wireless RF communication. Navigation using a sensor network is embedded in the environment which eliminates the need for a map or localization where sensor nodes act as signposts/beacons for the robot to follow. These methods translate the different sensory inputs into reliable estimates that can be used by other navigation systems. Obstacle avoidance is designed by using Fuzzy logic (FL) and artificial neural network (ANN) in robot navigation especially in real time implementation. It prevents the mobile robot from collision and damaging itself [Somia et al. (2019)]. Main issue in mobile robot is navigation in an uncertain and complex environment and considerable research has been done for making an efficient algorithm for the mobile robot navigation [Anish et al. (2016)]. Among them, adaptive control and behavior-based control are most popular control algorithms and these techniques drive the research in robot navigation area. Adaptive navigation control is a method using pre-defined equations that represent the robot's path to reach targets and show strong ability in well-known environment.

When dealing with mobile robot navigation, the main problem of collision avoidance is always encountered. So the main task of navigation of mobile robot is to find collision-free trajectory from a start to goal configuration. The optimal solution to prevent a collision can be defined by the planned path. The real time implementation of a mobile robot navigation schemes are used for path planning and obstacle avoidance where vision sensing is employed as primary sensor for path planning in a static or dynamic environment. Sensors like infrared, Sonar are used as secondary sensors for real mobile robot navigation with obstacle avoidance capability. The navigation strategy has been described based on prior environmental information needed for path planning. Without any a priori knowledge of an environment, it is almost impossible to determine the true shortest path for navigation among all possible paths. It is potentially possible to determine such paths by employing standard graph-search techniques, such as Dijkastra's algorithm [Sun et al. (2006), Wang et al. (2011)] and A* algorithm [Guruji et al. (2016)], D* algorithm [Chu et al. (2012), Saranya et al. (2016)] etc. A mobile robot computes a map of a previously unknown environment while localizing itself within that map is referred to as Simultaneous Localization and Mapping (SLAM) problem. Vision-based mobile robot's simultaneous localization and mapping navigation has been the source of various research contributions due to its rich sensory performance and vision sensor cost efficiency [Davison (2007)]. It is generally considered to be one of the most challenging problems in mobile robotics and one of the most significant that must be solved before truly autonomous robot become a reality.

2.2.1 Sensor Based Navigation

Sensors are a crucial component of a robotic system because they provide the information required to the system of its own status and to communicate properly in surrounding environment. Nowadays, mobile robots are equipped with several sensors and other accessories which add to the sophistication and flexibility and help in developing overall capability and intelligence of the system. While dealing with mobile robot navigation, sensors are usually used for positioning and obstacle avoidance. There are different sensors, such as infrared sensors (IR), ultrasonic sensors, sonar sensors, RFIDs, GPS, inertial sensors, laser range finders, vision sensors, which significantly differ with regard to accuracy, precision, cost and technology used.

IR sensors are also known for their non-linear behavior and their reflectance dependency on the surface of a target [Benet et al. (2002), Gorostiza et al. (2011)]. In addition, ultrasonic sensors can be used in opaque / not transparent conditions where optical sensors are difficult to operate [Kim et al. (2007)]. These sensors are of low cost, low power consumption and low computational complexity. However, ultrasonic sensors are not very accurate to test [Alonso et al. (2011)]. Sonar sensors are computationally affordable, and their data are simple to read, but the reliability of their data is low due to the environmental disturbances. Laser range finders provide better reliability, instantaneous measurement, superior range accuracy, and precise angular resolution than sonar, with finer directional resolution, but at much higher cost [Guo et al. (2009]. The laser range finder has a disadvantage that the scan may be prone to missing transparent objects, such as glasses and windows [Jung et al. (2020)]. Inertial navigation sensors such as accelerometers and gyroscopes provide orientation and trajectory measurements of the mobile robots, but provide no information about the obstacles in the environment that the robot is traversing [Qazizadaa and Pivarčiováa (2016)]. These sensors offer positional estimates but are prone to drift, leading to a navigation error that increases linearly with time. This unbounded error growth makes it difficult to operate autonomously over long-term with internal devices only.

A vision system is considered as a passive sensor and has the fundamental advantages over the active sensors such as infrared, laser, and sonar sensors. Passive sensors such as cameras do not alter the environment by emitting lights or waves in acquiring data and also the obtained image contains more information than active sensors [Hebert (2000)]. All these sensors acquire less information about the physical environment than a camera can capture and with the continued growth of faster and cheaper computing power, that potential is now being used for designing real-world vision based navigational systems. Hence vision as a sensing mechanism for mobile robots offers very attractive potential for solution [Arturo et al. (2010)].

Vision based Navigation is performed by vision sensors which can gain a wide range of information about the environment, including color, texture and other visual information, compared to other conventional sensors [Yihuan et al. (2015), Yuncheng et al. (2018)]. In modern robotic systems these sensors have a well-deserved area of research. This research focus is solely on visual navigation does not prohibit the use of other sensors. Vision sensors are cheaper and simpler to install, which makes vision-based navigation a main research point [Haythem et al. (2011)]. The vision sensor used to estimate the trajectory or the path of the mobile robot to overcome some limitations (drift due to noise and bias) can be categorized into various camera types such as stereo, monocular, omnidirectional and RGB-D. Usually the researchers are using the following standard vision sensors: (a) fish eye cameras [Courbon et al. (2007), Komatsu et al. (2020)], (b) stereo vision [Shan and Wang (2013), Sankowski et al. (2017)], (c) an RGB-D [Payá et al. (2017), Shuai et al. (2017)] and (d) monocular camera [Chang et al. (2012), Hwang and Song (2011), Shishira et al.

(2020)] respectively. Monocular vision systems are affected by the inconsistency of image scale where the surface is irregular or uneven, then the image scaling factor to estimate is difficult [Villanueva-Escudero et al. (2014), Aqel et al. (2016)]. The representation of the environment is another important issue for perception of static or dynamic environment. [Geiger et al. (2013), Puente et al. (2013)] used stereo Vision or LIDAR sensor for acquiring 3D data and visual odometry for ego-motion estimation.

2.3 Object Detection and Tracking

Object detection, a basic condition for initializing a tracking procedure, refers to locating the object of concern in the frame of a video or successive image sequence. Image sequence object detection and tracking continues to remain a computer vision problem because of various factors such as occlusion, dynamic movements, changes of location, variations of lighting and background clutter. While considerable progress has been reported in recent years and certain established techniques are currently included in many consumer electronics or incorporated in mobile robot navigational technologies, still far from achieving human efficiency, particularly as regards the open-world situation. Object detection and tracking behavior has been widely studied for many years since the subject is intensively used in computer vision applications such as automated surveillance, vehicle navigation and autonomous robot navigation [Dong and Lin (2010), Hashmi (2016)]. Most common approaches for vision based object detection and tracking rely on high-resolution images and matching to a priori known feature descriptors, both in controlled indoor and outdoor scenarios [Ohnishi and Imiya (2013)].

2.3.1 Static Object Detection

This involves locating objects in the frame of a video or successive image sequence. Since tracking is the problem of estimating the path or trajectory of an object in the image plane as it moves around a scene, therefore every tracking method requires an object detection mechanism either in every frame or when the object first appears in order to determine its behavior. It is clear that the navigation algorithms cannot be developed in isolation from object detection, as the action of the path tracking is inherently related to the object detection.

The preliminary object detection sets the framework for much of the research model in terms of training and evaluation procedures and classification techniques. The detection of objects has been demonstrated as a prominent module for various important applications which consist in determining the scale, location, pose, if any, present in an image or videos. There are typically two approaches widely used for object detection to initialize a tracking process:

(i) Manually locate the object in the first frame and allow the device to detect features like edges to track the object in the next frames. Feature-based detection approaches represent environment that employ a variety of geometric attributes, such as points, planes and corners and express those in a semantically precise map. The feature based detection output is a number of key points which indicate the position in the image with the corresponding scales and orientations.

(ii) Object is automatically detected using a collection of features like colors, motion, edges.

In Addition to the above two approaches, object-detection techniques are divided into four categories: point based, background based, segmentation based and supervised for object detection. Point based object detection detectors [Amin & Dzulkifli (2013), Bay et al. (2006)] uses an approach based on the center point, which is faster and more accurate than other bounding box. Properties like object size and position are extracted from features of the image at the center location. Background Subtraction [Kim and Kwon (2015), Thierry (2014)] in object detection techniques is commonly used in video sequences having static background that separates the image in the foreground. The foreground may contain of static or moving objects such as moving people, cars while the background contains static objects, like road, building, trees, stationary cars, etc. In this technique, when objects of interest are not found in the scenario, a reference background image is capture first. By subtracting the current image frame from the reference background image, the moving object is detected. Segmentation based [Keiichi and Morimichi (2001), Vese (2003), Wang et al. (2007)] object detection provides a complete description of the scene using dynamically evolving decompositions that explain each pixel both semantically and geometrically.

It is a very challenging task due to various factors such as highly complex background, arbitrary object orientations, high input resolution, etc. In higher-level tasks, most errors like tracking of object are often caused by incorrect detection. In addition, many problems arise for object detection and tracking, just to mention a few: due to the occlusion of the object to the scene, complex object motion, the specifications for processing in real time, and the object's incorrect shape.

2.3.2 Detection of a Dynamic Object

This can be acquired by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are targeted for further processing. In addition, the foreground objects were separated from the background by the image pixel intensities. Object detection can also be performed by learning different object views automatically from a set of examples by means of supervised learning mechanism. There are several common object detection methods described in [Mozos et al. (2007)]. From available object detection methods, the Viola Jones detection algorithm is robust. The success of the Viola-Jones detector has been illustrated in [Peleshko & Soroka (2013)] paper where the author used Haar-like features. [Felzenszwal et al. (2010)] had suggested that the HOG feature extracted has better detection rate with support vector machines (SVM) classifier for training and combine this with the Cascade classifier. The HOG feature has been widely recognized as a successful feature extraction method for object detection, which can be successfully implemented for real-time detection comparable to the rectangular Haar-like detectors used by Viola-Jones. Results from [Tian et al. (2013)] suggest that HOG features are more accurate than the Haar-like features found in Viola-Jones.

A unique method for the detection of moving objects in surveillance applications was proposed by [Elham and Davud (2014)] which combine adaptive filtering technique with the Bayesian change detection algorithm. [Guo et al. (2012)] proposed an approach to object detection in video frames to track objects. The results of the simulation show that this technique was efficient and accuracy, robust feature encoding for object detection and good performance of classifier for generic object classes detection. It is also important to concentrate on increasing precision in the classification of real time objects.

Due to the complexity of the environment and the noise effects on the images are high and lead to poor quality, low resolution and the object is occluded.

2.3.3 Object Tracking

Object tracking is the process of locating an object or multiple objects over time using a camera [Verschae et al. (2015)]. For moving object, there are primarily two sources of information: visual features (e.g. color, texture and shape) and motion information which depends on the elements observed in the environment. The problems in object tracking can arise due to abrupt object motion, changing appearance patterns of the object and the scene, non-rigid object structures, full and partial object, complex object shape and camera motion [Chen (2014)]. The solution to these problems depend on the environment in which the detection and tracking are performed and the end use for which the tracking information is being sought. The problem can be simplified by using the vast knowledge about the moving object appearance, shape and size. The research area pertaining to moving object detection and tracking is closely related to other lines of research that analyze mobile robot motion and recognition from images and video [Tao and Yu (2009), Li et al. (2010), Kryjak and Gorgoń (2011)]. Optical-flow-based solutions estimate features or the motion of objects within a real image sequence [(Güzel (2013), Ohnishi and Imiya (2013)]. Here the differences in optical flow of mobile robot are computed from the real-time images [Jung and Sukhatme (2010)]. The accuracy and efficiency of gradient based optical-flow method is better in comparison with other optical-flow based solutions [McCarthy and Bames (2004)]. The focus is on detection and tracking and do not explicitly consider context such as the environment, interactions among objects. The motivation in studying this problem is to create a visual monitoring system with real-time moving object detection, tracking and activity analysis capabilities. In this work, the prime aim is to locate the position of the mobile robot in a video which is basically an object detection process to find the presence of object in a 2D frame. A webcam based system to detect moving objects and to generate accurate track has been developed. After setting up the system the problem associated with automatic tracking due to variation in distances between the camera and the objects in different parts of the scene are solved.

The present survey has analyzed 146 articles, from which 45 from IEEE digital library, 23 from Elsevier, 28 were from springer digital library, 15 from Science direct, and 35 from Google scholar (Figure 2.1). The following are discussed from those articles: object detection and tracking, static and moving object tracking methods specifically as a video sequence.

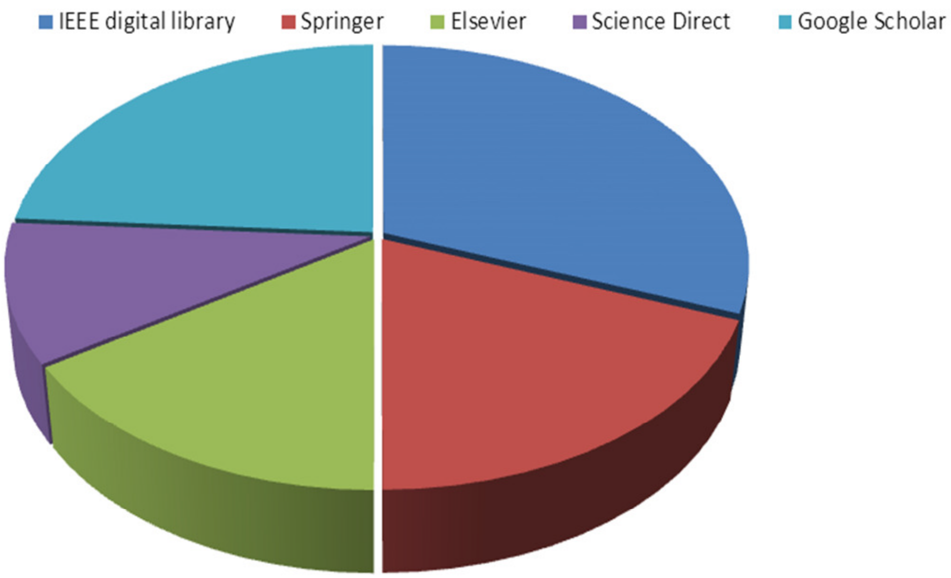

Figure 2.1: Number of research article referred

2.4 Navigation using Vision Sensor

Advancement in camera sensing and computer vision technology have been a very active research field within the robotics community with developments in object detection using monocular vision, stereo vision and sensor fusion with vision. Vision sensor is thought to be important to mobile robots for improving robustness, cost savings and reduced power consumption. The detection and tracking of static and dynamic objects using vision sensor has been explored, with an emphasis on published research paper after the year 2000. In addition, the focus is on the use of monocular and stereo cameras for above purposes. Researchers have been working on these aspects considering the real-time constraints of navigation techniques that are susceptible to noise and changes in dynamic environments, as it only considers current measurements.

There have been wide arrays of approaches for the development of a robotic framework for object clean-up in different situations which differ on the use of sensors for scene understanding and path planning. In [Ma et al. (2008)], RFID tags have been utilized for robot localization by dividing the room using a squared tile approach. However, over larger areas, placing multiple RFID tags will be expensive and cumbersome. On the other hand, [Buiu (2008)] discusses the use of ultrasonic sensors for obstacle avoidance and use of stereoscopic vision for creating a 3D virtual map for navigation. In [Hamel and Kress (2001)] tele-operated robots are implemented which are used in the 3-D modeling of structures using stereo and laser imaging. However, it places a high focus on the use of a human-robot interface for error correction in the 3-D models. The proposed idea focuses on a completely autonomous solution to the problem of object clean-up with minimal to zero human intervention.

Vision-based path planning has been highly successful for global-path planning approaches. A 3D panorama has been projected onto a 2D map and pre-processed further to form a unified obstacle map [ElHalawany et al. (2012)]. A modified depth-first A* algorithm is produced for optimal robot path planning for NASA Planetary Rovers. In [Gavrilut et al. (2006)], the authors have discussed the use of vision-based cellular neural networks for real-time path planning. However, the efficiency of this approach has not been compared with standard probabilistic or deterministic path-planning methods. In [Ziaei et al. (2014)], a global overhead camera is used to capture data and plan the path using artificial potential fields. A model-dependent CAD approach has been used for detecting the 3-D pose of an obstacle. However, this can be avoided by using camera calibration for finding the pose of objects which are moving in the same plane.

Present work proposes a multi-target A* approach for completing the task of obstacle/object clean-up. There has been a great focus on previous work for using a modified A* algorithm for improving the efficiency of the path planning process. In general, the practical implementation of A* algorithm treats obstacles as point objects, which does not capture the full dimensions of the obstacles. In [Goyal and Nagla (2014)], the size of the obstacles has been increased by (2n+1) to capture the obstacle dimensions in a better way. However, for small sized maps, this may close down many potential paths and lead to failure in detection of an optimal path to the goal.

All these approaches are only focused in single-target cases in general and do not consider the case of multi-target path planning for manipulation. The authors [Berenson et al. (2008), Gochev et al. (2012), Wolfe et al. (2010)] work on modeling the problem of mobile manipulation by combining path planning and robot manipulator configuration space and then searching for the optimal configurations to grasp different objects. A combination of a co-evolutionary algorithm and two directional RRTs is used [Berenson et al. (2008)] to find optimum grasp and robot orientations. A heuristic search based algorithm with adaptive dimensionality is used to increase computational gains for mobile manipulation tasks [Gochev et al. (2012)]. The [Wolfe et al. (2010)] focuses on using a hierarchical planning approach which has been identified as a problem of optimization.

With the growth in RGB-D sensor technology, robots have been able to learn spatial data and recognize objects more efficiently. In [Ferreira et al. (2014)], the author used RANSAC based approach to separate planes from a 3-D point cloud and utilized these measurements to find the volume of boxes. Whereas literature [Wong (2017), Wong et al. (2013), Scholz et al. (2015)] implemented the notion of modeling object characteristics using their type, pose and metrics. Semantic information [Wong (2017)] is implemented in an object by applying low-level features from object detection. The Dirichlet mixture model [Topkaya et al. (2013)] has been developed for detecting and tracking unknown number of objects metrics and spatial information is considered an important factor for mobile manipulation. The author [Wong et al. (2013)], discussed an approach for mobile manipulation of occluded objects using Bayesian inference models to detect different object arrangements based on metric information gained from previously trained models. On the other hand, [Scholz et al. (2015)] focuses on a physics-based online approach for mobile manipulation planning by estimating the object dynamics. A physics-based reinforcement learning and physics-based regression approach have been utilized to estimate the object dynamics.

Mobile robot navigation in the real world involves dealing with several challenges as the environment is an inherently uncertain and dynamic. So accurate models for path planning are difficult to obtain and time for application is usually very limited. Path planning algorithm with collision avoidance technique not only needs to detect the obstacles but also perform quantitative measurement which concerns dimensions of the obstacles. The mobile robot path planning can be divided into two main categories which are local path planning and global path planning [Ziaei et al. (2014)]. Global path planning requires complete information of obstacles with the environment to generate complete path from start to goal even before the robot starts motion.

Use of vision sensors for detection and tracking has been discussed in literature for various implementations [Qingchang and Qiao (2015)]. For computer vision, camera calibration is required to relate the pixel coordinates to the real-world coordinates. It is necessary to calibrate the vision system if the real-time scene is used for measurement applications. In robotic applications, most of the calibration procedure is to accurately measure the position coordinates and the camera image pixel operation which quantifies an object based on a number of images [Qi et al. (2010), Zhang (2016)].

In this work, the required information from each image has been extracted to calculate the correct values for the parameters to identify the obstacles in the real-time environment. The goal of image segmentation is to divide an image into multiple segments in order to simplify further image analysis such as feature extraction, object detection and object recognition [Ostovar et al. (2018)]. Thresholding (T) is the simplest segmentation method which separate out regions of an image $f(x, y)$ corresponding to objects that can be analyzed. Thresholding is distinctly a solution where gray image can be partitioned based on threshold value ([Židek and Hošovský (2014)]. The image is converted into binary image based on whether the image pixels fall below or above the threshold value (T). This separation is based on the variation of intensity between the object pixels and the background pixels [Li and Birchfield (2010), Xie and Lu (2013)]. If the distance in both the x and y direction between two pixels is lower than a specified threshold, these pixels are considered as part of the same object. Erosion is a set of operations that process images based on shapes where the minimum pixel value is computed. With this procedure, the areas of dark regions grow in size and bright regions are reduced [Jawas & Suciati (2013)]. This approach does not provide effective segmentation of the image and hence limits the classification of the various objects of the image if the image contains multiple regions or parts or color [Zhang et al. (2008)]. Clustering algorithms can operate on gray-tone images, color images, or multi-spectral images, making them easily adaptable to the computer vision domain [Yuheng and Hao (2017)]. The clustering algorithm partition 'q' no of images into 'p' clusters of smaller parts and each image belongs to the corresponding clusters with the centroid, mean intensity and area so that the overall processing time is significantly reduced [Ren et al. (2014), Sharma and Suji (2016)]. The result of clustering algorithm provides a coarse separation of object which is included in a scene [Bailer et al. (2010), Dhanachandra et al. (2015)].

Numerous researches on navigation using vision sensors have been proposed in recent time [Dayoub et al. (2013), Mane and Vhanale (2016), Kanellakis and Nikolakopoulos (2017)]. If the environment around the navigating mobile robot is dynamic, then some information like map data about environment for robust and reliable navigation are needed. The robot may get the information about 2D and 3D environment using vision sensor. While working in a dynamic environment the sensor updates the environment information and update the path of the mobile robot during navigation [Baslan et al. (2018)].

The map of the environment can be of various kinds. In the case of metric maps, the environment is divided into square-shaped cells of equal size [Gatesichapakorn et al. (2019)]. Upon this the obstacles may be defined as black cells while free space is defined as blank cells. These maps require extensive computational time and relatively large memory for their generation. The second kind is the topological map (graph-based), it is relatively better than the metric map in terms of computational time consumed and memory consumption but on the other hand, it is more difficult in constructions; also, the position of the robot cannot be located specifically on it [Kamarudin et al. (2014)]. The general approach is that the goal is selected based on the information gain from the surrounding of the mobile robot and then the path is planned to reach the target location.

Path planning plays a key role for the navigation of mobile robot and path planning algorithms are used to find the optimal path which is decided on various factors such as dynamic or static obstacles, illumination etc. Most of times there are multiple ways to reach the goal point. Among all the variants of path planning algorithms, D* lite algorithm is a fast path re-planning algorithm which adapts to unknown or moving obstacles in the environment [Koenig and Likhachev (2002), Stachniss et al. (2008)].

For tracking a mobile robot, the robot has to be detected first in the environment. Object tracking is a challenging task for researchers with respect to some constraints such as object deformation, noise in video or image sequence, motion blur, illumination change. Locating a target object in each video frame image accurately and robustly requires a range of information including object specifics and the background surrounding target object. The aim of object tracking is to estimate the object location in the image sequence when initial position of the object is given in the first frame. Whereas tracking efficiency depends on the implementation of algorithm, in different platforms such as MATLAB or Python. The platform based limitations are avoided using ROS open source application discussed and examined in this work.

In this work, tracking-learning-detection (TLD) based frame work has been discussed. This framework decomposes the long-term tracking task into three sub-tasks: tracking, learning and detection [Xiaoyan and Dan (2017)]. Once the vision sensor detects the target within its view range, it can track the mobile robot in real time for a long term and move towards the target precisely. There are few papers discussed about robustness of TLD algorithm of fast motion of selected object and illumination changed [Kalal et al. (2010), Hu et al. (2013)]. Detector performs full scanning of the image with respect to past observed appearances. Learning observes performance of both, the tracker and the detector which identifies errors of the detector and generates training examples to avoid the errors. Here the tracker and detector are working simultaneously which depends upon learning module to update the estimated location of the object.

Open source software overcome from complex software problems by enabling developers to freely add and modify the source code which can be reused on different hardware platforms. ROS applications communicate using publish-subscribe architecture, where the goal is to include various functionality such as object localization, path planning, object recognition [Murat et al. (2017)]. The ROS is composed of reusable libraries that are designed to work independently.

In this work, the mobile robotic system uses ROS framework for solving path planning and tracking problem. It is driven by data streamed from the vision sensor and different hardware abstraction, low-level device control components of the mobile robot process the data in different ways. The center of the system is a user controlled graphical interface for providing feedback and control of the components [Estefo et al. (2016)]. In ROS, nodes which are called as program can communicate with other nodes by publishing/subscribing the desired topic [Cousins (2010)]. There are three essential components include in a node such as publishing, subscribing and topics. By creating a ROS-launch file that contains all the nodes, all the process can start at the same time.

In practical tracking applications process, the characteristics of both standstill objects and moving objects will be clearly seen in video sequences; thus generating noise because of the low light effects video sequences. The Kalman filter and improved Kalman filter method are an estimation tool, which use the dynamic description of the mobile robot along with recursive least squares estimation for processing measurements [Chengjian et al. (2015)]. Robot location states and field parameter estimation using Kalman filters algorithm uses the time update (prediction) and the measurement update (correction) [Li et al. (2010)]. Time update is to advance the state based on state equation until the next measurement is obtained. Measurement update is to incorporate the measurement from sensors based on measurement model. The mobile robot is supposed to move in a 2D coordinate system. Thus, Kalman filter estimate current noisy frame (x, y) by comparing with previous de-noised frames of object to be tracked. Since the standard Kalman filter uses one correct measurement, data association should be considered to classify true measurement and false measurements [Das et al. (2019)].

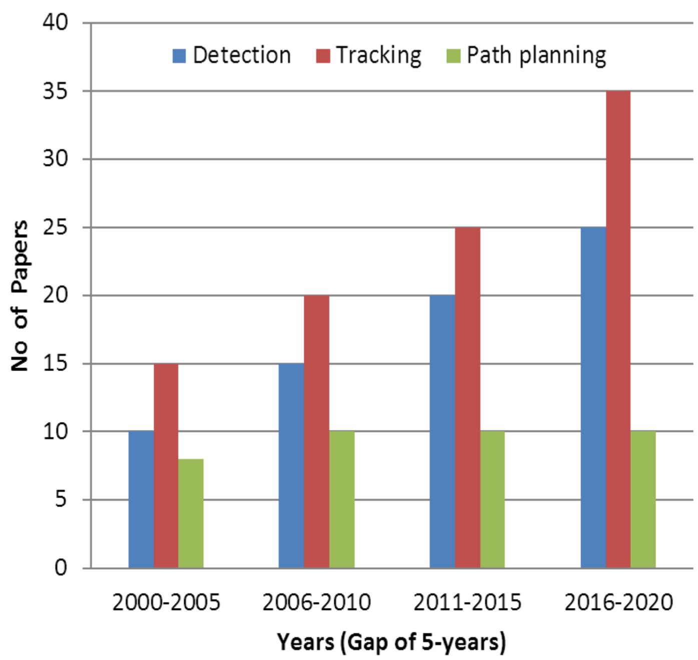

Figure 2.2: Reviewed article on navigation of mobile robot using vision sensor

The image in the foreground is segmented by a simple feature based object tracking technique for any object of interest. Then the variations in features for any consecutive frame for different directions of movement are determined. The location of the object in the next image frame is selected, which satisfies certain threshold conditions. The observation is also improved by local image statistics, primarily color histograms [Lee et al (2012)], to provide a wealthier description of the object detection. Histograms are often substituted in recent work by the online learning of unequal models of local texture. These models are effective when similar objects are near, particularly when a certain object is trained to differentiate from all other objects on the scene. A HSV color histogram is found to be appropriate for detection and tracking applications even though it monitors different objects in a dynamic scenario simultaneously.

2.5 Denoising Approaches and Overview of Kalman Filter

Kalman filtering is an algorithm that gives the measurements observed over time with an estimation of several unknown variables. Kalman filter is said to be an optimal estimator [Chen (2012)], due to its ability to calculate the error covariance of a system optimally by using the prediction method in a recursive manner to improve the system measurements over time.

In various applications such as robot localization [Zafari et al (2019)]; robot control [Ha and Park (2010)]; object tracking [Gunjal et al. (2018)]; path following [Sarra et al. (2018)]; visual navigation [Dao et al. (2005)]; visual servoing [Janabi-Sharifi and Marey (2010)]; data estimation and prediction where Kalman filters have demonstrated their usefulness, since it is reasonably easy and requires less computational power. While there is excellent literature [Simon (2006)] dealing with derivations and theory behind the Kalman filter, this section focuses on a more practical application perspective. The Kalman filter produces optimal estimation in state space formats based on linear dynamic systems with Gaussian additive noises. Navigation of mobile robot with a vision sensor/LIDAR will be provided as an implementation example of the Kalman filter. From the point of view of motion planning, it is necessary to define the outside environment, where relative distance measurements are needed to control a mobile robot. An example for implementing the Kalman filter is navigation where the mobile robot state, position, and velocity are estimated by using sensor output from an encoder unit and a vision sensor. States needed for robot self-location, map making, path computing, motion planning, and motion execution have been thoroughly studied in particular. For perception of remote perception, Kalman filtering technology used for vision-based navigation systems is useful.

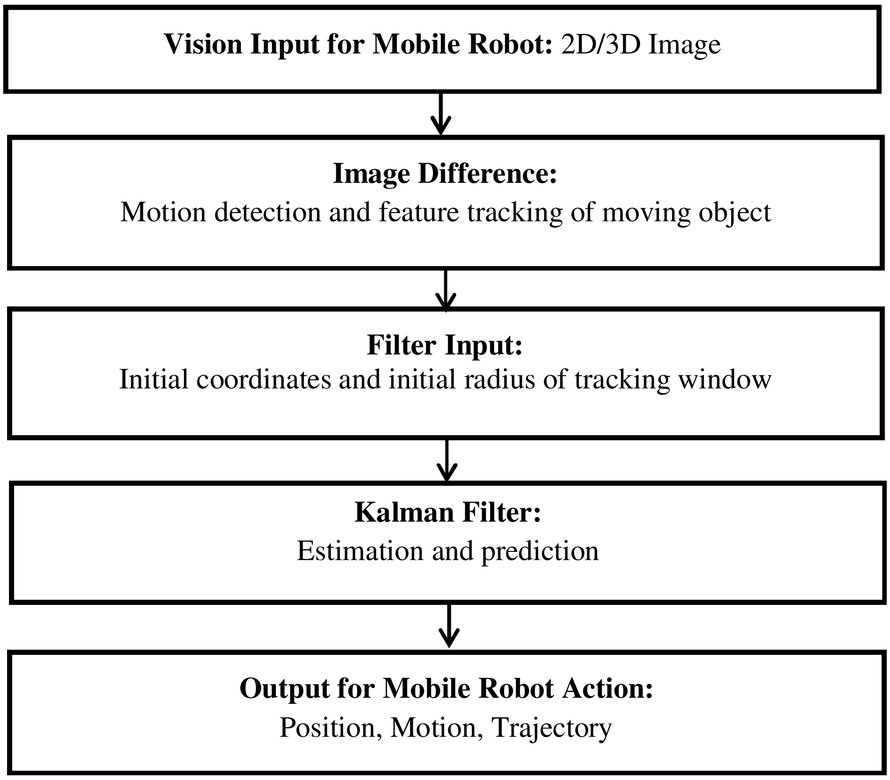

Figure 2.3: General procedure of the application of Kalman filters using vision sensor

Object tracking is the problem of estimating the position and other relevant information of moving objects in image/video sequences. In combination with the proposed detection and tracking algorithms, the Kalman filter has chosen by [Ahmed (2009)] to increase the estimating positions of noisy-occluded environment. The general procedure of such applications can be described using Kalman filter shown in Figure 2.3. The commonly used methods for solving vision based navigation for mobile robot problems are such as Kalman filter [Ali and Hassan (2014)]; Self-tuning Kalman filter [Macias and Exposito (2006)]; Steady-state Kalman filter [Deng et al.(2005)]; Adaptive Kalman filter (AKF) [Fakharian (2011)]; fuzzy Kalman Filter [Tseng et al. (2016), Khondker et al. (2018)];

Multi-dimensional kinematic Kalman filter [Lan et al. (2015)]; kinematic Kalman filter [Jeon et al. (2009)]; fuzzy logic controller Kalman filter [Ujwal et al. (2011)]; EKF [Choi and Lee (2010)]; Hybrid EKF [Cong et al. (2008)]; Modified covariance EKF [Xujiong et al.(2009)]; Iterative adaptive EKF [Zhang et al.(2006)].

Many researchers have been working on other localization issues, for example, [Miljković et al. (2015), Stephen et al. (2016), Ayadi et al. (2018)] authors implemented a new architecture that uses one SLAM monocular device for monitoring the mobile robot's uncontrolled motion. The main purpose of SLAM is to measure the position of the robot and the model of the surrounding map together. It plays a major role for robotics field and in particular a mobile robot field.

The Kalman filter generally derived using vector algebra as a minimum mean squared estimator. The Kalman filter [Welch and Bishop (2006)] assumes that the random process which has to be estimated is of the form:

$$\dot{x} = Fx + Bu + Gw \tag{2.1}$$

where, x-State vector which contains the system parameters of interest like position, velocity values; u-Control vector containing any control inputs like steering angle, braking force values; F-State transition matrix which applies the effect of each system state parameter at time t-1 on the system state at time t; B- control input matrix which applies the effect of each control input parameter in the vector u on the state vector; w- white noise vector containing known covariance.

When measurements are taken from the process at discrete moments in time, they occur according to the following relationship:

$$z = Hx + Du + v \tag{2.2}$$

where z- noise sample vector of measurements; H-transformation matrix that maps the state vector parameters connection between the measurement domain and the state vector; D- direct transmission vector of the input to the output, v-vector containing the measurement noise terms for each observation in the measurement vector.

This process can be modeled discretely in the following form, assuming there is not control inputs u to the system.

$$x_{k+1} = \Phi_k x_k + w_k \tag{2.3}$$

$$z_k = H_k x_k + v_k \tag{2.4}$$

The system error is defined as:

$$e_k^- = x_k - \hat{x}_k^- \tag{2.5}$$

where $\hat{x}_k^-$ - best estimate prior to receiving a measurement at time t_k

The error covariance matrix at this time is:

$$P_k^- = E\left[e_k^- e_k^{-T} \right] = E\left[\left(x_k - \hat{x}_k^-\right)\left(x_k - \hat{x}_k^-\right)^T \right] \tag{2.6}$$

where $E(\bullet)$ represents the expectation.

Now a linear blending of both the estimate and the measured value is taken.

$$\hat{x}_k = \hat{x}_k^- + K_k\left(z_k - H_k \hat{x}_k^-\right) \tag{2.7}$$

where $\hat{x}_k$ is the new updated estimate, Z is measured value, and K is a weighted value that determines the amount of error between the measured value and the best estimate.

By writing a posteriori state estimation as a linear combination of a priori estimation and the difference between actual measurement and measurement prediction weighted by kalman gain (K) is calculated to minimize a posteriori estimation error covariance. This gain is referred to as the Kalman gain which is capable of changing value over time. Now looking at the error covariance of this new updated estimate, which sets the following equation:

$$P_k = E\left[e_k e_k^T\right] = E\left[(x_k \hat{x}_k)(x_k \hat{x}_k)^T\right] \tag{2.8}$$

Now, after some algebra the following expression is obtained for the error covariance matrix:

$$P_k = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k \tag{2.9}$$

This is a general expression for updating the error covariance matrix, and it applies for any value of K. The resulting gain K is computed by the equation:

$$K_k = P_k^- H_k^- (H_k P_k^- - H_k^T + R_k)^{-1} \tag{2.10}$$

2.6 Gaps in Research

There are many attempts in the past to improve autonomous mobile robot navigation. Though a single elegant solution to the navigation problem still difficult to achieve, as various combinations of the many different methods may provide good results, both in simulation and in the real world problems. Thus a comprehensive literature review on autonomous mobile robot navigation concludes that:

The research endeavors have been for highly structured and static/dynamic environments.

- The challenges lie in real-time navigation in unstructured outdoor terrain, existence of dynamic obstacles, visual perception for visual control, visual sensors and incorporation of image information in the control loop.
- System integration of mobile robot platform should be examined using the vision sensor for the object detection, object segmentation and specified trajectory tracking problems. Mobile robots taking sharp turn or other complex maneuvers create problems considering various trajectory or path planning problems and these aspects should be handled to get restrained and smooth transition.
- The existing literature specify multiple techniques for solving problems such as motion planning, robot localization, target detection and static/dynamic obstacle avoidance in real time situations while the robustness and efficiency of many of these techniques is questionable.
- Path tracking control algorithm should be explored according to mobile robot velocity constraint such as heading and angular velocity in real world condition. There is no efficient long term path planning tracking techniques in dynamic uncertain real-time environments situation. Integration of path planning and tracking while navigating in a static and dynamic environment is rarely discussed.

2.7 Objectives of Proposed Research

Based on the identified gaps following objectives are enumerated below for further investigation.

- Development of mobile robot which can follow the specified path or the programmed path autonomously from the start location to its target location. Detection and tracking of the same is done using vision sensor. In addition to above development and use of integrated strategy for denoising the tracked path while following the specified path so that online control of mobile robot can be achieved.
- Development and implementation of path planning method for the developed mobile robot and use of vision sensor to identify and group the obstacles in the task space into handleable/non-handleable types. Determination of optimal path for mobile robot start and goal position using an algorithm.
- Development and implementation of navigation method for the designed mobile robot which has shape awareness capability in presence of static obstacles. Integration of vision sensor-based tracking and denoising method for the mobile robot while performing the task in the specified environment.
- Implementation of navigation method for a mobile robot with onboard stereo camera in presence of dynamic obstacles. Integration of vision sensor based tracking and denoising method for the mobile robot while performing the task in the specified environment.

The thesis attempts to address above objectives in various chapters. Chapter 3 discusses a vision sensor based approach to track the mobile robot while following a specified path and denoising the tracked trajectory for online surveillance. Chapter 4 develops and discusses a vision based path planning approach for object clean up in a task space. Similarly, Chapter 5 develops and presents an approach for vision sensor based tracking of mobile robot while navigating in a structured environment in presence of static obstacles and denoising the tracked path. Lastly Chapter 6 discusses an approach for vision based navigation and tracking of a mobile robot in an environment where the obstacles are static/dynamic and denoising of the tracked path.