Review

Autonomous Vehicles Navigation with Visual Target Tracking: Technical Approaches

Zhen Jia [1,*]**, Arjuna Balasuriya** [2] **and Subhash Challa** [3]

[1] United Technologies Research Center, No. 3239, Shenjiang Road, Shanghai 201206, P.R. China

[2] Department of Mechanical Engineering, Massachusetts Institute of Technology, MA, USA

[3] Faculty of Engineering, University of Technology, Sydney, NSW, Australia

E-mails: jiazhen@gmail.com, arjunab@mit.edu, subhash.challa@uts.edu.au

[*] Author to whom correspondence should be addressed.

Abstract: This paper surveys the developments of last 10 years in the area of vision based target tracking for autonomous vehicles navigation. First, the motivations and applications of using vision based target tracking for autonomous vehicles navigation are presented in the introduction section. It can be concluded that it is very necessary to develop robust visual target tracking based navigation algorithms for the broad applications of autonomous vehicles. Then this paper reviews the recent techniques in three different categories: vision based target tracking for the applications of land, underwater and aerial vehicles navigation. Next, the increasing trends of using data fusion for visual target tracking based autonomous vehicles navigation are discussed. Through data fusion the tracking performance is improved and becomes more robust. Based on the review, the remaining research challenges are summarized and future research directions are investigated.

Keywords: Computer Vision, Autonomous Vehicles, Mobile Robots, Target Tracking, Navigation, Sensor Data Fusion

1. Introduction

At present, the usage of autonomous vehicles is growing especially in applications such as manufacturing, hazardous materials handling, surveillance, etc. The basic task in any such application is the

perception of the environment through one or more sensors. Processing of the sensor input results in a particular representation of the unknown environment, which can then be used for navigating and controlling the vehicle. Autonomous vehicle navigation in a certain environment is thus a quest that many researchers have tackled over the years.

The general sensors used for autonomous vehicles include infra-red, sonar, laser, radar and so on [1]. For example, Patent [2] discusses a navigation and control system including an emitter sensor configured to locate objects in a predetermined field of view from a vehicle; Fujimura *et al.* [3] describe the techniques which make use of characteristics of infrared sensitive video data, in which heat emitting objects appear as hot spots. Compared to these types of sensors, vision sensors provide a whole new way for autonomous vehicles to create an image of the environment [4–7]. Video images plus specialized computer vision algorithms can provide high resolution information concerning the shape or range of nearby objects and environment. Coupled with the availability of increased computational power, visual sensor information becomes not only appealing but also easily attainable in real-time.

During the past ten years much research has gone into the area of computer vision for autonomous vehicles navigation [4]. Many algorithms and methods have been proposed, all with an ultimate common goal: to give intelligence to autonomous vehicles to interpret the visual information. If the research goal is to send an autonomous vehicle from one coordinate location to another, there is sufficient accumulated expertise in the research community today to design algorithms which could do that in a typical environment. But if the goal is to carry out the function-driven navigation, such as chasing or following moving targets, avoiding the obstacle which is somewhere in a given hallway and stopping at a stop sign (e.g. docking) under varying illumination and background conditions, it is still eons away. It is still the central research problem for vision based autonomous vehicle navigation that an autonomous vehicles must be aware of the position and dynamic information of the certain moving objects encountered in the environment.

Therefore this paper reviews the recent techniques in vision based target tracking for autonomous vehicles navigation. There are, of course, many approaches, and the publications list will be too long if including all of them. When people read such a broad survey paper, they will miss the key points and milestones. Thus this paper surveys only those contributions in the last decade that the authors believe are interesting and important.

One thing to be mentioned is that normally CCD (Charge-Coupled Device) cameras are used as vision sensors for autonomous vehicles navigation. CCD cameras' installation and maintenance costs are quite minimal and stereo CCD cameras can also provide the three-dimensional ($3D$) scene analysis [5]. For example, a system and method is presented in [8] for efficiently locating in $3D$ an object of interest in a target scene using video information captured by a plurality of CCD cameras. This system and method provide multi-camera visual odometry wherein pose estimates are generated for each camera by all of the cameras in the multi-camera configuration. The position and velocity of the target relative to the vehicle can be established continually by processing the stream of the cameras images, and this information can be used to navigate the vehicle. Generally the camera/object states in a tracking system can be divided into 4 categories [9]: 1) Stationary Camera, Stationary Object (SCSO), 2) Stationary Camera, Moving Object (SCMO), 3) Moving Camera, Stationary Object (MCSO), 4) Moving Camera, Moving Object (MCMO). In the case of visual target tracking by autonomous vehicles, both the camera
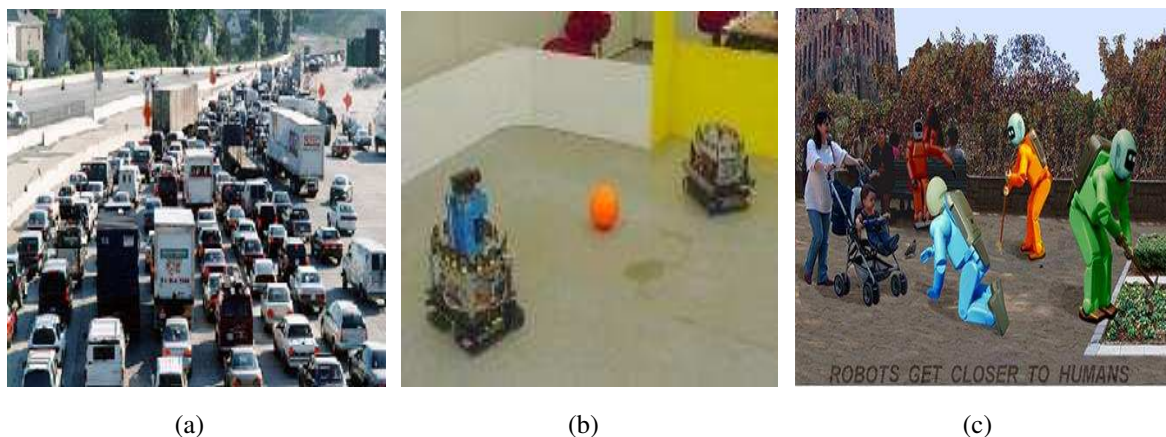
and object move with respect to each other and it is the MCMO state.

Before going into the algorithms details, the applications of visual target tracking for autonomous vehicles navigation are summarized in the next section.

1.1. Applications of Visual Target Tracking for Autonomous Vehicles Navigation

The moving target's position and velocity information can aid the autonomous vehicle to determine what constitutes its surroundings and what actions if necessary are to be taken. The potential applications of such visual target tracking systems are: autonomous vehicle navigation, map building, robot localization, path planning, obstacle avoidance, surveillance systems, intelligent transportation systems and human assistance mobile robots and so on [4].

Figure 1. Some applications of the visual target tracking system for autonomous vehicles navigation. (a) shows the general high way system (the image taken from http://www.tfhrc.gov/humanfac/presentation/). The visual target tracking based intelligent transportation system can be developed for road safety, vehicle avoidances and so on. (b) shows the mobile robots soccer (the image taken from the web site of Automation Laboratory at University of Mannheim). (c) shows the human assistance mobile robots (the image taken from the proceedings cover of IEEE ICRA 2005). The visual target tracking can help autonomous vehicles (mobile robots) to fulfill the tasks in all these applications.

(a)   (b)   (c)

In the intelligent transportation systems (Figure 1), the extraction and tracking of objects of interest are the necessary prerequisites for the development of intelligent autonomous vehicles or mobile robots. In the case of forward collision avoidance, visual moving objects tracking helps to distinguish the potential collision threats in terms of their relevance to the intended path of the vehicle [10]. Apart from such applications visual target tracking can also provide assistance to human drivers. One such application can be in drowsy driver warning systems. The knowledge of moving objects around the vehicle enables a driver assistant system to alert a driver of potential collisions and dangers [10]. Other applications in the intelligent transportation systems can be the design of an optimal trajectory for one vehicle under normal conditions, in order to overtake a single, slower-moving vehicle on a predetermined road [11].

Path planing for autonomous vehicles is to plan in real time a collision-free path in the presence of dynamically moving objects and with a limited sensing range [12]. Path planning is a fundamentally important issue in robotics, which also requires the autonomous vehicle to be aware of the objects in the environment. Complete coverage path planning (CCPP, which is also called region filling or area covering) of cleaning robots is a special type application of path planning in a two-dimensional ($2D$) environment. Here visual target tracking can help the autonomous vehicle pass through every area in the workspace and avoid obstacles [13]. Patent [14] explains an arrangement for obstacle detection in autonomous vehicles wherein two significant data manipulations are employed in order to provide a more accurate read of potential obstacles and thus contribute to more efficient and effective operation of an autonomous vehicle; in Patent [15] a mobile robot is equipped with a range finder and a stereo vision system. The mobile robot is capable of autonomously navigating through urban terrain, generating a map based on data from the range finder and transmitting the map to the operator, as part of several reconnaissance operations selectable by the operator.
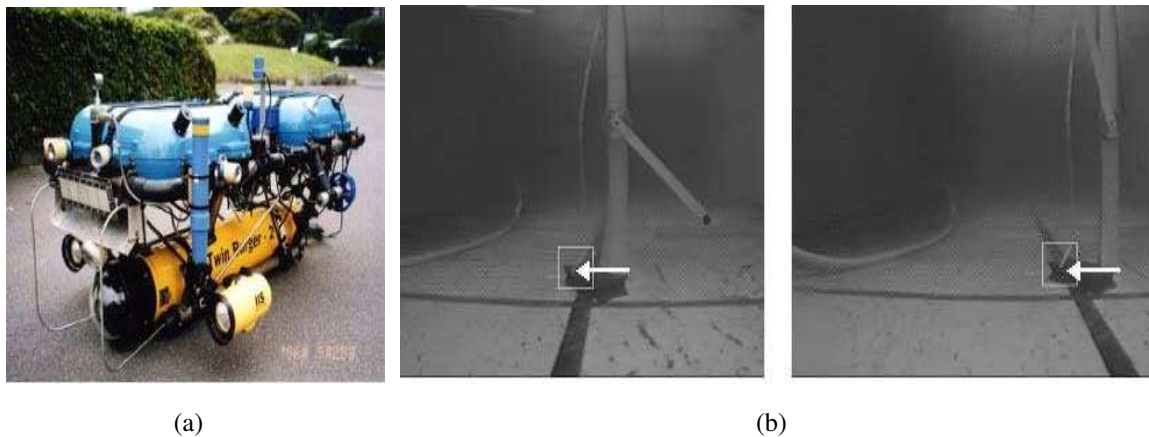
There are other applications where visual target tracking systems are used such as in the human assistance mobile robotics (Figure 1). Mobile robots equipped with mechanisms for communication, interaction, and behaviors are employed more and more outside traditional manufacturing applications, such as in museums or exposition areas [16, 17]. In order to operate mobile robots in the same environment as a human, obstacle avoidance techniques in a dynamic environment are required. Such collision avoidance problems between human and robots can be called as interaction based on the position information [18], which can be solved by visual target tracking.

In security or surveillance applications, it is often more important to carefully position multiple sensors in order to cover a large structured environment. In this content, autonomous vehicle-based trackers are attractive; they can potentially reduce the number of sensors needed in the tracking network, and they should be able to adapt to the movements of the targets or the dynamic changes in an environment by re-positioning themselves in responses [19].

In the underwater environment visual target tracking can be used for Autonomous Underwater Vehicles (AUV) navigation. For example in order to study the behavioral patterns of unknown underwater life forms, it is necessary to observe them carefully for a longer period. It is therefore important for AUVs to stay close to the object being observed by moving with it such as fish following. Also, for a reliable functioning of underwater man-made systems, it is necessary to maintain them by routine observations. Therefore, the underwater exploration and maintenance requires the AUV to have proper observations and as a result close tracking and following along or with these objects is necessary [21] (Figure 2). Arjuna and Tamaki [20] propose a sensor fusion technique for Autonomous Underwater Vehicles (the test bed vehicle Twin-Burger 2 in University of Tokyo (Figure 2) to track and follow underwater cables through video images. Also Arjuna *et al.* [22] propose the vision based tracking system for underwater docking. There are also many other robotics applications requiring the visual obstacle avoidance and/or object identification in unstructured underwater environments. For example the U.S. Navy is investigating the use of autonomous underwater vehicles to accurately detect and classify underwater mines prior to beach landings in the hostile territory [23].

Some other applications of visual target tracking systems for autonomous vehicles navigation can also be in the following areas:

Figure 2. Some applications of the visual target tracking system for AUV navigation. (a) shows the test bed AUV (Twin-Burger 2) for the underwater cabal following (the image taken from [20]). (b) shows the underwater images of AUV docking (the image taken from [21]).

(a)   (b)

- The mobile robots soccer [24, 25] (Figure 1).

- Obstacle avoidance for the robot manipulators in factory automation doing repetitive and dull work [26].

- Obstacle avoidance for the robotics motion planning [27].

- The motion control of autonomous vehicles in the car manufacturing industry. Many car manufacturers plan to equip their vehicles in the near future with computer-aided visual target tracking capabilities for parallel parking or automatic stop-and-go mode in traffic jams [28].

In order to successfully achieve all the above applications, it is very useful and necessary for autonomous vehicles to have the knowledge of the $3D$ dynamic information of the objects of interest in the environment. Therefore, it is very beneficial to write such a survey paper to highlight and summarize the recent interesting techniques on this topic. This rest of paper is organized as follows: In Section 2, the algorithms for autonomous land, underwater, aerial vehicles are reviewed separately. Next data fusion based methods for autonomous vehicles navigation are reviewed in Section 3. In Section 4, based on the previous reviews, remaining research problems are concluded and future research directions are identified. Section 5 draws the conclusion of this paper.

In our paper various object tracking methods will not be fully covered, because it is really out of scope of our main topic for autonomous vehicles navigation. However Yilmaz *et al.* [29] survey most of state-of-art object tracking methods. In their paper, they also mention that object tracking can be widely used for vehicle navigation, such as video-based path planning and obstacle avoidance capabilities. They review: object representation methods; feature selection methods for tracking; object detection methods and object tracking methods such as Point Tracking, Kernel Tracking, Silhouette Tracking. This paper can be good supplements for our survey paper.

2. Vision based Target Tracking for Autonomous Vehicles Navigation

The targets of interest can normally be the moving cars, moving persons or any other moving objects which autonomous vehicles need to track for navigation. Here the visual target tracking methods are reviewed in three different categories based on the applications in the land, underwater and aerial environments. Generally autonomous land vehicles are more widely developed and used and more navigation algorithms are developed for them. So in this section the literature survey focuses more on the visual target tracking algorithms for autonomous land vehicles navigation.

2.1. *Visual Target Tracking for Autonomous Land Vehicles Navigation*

In general applications of visual target tracking for Autonomous Land Vehicles navigation include some sorts of landmarks tracking, people following, multiple mobile robots cooperation, vehicles localization and map building, visual target tracking with pan-tilt cameras platforms and so on. According to their applications, different methods are categorized and reviewed as follows. It should be pointed out that the classification in this part is not absolute because algorithms from different categories can be intergraded together to achieve the navigation goal.

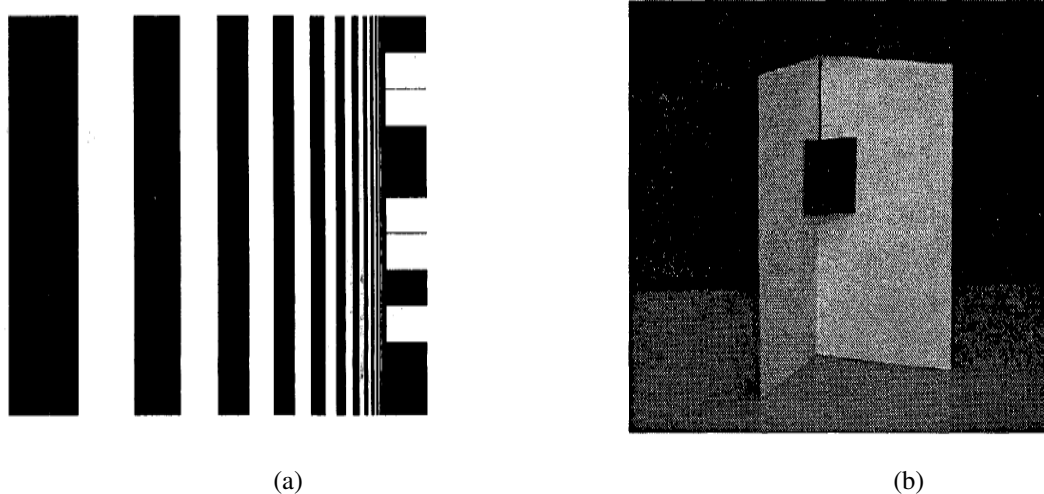Visual Landmark Tracking for Autonomous Land Vehicles Navigation

Visual landmark tracking is one kind of vision-based methods for autonomous land vehicles navigation. Landmarks are divided into two classes: natural or artificial. Generally natural landmarks are selected in the scenes in consideration of their particular characteristics. Autonomous vehicles learn those characteristics or keep the features of the landmarks in memory and recognize them using neural network or some matching techniques while they move [30–32]. Mosaic images of outdoor environments are also used for the image matching based method in [33].

On the other hand an artificial landmark is often designed with a specific pattern or color in consideration of its detection algorithm. For example, a landmark that has a bar code or a specific shape pattern such as the sine waves has been proposed. Recently, Briggs *et al.* [34] use the self-similar gray pattern landmarks for navigation and localization aids (Figure 3). In [35] Jang *et al.* propose a simple artificial landmark model, which can be used for the self localization of indoor mobile robots (Figure 3). In their paper an effective visual detection and tracking algorithm for this landmark is proposed. A pair of color points neighboring each other have been used as a sample to represent the probability density in the Condensation algorithm. Under the assumption of affine cameras and only with the information of a single landmark in a single image, they have presented a localization algorithm to estimate the absolute position accurately. In [36] Wei *et al.* propose a visual landmark tracking algorithm for docking unicycle-like vehicles based on the hearing-only information. An omni-directional panoramic camera is used to detect visual landmarks around the docking station and provide bearing (or heading) data for each observed landmark. In their method a robust and computationally cheap visual blob detection algorithm is proposed. The artificial landmarks used consist of two adjacent large color blobs. Using two adjacent color blobs improves the noise rejection over the single color blob extraction algorithms. In [37] Breed describes method and system for enabling semi-autonomous or autonomous vehicle travel includes providing a vehicle travel management system which monitors the location of vehicles in a

travel lane and the location of the travel lane, creating dedicated travel lanes for vehicles equipped with the vehicle travel management system, and managing travel of vehicles in the dedicated travel lanes to maximize travel speed of vehicles and minimize collisions between vehicles.

Normally, the robust extraction of natural landmarks is a difficult task. And the artificial landmark methods that use peculiar contour, color or edge information highly depend on the low-level image processing results, which are influenced by the noise and de-focus phenomenon very much. Also, artificial landmark methods are not robust under the geometrical background variations such as the rotation, camera zoom and viewing direction changes.

Figure 3. (a) is taken from [34], which shows their self-similar landmark pattern with barcode for landmarks detection and tracking. (b) is taken from [35]. This figure shows the structure of landmark, which they use for the self localization of indoor mobile robots.

(a)   (b)

Human Following for Autonomous Land Vehicles Navigation

Several approaches are proposed to use visual target tracking for human following by autonomous vehicles.

In [38] Hirai *et al.* present a visual tracking system for a human collaborative mobile robot. The robot tracks the human back and shoulder to follow a person. Normally it is not easy to keep tracking the human back if the background is so complex and cluttered. They solve the problem by choosing the texture of clothes and the human shoulder images as the template patterns to be detected and identified. This tracking system requires to know the special visual features of the target (the texture of the clothes and human shoulder). Such tracking systems are not suitable to track unknown objects appearing in an unknown situation.

One kind of research work on autonomous vehicles human following can be found from Morioka *et al.* [39–41]. According to Morioka *et al.* [40]'s human-following work, an intelligent space (ISpace), which is an intelligent environment with many intelligent sensors, is provided. The autonomous vehicle

cooperates with multiple intelligent sensors, which are distributed in the ISpace. The distributed sensors recognize the target human and the autonomous vehicle, and give control commands to the robot. CCD cameras are used as a kind of sensors of DINDs (Distributed Intelligent Network Devices) for ISpace. Location information of the human and autonomous vehicle is obtained by stereo vision processing and human do not need to hold any special tags (Figure 5). However there is a major drawback with this kind of human-following methods. The system strictly needs the ISpace. Normally for an unknown or unstructured environment, such an ISpace is not available, thus in order to achieve the visual target tracking, the vehicle's onboard sensors should mostly provide the accurate target dynamic information.

In [17] Jensen *et al.* design an autonomous vehicle, which at the same time executes a pre-programmed tour in a public exposition and allows for complex, collaborative interactions with the non-experienced visitors. In their system, they have dedicated two tasks to gather the information of the human presence in a public environment: a color camera based face tracking and a motion tracking based on the information from the laser ranger finder. The main steps of visual face detection and tracking in their system include: the skin color detection; contour extraction and filtering; tracking. Information gathered from the face tracking together with the motion tracking helps to verify the presence of the visitors.

In [42] Nishiwaki *et al.* present a humanoid walking control system that generates body trajectories to follow a given desired motion on-line. They implement the system by making the autonomous vehicle track and follow a moving person based on the stereo vision feedback. The visual tracking consists of 3 parts: a stereo vision processing for target detection and $3D$ position estimation in the camera coordinates; a planning of the desired future torso movements during one step; a camera posture and gaze direction control with the self-motion compensation. While the autonomous vehicle and human are both moving, the color segmentation and thresholds are utilized to detect the relative human's direction. Then a real-time depth map generation algorithm is employed to measure the distance to the human.

Kwon *et al.* [43] present an efficient human following algorithm for an autonomous vehicle using two independent moving cameras. In order to control the camera's pan/tilt motions, they have presented an image-based PTU control algorithm using a lookup table that stores the correspondences between the camera pan/tilt angles required to keep a target in the center of the image frame and the pixel displacements produced by the target in the image plain (Figure 4). The major problem with this method is that the object tracking is accomplished with a simple color histogram based algorithm. Using color information of the person's specific appearance, it calculates the centers of masses of the segmented color-blobs in each of the two images that form the conjugate pair of images in a tracking sequence. The current viewing direction of each camera in their system is adjusted so that the center of mass becomes the center of the image frame. However, a change in illumination can induce shifts in the center of mass of the blob being tracked in the two camera images, which makes the target tracking fail.
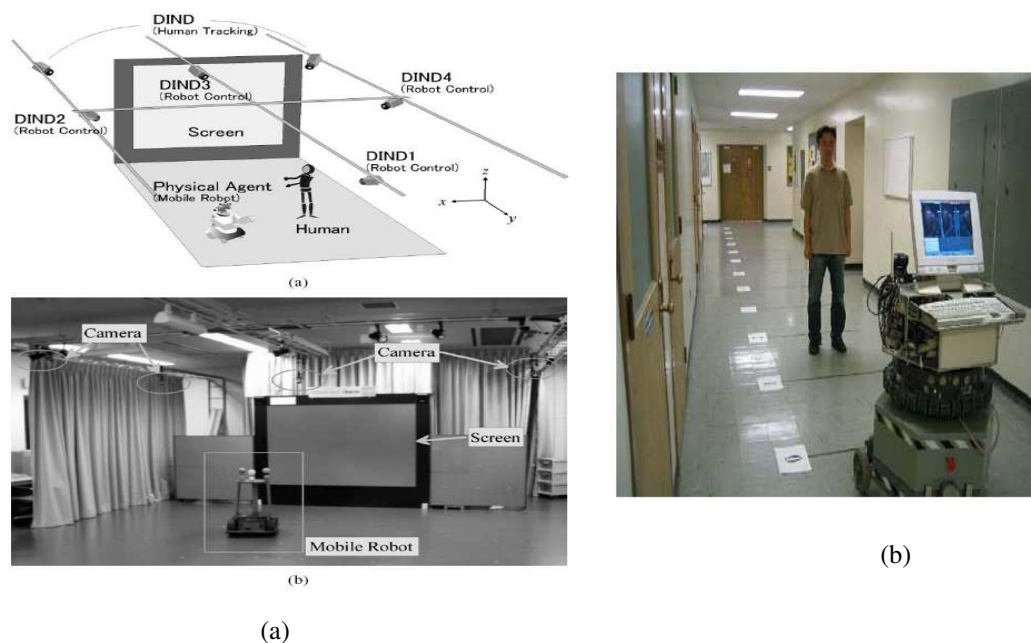
Ogata *et al.* [44] propose a tracking system employing a visually controlled aerial robot which recognizes the motion of the specified person. They propose the motion recognition technique employing MHIs and eigenspaces. The human region is extracted by its color information.

One example of optical flow based autonomous vehicles human following systems can be referred to Doi *et al.* [45]. They propose a real-time navigation system which observes the human behavior and reacts to those actions. The system detects body parts as the moving areas, and a face region or region specific to human is then extracted in the detected area based on the skin color or the cloth color of the

human.

In conclusion all these human following methods detect and track human's motion based on human's special visual features or the additional sensors in the environment. It limits the applications of such algorithms to track unexpected or different objects in an unstructured outdoor environment.

Figure 4. (a) is taken from [40]. This figure shows Morioka *et al.*'s human following algorithm's system setup and their experimental testing environment. Such an ISpace is specially designed for their algorithm. (b) is taken from [43], which shows their experimental environment for person following. Their system is designed specially based on the person clothes' color.

Visual Target Tracking for Autonomous Land Vehicles Localization and Map Building

Localization and target-tracking are both challenging yet essential and have a wide range of applications in mobile robotics. Localization can be defined as determining the position of an object within a reference coordinate system, and tracking consists of constructing a trajectory given a collection of spatially and temporally coherent localizations. Localization, mapping and moving object tracking serve as the basis for scene understanding, which is a key prerequisite for making a robot truly autonomous. Simultaneous localization, mapping and moving object tracking (SLAMMOT) involves not only simultaneous localization and mapping (SLAM) in dynamic environments but also detecting and tracking these dynamic objects [46]. Several methods have been proposed using visual target tracking for autonomous vehicles localization and map building.

In [47] Hajjawi *et al.* present an algorithm for visual position tracking of individual cooperative autonomous vehicles within their working environment. Initially, they present a technique suitable for

visual servoing of an autonomous vehicle towards its landmark targets. Secondly, they present an image processing technique that utilizes images from a remote surveillance camera for localization of the robots within the operational environment. In their algorithm the surveillance cameras can be either stationary or mobile. The supervisory control system keeps tracking of relative locations of individual autonomous vehicles and utilizes relative coordinates information of autonomous vehicles to coordinate their cooperative activities.

Burschka *et al.* [48] present a real-time mobile navigation system and an approach for the vision-based Simultaneous Localization and Mapping (SLAM) based on a second generation of the image processing library, XVision. They show how the multiple tracking and low-level image processing primitives for color, texture and disparity can be combined to produce vision-guided navigation systems. The applications they discuss make use of XVision capabilities to solve the temporal correspondence problem by tracking an image feature in a given image domain.
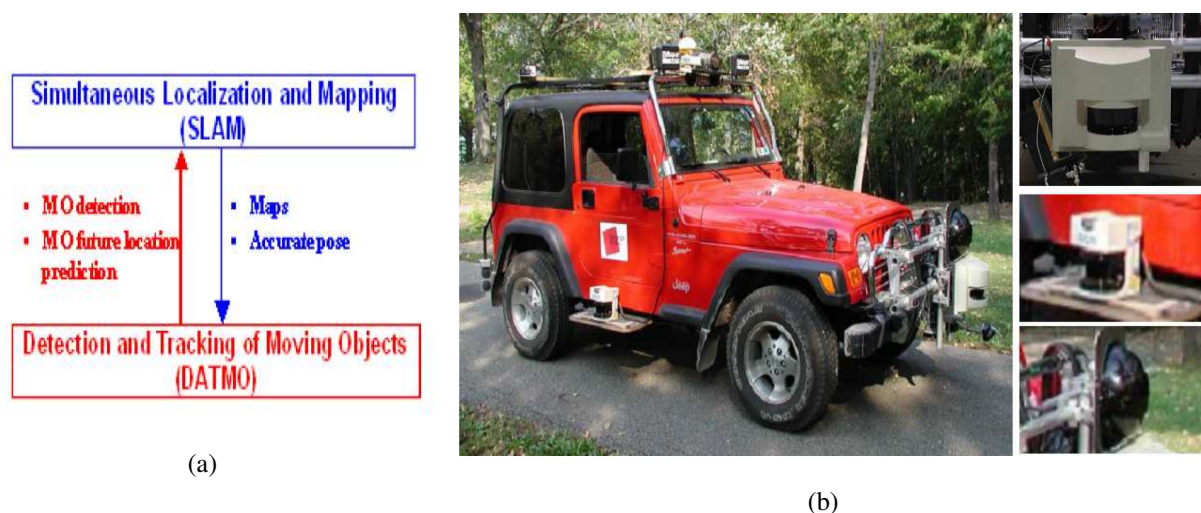
In [49] Dao *et al.* present a simple linear method for localizing an indoor mobile robot based on a natural landmark model and a robust tracking algorithm. The landmark model consists of a set of three or more natural lines such as baselines, door edges and linear edges in tables or chairs to take the advantages of fast landmark detection. The canny operator and Lucas-Kanade algorithm are used for the effectively detecting and tracking of the landmark model. Next, a quick localization method for autonomous vehicles from correspondent lines is proposed by adopting a linear technique. Based on assumptions on indoor environments, a complex nonlinear problem for the $3D$ pose determination using lines is converted to an iterative linear problem, which makes it possible to apply the proposed algorithm for real-time applications. However, in their system the line detection methods are not quite accurate and robust and with some additional condition constraints.

In the most recent work of [46], Wang establishes a mathematical framework to integrate SLAM and moving object tracking (Figure 5). He describes two solutions: SLAM with generic objects (GO), and SLAM with detection and tracking of moving objects (DATMO). SLAM with GO calculates a joint posterior over all generic objects and the robot. Such an approach is similar to existing SLAM algorithms, but with additional structures to allow for motion modeling of the generic objects. Unfortunately, it is computationally demanding and infeasible. Consequently, he provides the second solution, SLAM with DATMO, in which the estimation problem is decomposed into two separate estimators. By maintaining separate posteriors for the stationary objects and the moving objects, the resulting estimation problems are much lower dimensional than SLAM with GO.

Normally autonomous vehicle is assumed to move in the environment with the prior knowledge of its location. The SLAM and autonomous vehicles target tracking are considered separately [10]. However it is believed by many that a solution to the SLAMMOT problem would expand autonomous vehicles applications in proximity to human beings where autonomous vehicles work not only for people but also with people.

From the reviews in this part, the SLAM can be successfully integrated with visual target tracking methods to make autonomous vehicles work at high speeds under situations like the large crowded city urban environment [50]. As a future work, in order to make the vehicle fully autonomously navigate the environment, the SLAM problem can be considered more and more at the same time with the visual target tracking problem.

Figure 5. These two figures are taken from [46]. (a) shows the relationship between SLAM and DATMO. The simultaneous localization, mapping and moving object tracking problem aims to tackle the SLAM problem and the DATMO problem at once. Because SLAM provides more accurate pose estimates and a surrounding map, a wide variety of moving objects are detected using the surrounding map without using any predefined features or appearances, and tracking is performed reliably with accurate autonomous vehicles pose estimates. SLAM can be more accurate because moving objects are filtered out of the SLAM process thanks to the moving object location prediction from DATMO. SLAM and DATMO are mutually beneficial. The left of (b) shows the Navlab11 testbed, which is used to test their Simultaneous Localization, Mapping and Moving Object Tracking algorithm. The right of (b) shows the sensors used for the testbed (SICK LMS221, SICK LMS291 and the tri-camera system).

(a)

(b)

Visual Target Tracking with Pan-Tilt Camera Platforms in Autonomous Land Vehicles

Ego-motion estimation or pan-tilt cameras' motion control are one of the key issues in autonomous vehicles navigation, especially in demand for moving objects tracking.

In [51] Karlsson *et al.* develop a lightweight, robust real-time tracking system used on an experimental geo-referenced cameras platform. The purpose of their system is to study the benefits of combining image processing with navigation data that should be available from the control system of any AGV system. Their experiments show that by using a Kalman Filter the tracking algorithm can handle objects' large movements between images and it becomes more resistant to the occlusions. However, their tracking system is designed mainly based on the image brightness, which gives much less robust performance under environmental lighting changes.

The objective of the research work in [52] is to derive the orientation of a pan-tilt camera fitting a drone in order to track a target and to maintain its position in the middle of the image (Figure 6). To ensure real-time video operation, an algorithmic solution integrating a successive-step and multi-block

search method is implemented, thus allowing tracking with complex target displacements (Figure 6). The micro-controller uses this information to manage the camera orientation. With a certain regularity in the evolution of the target model, this system is sufficiently robust to track deformable targets in real images. However their technique has limitations when the target is close to the cameras. Also only a simple linear interpolation method is carried out. The drone localization and attitude are not considered in the algorithm. In addition, the implementation of a $2D$ visual camera control requires a priori knowledge of the $3D$ target model.

Tomono *et al.* [53] present a method of planning a path on which the autonomous vehicle with a pan-tile camera can find the target objects under spatial uncertainties. The object recognition is normally based on feature matching between models and the image data. However there are several problems in this method. For example, in the case that the salient features to identify the target object may center on their particular faces, the autonomous vehicle has to move around the object to find the features. If the possible locations of the target object are in a wide area, the robot has to move around the area to search the object. This paper addresses these problems by a probabilistic approach. Given an initial roughly-planned path, the proposed method optimizes it with respect to the travel time, high pass-ability, and a high probability of finding the target. The method defines a path evaluation function based on these factors and finds a suboptimal path by solving the nonlinear optimization problem of the path evaluation function. Their method combines the visibility constraint and the conventional constraints of travel time and collision avoidance for autonomous vehicles navigation.

Zhang *et al.* [54] develop a pan-tilt visual tracking system to dynamically track moving targets using vision-based control. The algorithm includes the color-based segmentation, data pre-processing and active parameters adjustment. Since computationally expensive techniques are inapplicable, they focus on the use of specific color properties to identify the objects of interest.

One of the major limitations in the pan-tilt cameras visual tracking system is that the cameras's movements are rather complex with pan and tilt motions. In the above algorithms, the cameras' complex motions are neglected or not considered enough in an accurate way. If the cameras' self motion can not be precisely identified and integrated into the tracking system, then the visual target tracking performance will not reach a satisfactory level.
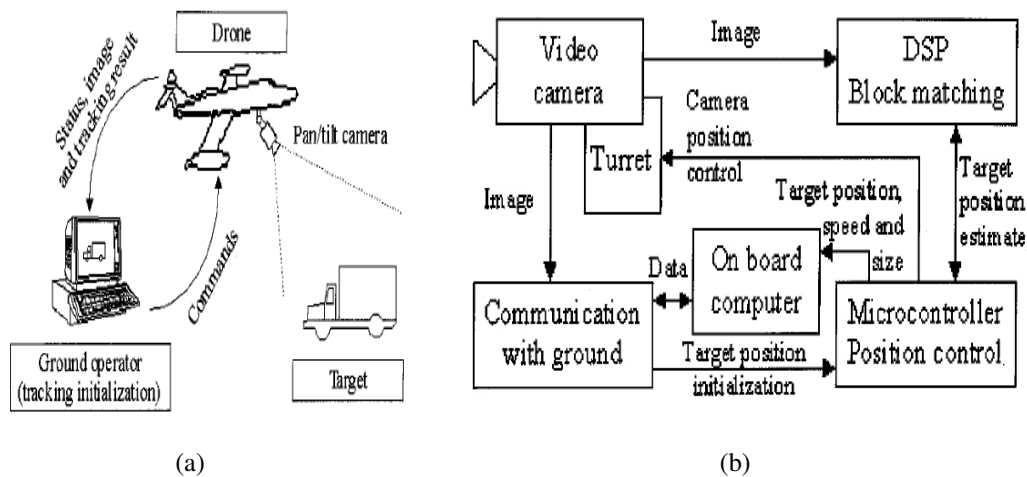
Visual Target Tracking for Multiple Mobile Robots Cooperation

Multiple mobile robots (autonomous vehicles) cooperation means that each mobile robot plans its path based on other robots' navigation information. The robots cooperate with each other to complete navigation tasks. Several methods based on the visual target tracking have been proposed in this area.

In [55] a real-time visual tracking algorithm for MRFS (Multiple Robot Fishes cooperation System) is described. They give a description of the operation process for the vision subsystem, and propose an adaptive segmentation method based on the color information. Color information is the foundation of their object identification. In MRFS, halls, obstacles and robot fishes are equipped with specified color properties.

In [56] the problem of estimating and tracking the motion of a moving target by a team of mobile robots is studied. Each robot is assumed to have a directional sensor with limited ranges, thus more than one robot (sensor) is needed for solving the problem. A sensor fusion scheme based on the inter-robot

Figure 6. These two figures are taken from [52]. (a) shows the overview of their proposed
system. (b) is the block diagram of their tracking system.

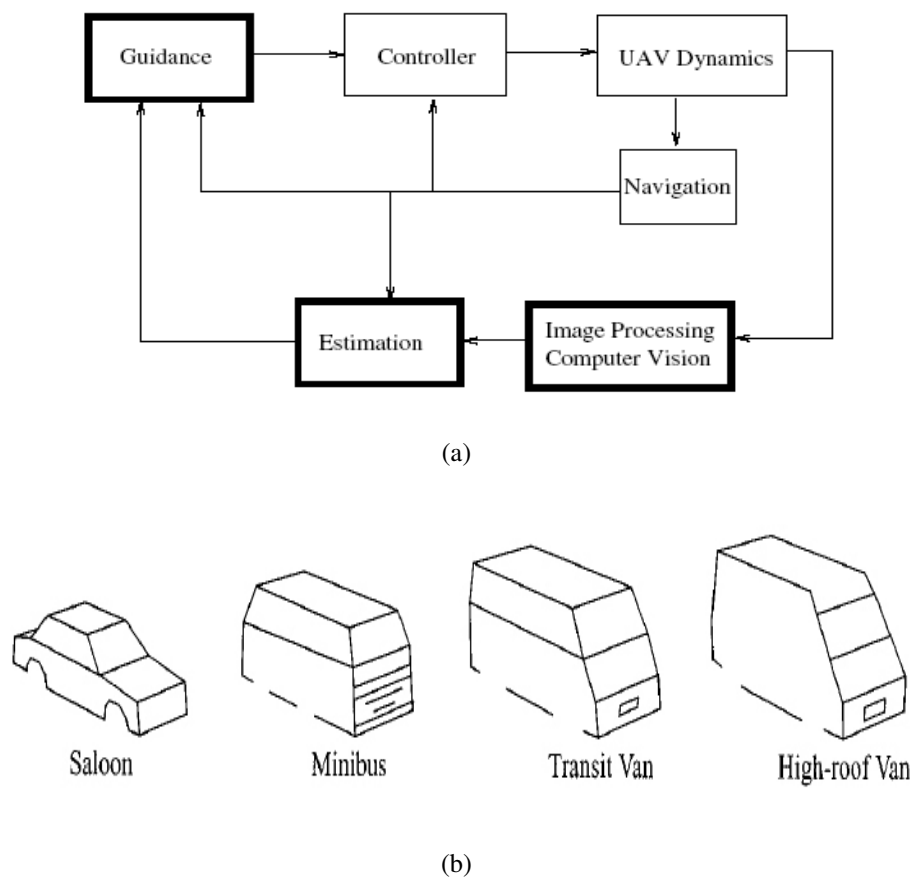(a)   (b)

communication is proposed in order to obtain the accurate real-time information of the target's position
and motion. Accordingly a hierarchical control scheme is applied, in which a consecutive set of desired
formations are planned through a discrete model and low-level continuous controls are executed to track
the resulting references.

Betser *et al.* [57] describe a tracking system relying on active contours for the target exaction and
an Extended Kalman Filter for the relative pose estimation (Figure 7). Their work represents the first
step towards treating the general problem for the control of several unmanned autonomous vehicles
flying in formulation using visual information. The Extended Kalman Filter is improved by introducing
additional image information available to the vehicle with a fixed forward-pointing monocular camera.
Active contours are used to track the follower in the image plain and provide the Kalman Filter with the
required input. There are several drawbacks in this kind of methods. They ignore the additional angle
measurements found in the equivalent $3D$ range estimations and also they do not consider the target's
acceleration effects.

In [58] Ng *et al.* have formulated an algorithm that has coordinated the movements of multiple robots
to follow a search tactic collectively in an unknown and cluttered environment. First they develop the
individual robot reactive behavior that make their coordinated movement possible. Their algorithm re-
quires every robot to be programmed with the same set of primitive behaviors: (1) obstacles negotiation;
(2) homing; (3) flocking and (4) searching, with obstacles negotiation being the most important and
searching being the least important. According to different environment stimulants, the robots adopt one
of these behaviors at a time according to their order of importance for the cooperation purposes.

Multiple mobile robots (autonomous vehicles) cooperation is a rather hot and promising research
topic. Many researchers already consider it as their future working directions, such as the multiple
mobile robots visual target tracking or multiple mobile robots SLAM. It can be foreseen that in the near
future more and more systems will be developed in this area.

Figure 7. (a) shows the system overview of [57]. (b) is taken from [59] and shows the car models used in their method for vehicles tracking. In [59], the object in the image is assumed to be an instance of one of these four models.

(a)

(b)

Other Visual Target Tracking Methods for Autonomous Land Vehicles Navigation

Traditional visual target tracking techniques usually ignore the presence of obstacles and focus on imaging and target recognition issues. The papers [60, 61] introduce a new visual tracking algorithm for autonomous vehicles navigation when the target moves unpredictably and no prior map of the environment exists. Their algorithm computes a motion strategy based exclusively on the current sensor information, while no global map or historical sensor data is required. The algorithm is based on the notion of escape risk and the computation of an escape-path tree. Their proposed algorithm governs the motion of autonomous vehicles based on current measurements of the target's position and the location of the local obstacles. Their approach is combinatorial in the sense that the algorithm explicitly computes a description of the geometric arrangement between the target and observer's visibility region produced by the local obstacles. The algorithm computes a continuous control law based on this description. However most failures of this system are due to the shortcomings of their simple visual target detection algorithms. Their method is also limited to a $2D$ workspace without considering the problems

in the $3D$ space.

Tracking targets robustly by vision is very difficult for autonomous vehicles running on irregular terrains in natural environment, because the image deformation caused by rolling and pitching of the camera, as well as the relative movement between the target and camera, affects the tracking ability greatly. One approach to cope with such problems is matching the target image with many affine transformed candidate images while tracking. But when the number of candidate images gets larger, such an approach is not available to the real-time tasks due to the computational cost. In [62] Ding *et al.* propose a new Robustness Analysis for Tracking (RAT) to improve the tracking ability. RAT is the analysis based on features of the object image, where three parameters: 'Detectability', 'Robustness for Depth (RBD)' and 'Robustness for Rotation (RBR)' are defined. Much more robust templates can be found by analyzing the object image using RAT before the tracking task is performed.

In [63] Burschka *et al.* present an approach for scene classifications in dense disparity maps from a binocular stereo map. The classification result is used for the indoor autonomous vehicle tracking and navigation purposes. The $3D$ model of the scene is also derived directly from the disparity image. The classification of the scene helps to decide, which objects are interesting and should be monitored as well as which behavior is appropriate depending on the current structure. The vehicle can switch from the wall following in hallway environments to localization based on the corner structures, etc. However, the shortcoming of this system is that the applied algorithms used for dynamical composition of tracking primitives are highly dependent on the current environmental structures.

In [64] Mori *et al.* propose a ball tracking and catching strategy called GAG (short for "Gaining Angel of Gaze") that enables an autonomous vehicle to track and catch a ball flying in the three-dimensional space. The mobile robot receives a visual feedback control scheme based on GAG and then the proposed scheme enables the robot to track and catch a ball flying in the three-dimensional space by using a monocular vision system.

In [65] Yu *et al.* propose a correspondence based method, which applies the Iterative Closest Point (ICP) algorithm to match feature points on the ground plain. Since the outliers in the scene contribute false measurements for estimation, they introduce a stereo vision-based method to detect free-space on the road plain. They extract the edge points in the free-space as primitives, which avoid the limit of the rigid scene hypothesis.

Other visual target tracking methods for autonomous land vehicles navigation are: the paper [66] deals with the problem of computing the motions of a robot observer in order to maintain the visibility of a moving target; Taking inspiration from the visual system of the fly, the paper [67] describes and characterizes a monolithic analog very large-scale integration sensor, which produces control signals appropriate for the guidance of an autonomous robot to visually track a small moving target; The visual vehicles following systems are also presented in [59, 68] (Figure 7).

Generally all the above methods have several common drawbacks: they normally do not take into account the complex $3D$ motions of cameras; they do not have appropriate target dynamic models for the tracking estimation; the target visual features excitations from images are based on simple low-level image processing methods or the target's special visual features, which are less accurate and robust; they can not track the target's $3D$ dynamics in the world coordinate.

A common problem in all the land navigation systems is caused by environmental shadows, changing

illumination conditions, changing colors, etc. This can impose serious limitations on the performance of a navigation system.

2.2. Visual Target Tracking for Autonomous Underwater Vehicles

In the underwater environment visual target tracking can be broadly used for Autonomous Underwater Vehicles (AUV) navigation. For example in order to study the behavioral patterns of unknown underwater life forms, it is necessary to observe them carefully for a longer period. It is therefore, important to stay close to the object being observed by moving with it such as fish following. Also, for a reliable functioning of underwater man-made systems, it is necessary to maintain them by routine observations. Therefore, the underwater exploration and maintenance require proper observations and, as a result the close navigation along or with these objects is necessary [21]. The visual target tracking system can be used for these autonomous underwater vehicles navigation applications.

Underwater pipe inspection is one example of a class of problems that bear many similarities with visual target tracking for autonomous land vehicles. Extracting the contours of a pipe is equivalent to extract and track the landmarks. In [7] Rives *et al.* take a visual servoing approach and devise a controller that take the inputs as the lines extracted from the image of a pope and uses this information to generate steering commands for the ROV Vortex vehicle. In [20] Arjuna and Tamaki propose a sensor fusion technique for the Autonomous Underwater Vehicle to track underwater cables. They propose a sensor fusion scheme using the dead reckoning position uncertainty with a $2D$ position model of the cable to predict the region of interest in the image. They solve two practical problems encountered in the optical vision based systems in underwater environments: first, the navigation of AUV when the cable is invisible in the image; second, the selection of the correct cable (interest feature) when there are many similar features appearing in the image.

Arjuna *et al.* [22] also propose the vision based tracking system for underwater docking using correlation based underwater images template matching.
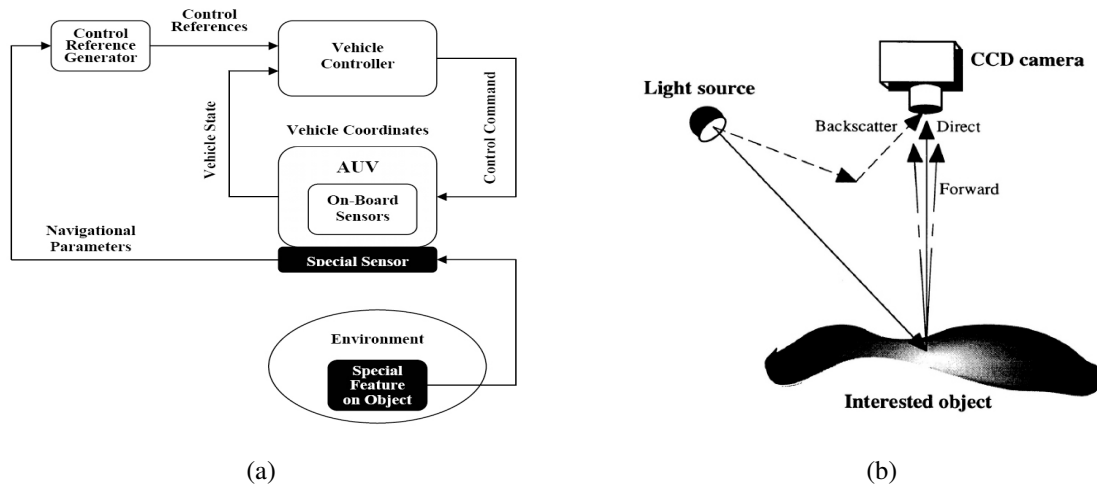
One of the most recent underwater visual target tracking methods is proposed in [21] for AUV navigation (Figure 8). In Yang's work the objects of interest are extracted from the images by using the dynamic properties (optical flow techniques) and their optical features (color, texture, shape and so on). The consecutive dynamic behavior of the objects of interest is then estimated based on the current dynamics. By using this predicted dynamics, the amount of data for the region of interest identification can be reduced. This also increases the speed of processing for the hardware available in the small and limited hulls of AUV. After the image dynamics and feature position information are fused with the AUV's other onboard sensor data, the navigation commands are derived for AUV to track the object.

2.3. Visual Target Tracking for Autonomous Aerial Vehicles

Development of Autonomous Aerial Vehicles (AAV) (such as the unmanned helicopters) has been an active area of research for several years. AAVs have been used as test beds to investigate problems ranging from control, navigation, and path planing to object tracking and following [69].

An early approach to the AAV landing [70] decouples the landing problem from vision-based target tracking. Nowadays several techniques have been implemented for vision based landing of an AAV on

Figure 8. These figures are taken from [21]. (a) shows the target-based AUV navigation scheme. Normal AUV visual target tracking systems are developed based on this scheme. (b) shows the components of an underwater image system.

(a)   (b)

stationary or moving targets [70]. The problem of landing as such is inherently difficult because of the instability of the AAV near the ground [71]. Also, since the dynamics of a helicopter is nonlinear, only an approximate model of the helicopter can be constructed [69].

In [72] a vision-based target tracking approach to the AAV landing is presented. Their landing pad has a unique shape which makes the problem of identification of the landing pad much simpler. In [73], a vision-based solution is given for the safe landing-site detection in the unstructured terrain, where the key problem is for the onboard vision system to detect a suitable place to land without the aid of a structured landmark, such as a helipad. The University of California, Berkeley (UC Berkeley) team has proposed a real-time computer vision system for tracking a landing target [71, 74] and have successfully coupled it with a helicopter controller to achieve landing [75].

In [76] Saripalli *et al.* have presented the design and implementation of a real-time vision-based system for detecting a landing target (stationary or in intermittent motions) and a controller to autonomously land the AAV on the target. Their method relies on the assumptions that the landing target has a well-defined geometric shape, and all the feature points of the landing target are coplanar. In [76] they use invariant moment descriptors for detecting and landing the AAV on the target. They do not impose any restriction on the shape of the landing pad except that it is planar. However there exists the problem of safely and precisely landing the AAV in the unstructured harsh $3D$ environment.

The visual target tracking approach from [69] differs from the prior approaches in two ways (Figure 9). First, they impose no constraints on the design of the landing pad except that it should lie on a two-dimensional plain. Secondly, they use moment descriptors to determine the location and orientation of the landing target. Their algorithm is not only able to detect and land on a given target, but also able to track a target which moves intermittently, and land on it.

In [77] Saripalli *et al.* present a vision-based algorithm designed to enable the AAV to land on a

moving target. The AAV is required to identify a target, track it, and land on it while the target is also in motion. They use Hu's moments of inertia for the precise target recognition and a Kalman Filter for target tracking. Based on the output of the tracker, a simple trajectory controller is implemented which (within the given constraints) ensures that the AAV is able to land on the target.
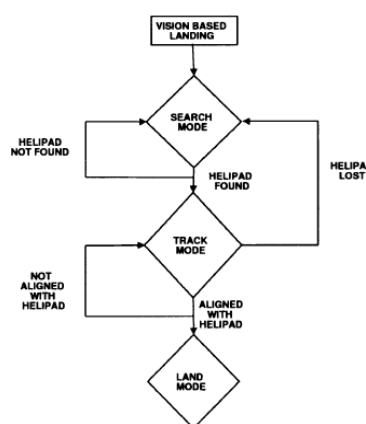
In [78] it has been shown that Autonomous Aerial Vehicles equipped with a 2-degree-of-freedom pan-tilt camera can be used for the long-term autonomous observation of stationary ground targets. The environmental aspects such as wind and sunlight have an influence on the quality of the images taken by the camera. These influences can set limits to the time span of continuous target observation, especially when the range of camera motion is mechanically limited. Two different camera limit specifications with the realistic applicability have been analyzed for these potential problems. Solutions to these problems have been found by applying circle-based flight maneuvers including commanded sideslip for heading adjustment.

In conclusion several limitations exist with the above visual target tracking algorithms for AAVs: Some of their estimators can only track the target in a single dimension; Some of them can not track the targets in all the six degrees of freedom; They all use the algorithms based on the intensity of the image for object detection; Some of them assume that the object is planar, which is quite restrictive in nature; A better object detector and a camera calibration routine can be integrated with these algorithms so that the coordinates of the tracked object obtained during the vision processing stage are much more accurate; These algorithms can only track single object and can not pursue and land on evasive targets.
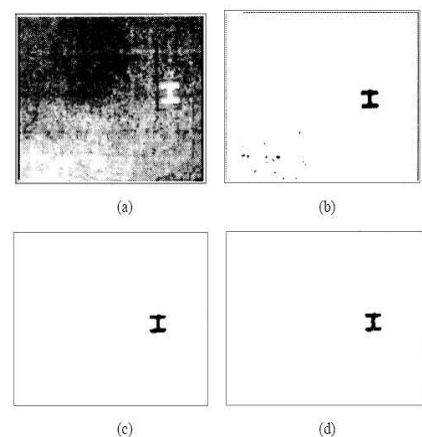
Figure 9. These figures are taken from [69]. (a) shows the USC autonomous aerial vehicle tracking and reconnaissance (AVATAR) after landing on a helipad. (b) shows the state transition diagram for AAV landing in [69]. (c) shows the image processing results (image (a) is captured in flight from the downward-pointing camera on the helicopter; (a) image from onboard camera; (b) thresholded and filtered image; (c) segmented image; (d) final image).

(a)   (b)   (c)

Table 1. The Classification of Data Fusion Algorithms from [79].

Estimation methods	Non-recursive:
	*Weighted Average
	*Least Squares
	Recursive:
	*Kalman Filter
	*Extended Kalman Filtering
Classification methods	*Parametric Templates
	*Cluster Analysis
	*Learning Vector Quantization (LVQ)
	*K-means Clustering
	*Kohonen Feature Map
	*ART, ARTMAP, Fuzzy-ART Network
Inference methods	*Bayesian Inference
	*Dempster-Shafer Method
	*Generalized Evidence Processing
Artificial intelligence methods	*Expert System
	*Adaptive Neural Network
	*Fuzzy Logic

3. Fusion based Visual Target Tracking for Autonomous Vehicles Navigation

Since each navigation method has its own merits and demerits, it is necessary to have a fusion methodology to combine their advantages and also compensate for their disadvantages. Also autonomous vehicles are equipped with different sensors (such as inertial motion sensors, camera, laser scan, sonar, radar, GPS and so on). The sensors data are redundant and complementary. Fusing these information can lead to a robust feature extraction and target tracking performance. Therefore in this section the general methods of sensor data fusion for visual target tracking by autonomous vehicles are reviewed.

In Table.1 Luo *et al.* [79] review the general data fusion methodologies. Data fusion algorithms can be broadly classified as follows: estimation methods, classification methods, inference methods, and artificial intelligence methods. Data fusion has been widely used in visual target tracking for autonomous vehicles navigation.

In the article of [80], a multi-sensor fusion approach between an omnidirectional vision system and a panoramic range finder is presented to dynamically localize an autonomous vehicle. These two sensors provide some complementary and redundant data which enable to construct a robust sensorial model which integrates an important number of significant primitives. Based on this model, they treat the problem of maintaining a matching and propagating uncertainties on each matched primitive. In details their localization method is based on the tracking of significant landmarks of the environment and integrating a multi-level uncertainty propagation stage based on the usage of the Dempster-Shafer theory. For

this purpose, they have developed a robust multi-criteria method which solves two problems linked to the tracking: the propagation of an uncertainty concerning the landmark tracks and the treatment of the apparition and momentary disappearance of a track.

In [81] Derrouich *et al.* present a new hybrid approach for range estimation that combines inertial and visual based technologies, which allows them to calculate the image-space distance between the autonomous vehicle and the edge lines of the $3D$ environment. Two frames obtained from the moving CCD video cameras and the outputs of the inertial tracking system which report the relative changes of orientations and accelerations between the two frames are integrated to estimate the image-space distance of different $3D$ points.

Using the Cellular Nonlinear Network (CNN) tracking method, with a combination of inertial and visual tracking technologies, Derrouich *et al.* [82] have presented a direct trigonometric range estimation method from a stream of images in the static scene for visual target tracking.

Marin-Hernandez *et al.* [83] describe how to integrate tracking functions on an autonomous vehicle in different situations: landmark tracking to guarantee the real-time robot localization, target tracking for a sensor based motion or obstacle tracking to control an obstacle avoidance procedure. These objects such as landmarks, targets, and obstacles have different characteristics. Moreover, normal tracking methods are not robust enough to work under varying environmental conditions. Tracking often fails when illumination undergoes significant changes. Some other situations where tracking methods fail are due to the cluttered background, changes in the model pose, occlusions of the target image or motion discontinuities in the object dynamics. Thus under the real-world environmental conditions, a single tracking method cannot deal with all the different tasks and situations presented in it. Several methods must be integrated on autonomous vehicles. In their paper, a tracker controller is in charge to activate a specific method with respect to the subject type and a transition model used to recover the tracking failures. This method is validated using four trackers based on: template differences, set of points, lines and snakes. This method requests the tracker to switch, either if a target is lost, by one quick but not so robust method, and a more robust method is activated to find again the target using local image measurements, or if a target switch is necessary, for example when the robot must turn around an edge after a corridor following execution. In such a way every target is associated with the more adapted tracker and with a recovery procedure in case of failures.
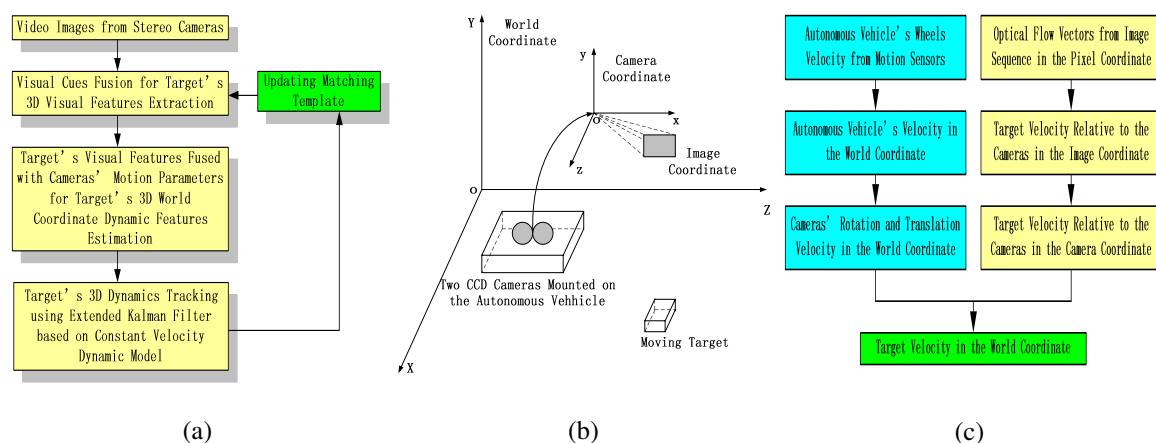
In the paper of [84] a multiple-sensor multiple-target tracking (MS-MTT) approach for the Autotaxi system is proposed. The system consists of two basic components: the sensor-level tracking and multiple-sensor track fusion center. Each sensor in the sensor level is considered as an intelligent one which generates its own track files. Thus, the task of the fusion center is to fuse the local track files to produce a more accurate and reliable single system track file. This is performed in three stages: data alignment, track-to-track association and track fusion. In their paper a decentralized sequential data association and fusion structure, referred to as the Sequential Minimum Normalized Distance Nearest Neighbor (SMNDNN) with the majority decision making MDW/OR logic method, has been considered for the Autotaxi application.

Another example of the implementation of sensor fusion algorithms for autonomous vehicles target tracking is proposed by [79]. The experimental setup of their system consists of one autonomous vehicle and one multi-sensor based electrical wheelchair. The autonomous vehicle (Nomad 200 platform) is

a three-wheel mobile platform equipped with a vertical sliding manipulating arm and other sensory modules. The experimental target is the multi-sensor based electrical wheelchair which is developed by them. Their system contains two major agents for the local decisions. One is the target-tracking agent whose inputs are the target position measurements from the fusion of the ultrasonic and vision sensors and the other is the collision-avoidance agent whose inputs are the surrounding range measurements from the fusion of the 16 ultrasonic sensors. The final decision calculated by fusion of these two local decisions is the absolute driving velocity of the autonomous vehicle.

Most recently Jia *et al.* [85, 86] propose a novel data fusion scheme for visual object identification and tracking by autonomous vehicles. In their scheme, image motion vectors fields, color features, visual disparity depth information and cameras motion parameters are fused together to identify the target $3D$ visual and dynamic features. Their paper also presents a detailed description of the $3D$ target tracking algorithm using an Extended Kalman Filter with a constant velocity dynamic model. The details of the their proposed system can be seen from Figure 10.

Figure 10. These figures are taken from [85]. Jia *et al.*'s proposed data fusion based target tracking system is described schematically in (a). The cameras are mounted on an autonomous vehicle moving in the environment as shown in (b). Estimation of the $3D$ velocity of the interested target in the world coordinate can be achieved using the procedures explained in (c).

In conclusion, data fusion improves the system performance significantly and make robust estimations for vision based target tracking by autonomous vehicles. Based on the Table.1 and above reviews, data fusion algorithms can be generally developed in two different ways: (1) fusing data from different sensors or different data from one sensor; (2) integrating different methods. Data fusion shows many promising advantages and should be paid more attentions to in the future research work.

4. Remaining Research Challenges and Future Research Directions

Although numerous researchers have made efforts to develop visual target tracking systems for autonomous vehicles navigation, many issues are still open and deserve further research, especially in the

following areas.

First, as the visual tracking system designed for autonomous vehicles works mainly in an unknown environment, the target's specific features are generally unavailable or unpredictable. Also as the cameras are moving with the vehicle, the visual features of the environment may also change continuously. For example as the light attenuates exponentially with the distance in the air in an outdoor environment, it makes the quality of the camera images poor. In such a situation the color of the object of interest will be attenuated and the normal tracking algorithms will meet great difficulties. Also there may be many unknown colorful objects in the same environment, which can be misinterpreted for the object of interest. So it is not suitable to develop a tracking system mainly based on the target's special visual features or models such as color or shape.

Secondly, in the normal visual target tracking systems the target's relative dynamics to the cameras is estimated for navigation purposes. This relative information is obtained from the vehicle's onboard sensors without considering the vehicle's self movement. The relative information is not available for other applications in the same environment. Also the relative information is less accurate due to the vehicle's self complex rotational and translational motions (such as pan and tilt motions), which are normally ignored in the estimation processes.

Thirdly, as the autonomous vehicle is working in both the structured and unstructured environment, it is not always applicable to put additional sensors in the environment to help autonomous vehicles track the objects and navigate the environment. Also when the autonomous vehicle works in an unknown or unexpected environment alone, it doesn't have a prior knowledge of itself location or the environment conditions. The $3D$ visual target tracking will be difficult when there are very few initial or beforehand information that can be used.

Fourthly, in the visual target tracking, the dynamic information aids the tracking process in the presence of occlusions and measurement noises [87]. However, the tracking is complicated by the fact that natural moving targets do not exhibit one type of motion but rather have complex, unknown, highly nonlinear or time-varying dynamics. The single model approaches do not make full use of the information available and often rely on somewhat ad hoc methods of incorporating uncertainty about the mode estimates (like adjusting the filter noise covariance in proportion to the variance in the model estimate [85]), so the traditional single target linear dynamic models are not quite applicable.

Fifthly, according to the object tracking survey paper [29], there are still some major remaining research challenges: 1. One challenge in tracking is to develop algorithms for tracking objects in unconstrained environments, for example, videos obtained from crowded streets. 2. An important issue that has been neglected in the development of tracking algorithms is integration of contextual information. For example, in a vehicle tracking application, the location of vehicles should be constrained to paths on the ground as opposed to vertical walls or the sky. A tracker that takes advantage of contextual information to incorporate general constraints on the shape and motion of objects will usually perform better than one that does not exploit this information. 3. Most of tracking algorithms require off-line training information about the target and/or the background. Such information is not always available. Moreover, as the object appearance or background varies, the discriminative features also vary. Thus, there is a need for online selection of discriminative features. 4. In a similar vein, most tracking algorithms use pre-specified models for object representation. The capability to learn object models online will greatly

increase the applicability of a tracker. 5. Among different probabilistic state-space methods including Kalman Filters, Joint Probabilistic Data Association Filter (JPDAFs), Hidden Markov Models (HMMs), and Dynamic Bayesian Networks (DBNs), DBNs are probably the most general method for representation of conditional dependencies between multiple variables and/or image observations. However there is a need for more efficient solutions for inference before DBNs are more commonly used in tracking applications.

Sixthly, for normal data fusion algorithms, all the sensors data can not arrive at the same time and different sensors have different data capturing and processing time. As a result, the fusion algorithm receives different measurements in different time stamps. There is the Out-Of-Sequence Measurements "OOSM" (or called delayed measurements) problem [88]. In Bar-Shalom's paper [88], it is also known as the "negative-time measurement update" problem. In the normal multi-sensor fusion based target tracking systems, the OOSM problem is often ignored or just simply solved by using some human reasoning methods like [89]. In order to make the accurate target state estimation, this Out-Of-Sequence Measurements (OOSM) problem must be solved.

Since there are the above remaining challenges, future work must be investigated before being able to establishing a system that produces quality visual target tracking results while overcoming all the challenges implicated by operation in the real-time environment. Following research directions are suggested for future further considerations:

- To make more general applications in different environments under both known or unknown situations, the target tracking scheme which only rely on the autonomous vehicle's onboard sensors can be developed as a better solution to detect and track the object's $3D$ dynamics.

- A dynamically changing background is considered in visual target tracking for autonomous vehicles navigation. The robust target tracking system needs to be developed in the MCMO state.

- Instead of using the target's dynamics relative to the cameras, a tracking algorithm which can accurately derive the target's $3D$ dynamic information in the world coordinate can be developed to give more applicable and accurate target tracking results.

- At present, many tracking algorithms are achieved by placing a specific visual property on the objects of interest. The limitation is mainly due to the finite object features or the models information. The research aiming at tracking the targets without placing any special property on the objects of interest can be considered for autonomous vehicles navigation tasks.

- In a MCMO state, the visual features of the object and the background are changing continuously. It is necessary to make the system have the ability to adapt to these changes, which means that the tracking algorithms should be developed having some learning or adaptive estimation strategies.

- The ideas of data fusion can be more thoroughly studied and applied in visual target tracking for autonomous vehicles navigation. With successful data fusion, the tracking system can easily and accurately achieve the tracking requirements for autonomous vehicles navigation.

- In order to accurately track the object's $3D$ dynamics in different situations, there are several well-known tracking problems to be solved.

1. First, SLAM can be integrated more closely at the same time with visual target tracking to make autonomous vehicles have total self-intelligence independent of the environment information.

2. Second, more complex and adaptive multiple dynamic models algorithm can be developed to approximate the target's motion properties.

3. Third, in order to make the accurate estimation, the OOSM problem must be considered for the real-time testing.

- Normally the visual target tracking system for autonomous vehicles is designed based on data fusion of optical features from the CCD cameras and the vehicle's inertial motion sensors data. In the future, CCD cameras fused with other sensors such as radar, laser, sonar or GPS can be more widely introduced into the tracking schemes. CCD cameras can be used to derive the target's $2D$ visual information, while the radar, laser or GPS can help to obtain the target's $3D$ depth information.

- For the general object tracking part, here there are several future research directions:

  1. Robust tracking algorithms under crowded conditions to better handle the problems such as occlusions or lost tracker.

  2. Tracking algorithms with the integration of contextual information.

  3. Online estimation of discriminative features and models for object tracking.

  4. More robust probabilistic state-space methods for tracking.

Finally and most importantly, due to large amount of the visual data, it is very difficult to realize the real-time processing on the hardware placed in small limited hulls in autonomous vehicles. Fast and robust tracking algorithms are required for the real-time performance.

5. Conclusion

Nowadays, autonomous vehicles are widely used and the need of interpreting the environment in thorough and smart ways is increasing for autonomous vehicles to fulfil different tasks in different applications. This paper shows that vision based target tracking can be an alternatively better solution for autonomous vehicles navigation.

Much has been accomplished in the vision based target tracking for autonomous vehicles navigation. This paper surveys the most interesting and important algorithms proposed in the last decade. First, visual target tracking schemes for autonomous land, underwater and aerial vehicles are reviewed separately. Second, since data fusion methodologies are widely used, data fusion based visual target tracking algorithms for autonomous vehicles navigation are reviewed.

Based on the reviews, remaining researching challenges are concluded and several future research directions are investigated. It can be predicted that in the near future great processes will be made to make autonomous vehicles fully automatically sense the environment and track the object of interest.

Data fusion makes better use of the redundant sensor information and estimates the target states more accurately. There is no doubt that data fusion will continuously be a promising and hot research topic.

More research work should be carried on to develop robust fusion algorithms for autonomous vehicles navigation.

References and Notes

1. Mattos, L.; Grant, E. Passive sonar applications: target tracking and navigation of an autonomous robot. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation.* May, 2004, Vol. 5, pp. 4265–4270.

2. Trepagnier, P.; Nagel, J.; Kinney, P.; Dooner, M.; Wilson, B.; Schneider, C.; Goeller, K. Navigation and control system for autonomous vehicles. *Patent: WO2008048707* **2008**.

3. Fujimura, K.; Liu, X. *Patent: US20087366325* **2008**.

4. DeSouza, G.; Kak, A. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2002**, *24*, 237–267.

5. Forsyth, D.; Ponce, J. *Computer Vision: A Modern Approach.* Prentice Hall, USA, 2002.

6. Li, T.-H.; Chang, S.-J.; Tong, W. Fuzzy target tracking control of autonomous mobile robots by using infrared sensors. *IEEE Transactions on Fuzzy Systems* **2004**, *12*, 491–501.

7. Rives, P.; Borrelly, J.-J. Underwater pipe inspection task using visual servoing techniques. In *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems, 1997. IROS '97.* September, 1997, Vol. 1, pp. 63–68.

8. Samarasekera, S.; Kumar, R.; Oskiper, T.; Zhu, Z.; Naroditsky, O.; Sawhney, H. Unified framework for precise vision-aided navigation. *Patent: US20080167814* **2008**.

9. Nadimi, S.; Bhanu, B. Multistrategy fusion using mixture model for moving object detection. In *Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems.* August, 2001, pp. 317–322.

10. Kodagoda, K. *Multi-Sensor Integration and Fusion in Mobile Robot Navigation.* Ph.d. thesis,, School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang, China, 2003.

11. Shamir, T. How should an autonomous vehicle overtake a slower moving vehicle: Design and analysis of an optimal trajectory. *IEEE Transactions on Automatic Control* **2004**, *49*, 607–610.

12. Qu, Z.; Wang, J.; Plaisted, C. A new analytical solution to mobile robot trajectory generation in the presence of moving obstacles. *IEEE Transactions on Robotics* **2004**, *20*, 978–993.

13. Yang, S.; Luo, C. A neural network approach to complete coverage path planning. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics* **2004**, *34*, 718–724.

14. Whittaker, W.; Johnston, J.; Ziglar, J. Obstacle detection arrangements in and for autonomous vehicles. *Patent: WO2008070205* **2008**.

15. Yamauchi, B. Autonomous mobile robot. *Patent: WO2008013568* **2008**.

16. Graf, B.; Hans, M.; Schraft, R. Mobile robot assistants. *IEEE Robotics and Automation Magazine* **2004**, *2*, 67–77.

17. Jensen, B.; Froidevaux, G.; Greppin, X.; Lorotte, A.; Mayor, L.; Meisser, M.; Ramel, G.; Siegwart, R. The interactive autonomous mobile system robox. In *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robotics and Systems.* October, 2002, Vol. 2, pp. 1221–1227.

18. Katsura, S.; Ohnishi, K. Human cooperative wheelchair for haptic interaction based on dual com-

pliance control. *IEEE Transactions on Industrial Electronics* **2004**, *51*, 221–228.

19. Jung, B.; Sukhatme, G. A region-based approach for cooperative multi-target tracking in a structured environment. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robotics and Systems*. October, 2002, Vol. 3, pp. 2764–2769.

20. Balasuriya, A.; Ura, T. Underwater robots for cable following. In *Proceedings. 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2001, Vol. 4, pp. 1811–1816.

21. Yang, F. *Vision-Based Target Tracking and Navigation System for Autonomous Underwater Vehicles*. Master thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang, China, 2001.

22. Balasuriya, B. *Computer Vision for Autonomous Underwater Vehicle Navigation*. Ph.D. Thesis, Department of Naval Architecture and Ocean Engineering, Faculty of Engineering, University of Tokyo, Tokyo, Japan, 1998.

23. Scholz, G.; Rahn, C. Profile sensing with an actuated whisker. *IEEE Transactions on Robotics and Automation* **2004**, *20*, 124–127.

24. Marques, C.; Lima, P. Avoiding obstacles-multisensor navigation for nonholonomic robots in cluttered environments. *IEEE Robotics and Automation Magazine* **2004**, *11*, 70–82.

25. Skrzypczyk, K. Game theory based target following by a team of robots. In *Proceedings of the Fourth International Workshop on Robot Motion and Control*. June, 2004, pp. 91–96.

26. Zhang, Y.; Wang, J. Obstacle avoidance for kinematically redundant manipulators using a dual neural network. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics* **2004**, *34*, 752–759.

27. Ogren, P.; Leonard, N. A convergent dynamic window approach to obstacle avoidance. *IEEE Transactions on Robotics* **2005**, *21*, 188–195.

28. Lamiraux, F.; Bonnafous, D.; Lefebvre, O. Reactive path deformation for nonholonomic mobile robots. *IEEE Transactions on Robotics* **2004**, *20*, 967–977.

29. Yilmaz, A.; Javed, O.; Shah, M. Object tracking: A survey. *ACM Computing Surveys (CSUR)* **2006**, *38*.

30. Dellaert, F.; Burgard, W.; Fox, D.; Thrun, S. Using the condensation algorithm for robust, vision-based mobile robot localization. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Ft. Collins, CO, USA, 1999, Vol. 2, pp. 558–594.

31. Murrieta-Cid, R.; Briot, M.; Vandapel, N. Landmark identification and tracking in natural environment. In *Proceedings of 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 1998, Vol. 1, pp. 179–184.

32. Sim, R.; Dudek, G. Mobile robot localization from learned landmarks. In *Proceedings of 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 1998, Vol. 2, pp. 1060–1065.

33. Li, S.; Tsuji, S. Finding landmarks autonomously along a route. In *Proceedings of the 11 IAPR International Conference on Computer Vision and Applications*. September, 1992, pp. 316–319.

34. Briggs, A.; Scharstein, D.; Braziunas, D.; Dima, C.; Wall, P. Mobile robot navigation using self-similar landmarks. In *Proceedings of ICRA '00. IEEE International Conference on Robotics and Automation*. April, 2000, Vol. 2, pp. 1428–1434.

35. Jang, G.; Lee, S.; Kweon, I. Color landmark based self-localization for indoor mobile robots. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 1, pp. 1037–1042.

36. Wei, R.; Mahony, R.; Austin, D. A bearing-only control law for stable docking of unicycles. In *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2003, Vol. 3, pp. 3793–3798.

37. Breed, D. S. Autonomous vehicle travel control systems and methods. *Patent: US20080161986* **2008**.

38. Hirai, N.; Mizoguchi, H. Visual tracking of human back and shoulder for person following robot. In *Proceedings of the 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*. July, 2003, pp. 527–532.

39. Lee, J.-H.; Morioka, K.; Hashimoto, H. Physical distance based human robot interaction in intelligent environment. In *Proceedings of IEEE 2002 28th Annual Conference of the IECON [Industrial Electronics Society]*. November, 2002, Vol. 4, pp. 3068–3073.

40. Morioka, K.; Lee, J.; Hashimoto, H. Human-following mobile robot in a distributed intelligent sensor network. *IEEE Transactions on Industrial Electronics* **2004**, *51*, 229–237.

41. Morioka, K.; Lee, J.; Hashimoto, H. Physical agent for human following in intelligent sensor network. In *Proceedings of 2002 IEEE/RSJ International Conference on Intelligent Robotics and Systems*. October, 2002, Vol. 2, pp. 1234–1239.

42. Nishiwaki, K.; Kagami, S.; Kuffner, J.; Inaba, M.; Inoue, H. Online humanoid walking control system and a moving goal tracking experiment. In *Proceedings Of the 2003 IEEE International Conference on Robotics and Automation*. September, 2003, pp. 911–916.

43. Kwon, H.; Yoon, Y.; Park, J.; Kak, A. Person tracking with a mobile robot using two uncalibrated independently moving cameras. In *Proceedings of 2005 IEEE International Conference on Robotics and Automation*. April, 2005, pp. 2877–2883.

44. Ogata, T.; Matsuda, S.; Tan, J.; Ishikawa, S. Real-time human motion recognition by aerial robot. In *Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics*. October, 2004, Vol. 6, pp. 5290–5295.

45. Doi, M.; Nakakita, M.; Aoki, Y.; Hashimoto, S. Real-time vision system for autonomous mobile robot. In *Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication*. September, 2001, pp. 442–449.

46. Wang, C.-C. Simultaneous localization, mapping and moving object tracking. *Ph.D. Thesis, Robotics Institute, Carnegie Mellon University* **2004**.

47. Hajjawi, M.; Shirkhodaie, A. Cooperative visual team working and target tracking of mobile robots. In *Proceedings of the Thirty-Fourth Southeastern Symposium on System Theory*. March, 2002, pp. 376–380.

48. Burschka, D.; Hager, G. Principles and practice of real-time visual tracking for navigation and mapping. May, 2004, pp. 1–8.

49. Dao, N.; You, B.-J.; Oh, S.-R.; Hwangbo, M. Visual self-localization for indoor mobile robots using natural lines. In *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2003, Vol. 2, pp. 1252–1257.

50. Wang, C.-C.; Thorpe, C. Simultaneous localization and mapping with detection and tracking of moving objects. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 3, pp. 2918–2924.

51. Karlsson, H.; Nygards, J. Robust and efficient tracking in image sequences using a kalman filter and an affine motion model. In *Proceedings of 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2002, pp. 98–104.

52. Canals, R.; Roussel, A.; Famechon, J.-L.; Treuillet, S. A biprocessor-oriented vision-based target tracking system. *IEEE Transactions on Industrial Electronics* **2004**, *49*, 500–506.

53. Tomono, M. Path planning with target finding by single camera under map uncertainty. In *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*. June, 2003, pp. 465–470.

54. Zhang, H.; Yuan, K.; Liu, J.-D. A fast and robust vision system for autonomous mobile robots. In *Proceedings of 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*. October, 2003, Vol. 1, pp. 60–65.

55. Yu, J.; Wang, S.; Tan, M. A parallel algorithm for visual tracking of multiple free-swimming robot fishes based on color information. In *Proceedings of 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*. October, 2003, Vol. 1, pp. 359–364.

56. Mazo, M. J.; Speranzon, A.; Johansson, K.; Hu, X. Multi-robot tracking of a moving object using directional sensors. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*. May, 2004, Vol. 2, pp. 1103–1108.

57. Betser, A.; Vela, P.; Tannenbaum, A. Automatic tracking of flying vehicles using geodesic snakes and kalman filtering. In *Proceedings of 43rd IEEE Conference on Decision and Control*. December, 2004, Vol. 2, pp. 1649–1654.

58. Ng, W.; Leng, G.; Low, Y. Coordinated movement of multiple robots for searching a cluttered environment. In *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2004, Vol. 1, pp. 400–405.

59. Tan, T.; Baker, K. Efficient image gradient based vehicle localization. *IEEE Transaction on Image Processing* **2000**, *9*, 1343–1356.

60. Gonzalez-Banos, H.; Lee, C.-Y.; Latombe, J.-C. Real-time combinatorial tracking of a target moving unpredictably among obstacles. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 2, pp. 1683–1690.

61. Lee, C.-Y.; Gonzalez-Banos, H.; Latombe, J.-C. Real-time tracking of an unpredictable target amidst unknown obstacles. In *Proceedings of Seventh International Conference on Control, Automation, Robotics And Vision (ICARCV'O2)*. December, 2002, Vol. 4, pp. 596–601.

62. Ding, J.; Kondou, H.; Kimura, H.; Hada, Y.; Takase, K. Robust tracking for camera control on an irregular terrain vehicle. In *Proceedings of the 41st SICE Annual Conference on SICE 2002*. August, 2002, Vol. 2, pp. 1191–1196.

63. Burschka, D.; Hager, G. Scene classification from dense disparity maps in indoor environments. In *Proceedings of 16th International Conference on Pattern Recognition*. August, 2002, Vol. 3, pp. 708–712.

64. Mori, R.; Miyazaki, F. Gag (gaining angel of gaze) strategy for ball tracking and catching task.

In *Proceedings of 2002 IEEE/RSJ International Conference on Intelligent Robotics and Systems*. October, 2002, Vol. 1, pp. 281–286.

65. Yu, Q.; Yang, M.; Wang, H. Real-time vision based ego-motion estimation for outdoor mobile robot. In *Proceedings of the 5th World Congress on Intelligent Control and Automation*. June, 2004, Vol. 6, pp. 4720–4724.

66. Murrieta, R.; Sarmiento, A.; Hutchinson, S. On the existence of a strategy to maintain a moving target within the sensing range of an observer reacting with delay. In *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2003, Vol. 2, pp. 1184–1191.

67. Higgins, C.; Pant, V. A biomimetic vlsi sensor for visual tracking of small moving targets. *IEEE Transactions on Circuits and Systems I [see also IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications]* **2004**, *51*, 2384–2394.

68. Dang, T.; Hoffmann, C.; Stiller, C. Fusing optical flow and stereo disparity for object tracking. In *Proceedings of the 5th IEEE International Conference On Intelligent Transportation Systems*. September, 2002, pp. 112–117.

69. Saripalli, S.; Montgomery, J.; Sukhatme, G. Visually guided landing of an unmanned aerial vehicle. *IEEE Transactions on Robotics and Automation* **2003**, *19*, 371–380.

70. Miller, R.; Mettler, B.; Amidi, O. Carnegie mellon university's 1997 international aerial robotics competition entry. In *Proceedings of International Aerial robotics Competition*, 1997.

71. Sinopoli, B.; Micheli, M.; Donato, G.; Koo, T. Vision based navigation for an unmanned aerial vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*. Seoul, Korea, 2001, Vol. 2, pp. 1757–1764.

72. Shakernia, O.; Ma, Y.; Koo, T.; Sastry, S. Landing an unmanned air vehicle: vision based motion estimation and non-linear control. *Asian Journal of Control* **1999**, *1*, 128–145.

73. Garcia-Padro, P.; Sukhatme, G.; Montgomery, J. Toward vision-based safe landing for an autonomous helicopter. *Robotics and Autonomous System* **2001**, *38*, 19–29.

74. Sharp, C.; Shakernia, O.; Sastry, S. A vision system for landing an unmanned aerial vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*. Seoul, Korea, 2001, Vol. 2, pp. 1720–1727.

75. Shakernia, O.; Vidal, R.; Sharp, C.; Ma, Y.; Sastry, S. Multiple-view motion estimation and control for landing an unmanned aerial vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 3, pp. 2793–2798.

76. Saripalli, S.; Montgomery, J.; Sukhatme, G. Vision based autonomous landing of an unmanned aerial vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 3, pp. 2799–2804.

77. Saripalli, S.; Sukhatme, G. Landing on a moving target using an autonomous helicopter. In *Proceedings of the International Conference on Field and Service Robotics*. July, 2003.

78. Stolle, S.; Rysdyk, R. Flight path following guidance for unmanned air vehicles with pan-tilt camera for target observation. In *Proceedings of The 22nd Digital Avionics Systems Conference*. October, 2003, Vol. 2, pp. 8.B.3–81–12.

79. Luo, R.; Yih, C.; Su, K. Multisensor fusion and integration: Approaches, applications, and future research directions. *IEEE Sensors Journal* **2002**, *2*, 107–119.

80. Clerentin, A.; Delahoche, L.; Brassart, E.; Cauchois, C. Mobile robot localization based on multi target tracking. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*. May, 2002, Vol. 1, pp. 13–18.

81. Derrouich, S.; Izumida, K.; Shiiya, K. A combination of monocular ccd camera and inertial-sensor for range estimation. In *Proceedings of IEEE 2002 28th Annual Conference of the IECON [Industrial Electronics Society]*. November, 2002, Vol. 3, pp. 2191–2196.

82. Derrouich, S.; Izumida, K.; Murao, K.; Shiiya, K. The use of cnn models and vertical rectification for a direct trigonometric recovery of 3d scene geometry from a stream of images. In *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*. November, 2003, Vol. 2, pp. 1600–1604.

83. Marin-Hernandez, A.; Ayala-Ramireze, V.; Devy, M. Collaborative tracking for robotic tasks. In *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October, 2003, Vol. 1, pp. 272–277.

84. Escamilla-Ambrosio, P.; Lieven, N. A multiple-sensor multiple-target tracking approach for the autotaxi system. In *Proceedings of 2004 IEEE Intelligent Vehicles Symposium*. June, 2004, pp. 601–606.

85. Jia, Z.; Balasuriya, A.; Challa, S. Visual information and camera motion fusion for 3d target tracking. In *Proceedings of The Eighth International Conference on Control, Automation, Robotics and Vision*. December, 2004, pp. 2296–2301.

86. Jia, Z. Vision based target tracking for autonomous vehicles navigation. *Ph.D. Thesis, School of EEE, Nanyang Technological University, Singapore* **2006**.

87. Jeong, M.; Kuno, Y.; Shimada, N.; Shirai, Y. Two-hand gesture recognition using coupled switching linear model. In *Proceedings of the 16th International Conference on Pattern Recognition*. August, 2002, Vol. 3, pp. 529–532.

88. Bar-Shalom, Y. Update with out-of-sequence measurements in tracking: Exact solution. *IEEE Transactions on Aerospace and Electronic Systems* **2002**, *38*, 769–777.

89. Kosaka, A.; Meng, M.; Kak, A. Vision-guided mobile robot navigation using retroactive updating of position uncertainty. In *Proceedings of 1993 IEEE International Conference on Robotics and Automation*. May, 1993, Vol. 2, pp. 1–7.