International Journal of Advanced Engineering Application
Volume No.2 Issue No 1 Jan 2025
ISSN NO:3048-6807

Challenges and Solutions in Vision-Based Target Tracking for Autonomous Systems

Ahmed S. El-Khidir[1] and Fatima R. Al-Mahdi[2]
[1]Department of Electrical Engineering, University of Tripoli, Tripoli, Libya
[2]Department of Computer Science, University of Tripoli, Tripoli, Libya

Abstract:
This paper provides a comprehensive survey of advancements made over the past decade in the field of vision-based target tracking for autonomous vehicle navigation. The introduction begins by highlighting the motivations and wide-ranging applications of vision-based target tracking in autonomous vehicle navigation. These applications span various domains, underscoring the necessity for the development of robust algorithms capable of handling the diverse challenges faced by autonomous vehicles in dynamic environments. The discussion establishes that creating resilient vision-based tracking solutions is crucial for the efficient operation of autonomous systems. The review is organized into three primary categories: land, underwater, and aerial vehicles. Each category explores the specific techniques and methodologies developed for target tracking within its respective domain. For land-based autonomous vehicles, the focus is on approaches that manage obstacles, uneven terrains, and dynamic road conditions. In the context of underwater vehicles, challenges such as poor visibility, varying water conditions, and the need for energy-efficient operations are examined. For aerial vehicles, the discussion highlights the importance of precise tracking in three-dimensional space, which is critical for applications like surveillance, delivery, and disaster response. Additionally, the paper delves into the growing trend of integrating data fusion techniques to enhance the performance and robustness of vision-based target tracking systems. By combining data from multiple sensors and modalities, data fusion helps address limitations like occlusion, noise, and environmental variability, thereby improving tracking accuracy and reliability. Finally, the paper identifies several research challenges that remain unresolved, including issues like computational efficiency, real-time processing, and adapting to highly dynamic environments. It also explores potential future research directions, such as leveraging advancements in artificial intelligence, deep learning, and multi-modal data integration to further enhance the capabilities of vision-based target tracking systems for autonomous navigation.

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, 5, 6, 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$3D) scene analysis [5]. For example, a system and method is presented in [8] for efficiently locating in $3D$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].

(a) (b) (c)

Figure 1. Some applications of the visual target tracking system for autonomous vehicles navigation. (a) shows the general high way system. 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.

In intelligent transportation systems (ITS), object extraction and tracking serve as essential prerequisites for advancing intelligent autonomous vehicles and mobile robots (Figure 1). These processes enable various applications, including forward collision avoidance, where visual tracking of moving objects helps identify potential collision threats based on their relevance to the vehicle's intended path [10]. Beyond collision avoidance, visual target tracking can assist human drivers. For instance, it is instrumental in drowsy driver warning systems, where tracking the movement of objects around the vehicle allows driver assistance systems to alert drivers to potential collisions and other dangers [10].

Another application within ITS is the design of optimal trajectories for vehicles under normal driving conditions, such as overtaking a slower-moving vehicle on a predetermined road [11]. These applications demonstrate the multifaceted utility of object tracking in enhancing road safety and vehicle performance.

Path Planning and Autonomous Vehicle Navigation Path planning, a critical component of autonomous vehicle operations, involves the real-time design of a collision-free path while navigating dynamically moving objects within a limited sensing range [12]. This task is fundamental to robotics, requiring vehicles to be aware of their surroundings. One specialized application is Complete Coverage Path Planning (CCPP), also known as region filling or area covering, often used by cleaning robots in two-dimensional (2D) environments. In this context, visual target tracking ensures the vehicle traverses every area in its workspace while avoiding obstacles [13].

Patent innovations, such as those described in [14], present arrangements for obstacle detection in autonomous vehicles. These patents leverage advanced data manipulation techniques to enhance the accuracy of obstacle detection, thereby improving the operational efficiency of autonomous vehicles. For example, [15] details a mobile robot equipped with a range finder and stereo vision system, capable of autonomous navigation through urban terrain, map generation, and reconnaissance operations.

Human Assistance Mobile Robotics Visual target tracking extends to human assistance robotics, where mobile robots equipped with communication, interaction, and behavioral mechanisms operate in environments shared with humans, such as museums or exposition areas [16, 17]. In these dynamic environments, obstacle avoidance becomes critical to ensure safe human-robot interaction. Collision avoidance problems in such settings can be addressed using position-based interaction techniques facilitated by visual target tracking [18].

Security and Surveillance Applications For security and surveillance, efficient positioning of multiple sensors is often required to cover large, structured environments. Autonomous vehicle-based trackers offer a compelling solution, reducing the number of sensors needed in a tracking network. These trackers can adapt to target movements and dynamic environmental changes by repositioning themselves as needed [19]. Such capabilities make them invaluable for monitoring and response in security scenarios.

Underwater Environments and Autonomous Underwater Vehicles (AUVs) In underwater environments, visual target tracking is crucial for Autonomous Underwater Vehicle (AUV) navigation. For example, observing the behavioral patterns of marine life requires AUVs to follow and stay close to subjects, such as fish, over extended periods. Similarly, underwater exploration and maintenance tasks, like routine inspections of man-made systems, necessitate precise tracking and observation capabilities [21] (Figure 2).

Research by Arjuna and Tamaki [20] introduces a sensor fusion technique for tracking and following underwater cables using video images, tested on the Twin-Burger 2 vehicle at the University of Tokyo (Figure 2). Further developments by Arjuna et al. [22] propose vision-based tracking systems for underwater docking applications. Other applications include obstacle avoidance and object identification in unstructured underwater environments. For instance, the U.S. Navy is investigating AUVs to accurately detect and classify underwater mines, critical for pre-landing operations in hostile territories [23].

These examples illustrate the extensive potential of visual target tracking across various environments, highlighting its integral role in enabling autonomous navigation, safety, and functionality across a wide spectrum of applications.

(a) (b)

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]).

- 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 3*D*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

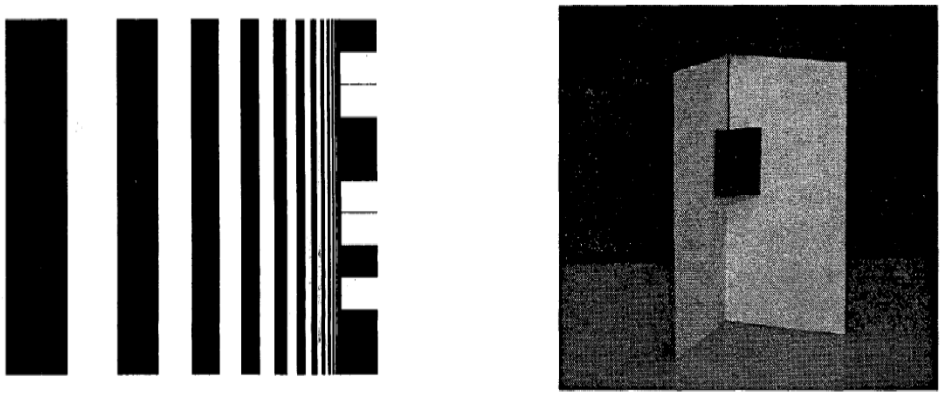
In general applications of visual target tracking for Autonomous Land Vehicles navigation include some sorts of landmarks tracking, people following, multiple mobile robots cooperation, vehicles localization and map building, visual target tracking with pan-tilt cameras platforms and so on. According 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, 31, 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.



(a) (b)

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.

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**, **40**, **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.

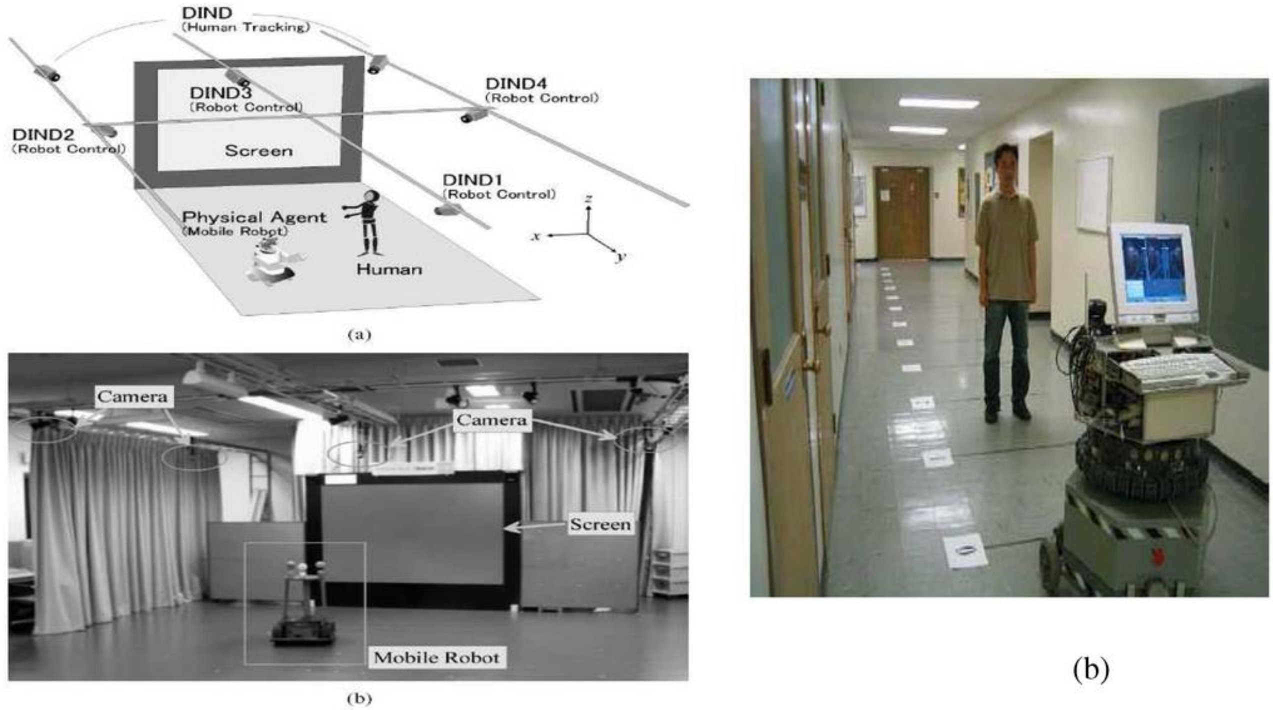
In [**42**] Nishiwaki *et al.* present a humanoid walking control system that generates body trajectories to follow a given desired motion on-line. They implement the system by making the autonomous vehicle track and follow a moving person based on the stereo vision feedback. The visual tracking consists of 3 parts: a stereo vision processing for target detection and $3D$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.



(b)

(a)

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 visionbased 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 3*D*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.

(a)

(b)

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).

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$2D visual camera control requires a priori knowledge of the $3D$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 specifical 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.

5. Conclusion

In recent years, autonomous vehicles have seen widespread adoption across various industries and applications. As their utilization grows, the demand for advanced techniques to interpret and interact with the environment becomes increasingly critical. Autonomous vehicles need to process environmental data in smart, efficient, and robust ways to perform diverse tasks reliably. This paper demonstrates that vision-based target tracking offers a promising and effective solution for autonomous vehicle navigation, with significant potential for improvement and application.

Over the past decade, considerable progress has been made in the development of vision-based target tracking systems for autonomous vehicles. This paper surveys some of the most notable and impactful algorithms from this period. First, it categorizes and reviews visual target tracking schemes tailored for land, underwater, and aerial vehicles, emphasizing the unique requirements and challenges associated with each domain. Second, it highlights the growing importance of data fusion methodologies, showcasing their role in enhancing the performance and robustness of visual target tracking systems for autonomous navigation.

Key Findings and Research Challenges

While notable advancements have been achieved, several challenges remain that must be addressed to fully realize the potential of autonomous vehicles. One such challenge is ensuring the systems' ability to operate effectively in diverse and unpredictable environments. Variations in lighting, weather, and terrain conditions can affect the performance of vision-based systems, and addressing these factors will require further innovation in robust algorithm design.

Another key challenge is the integration of data from multiple sensors using data fusion techniques. By combining redundant information from various sources, data fusion enhances accuracy and reliability in target state estimation. However, developing algorithms that can adapt dynamically to changing conditions and computational constraints remains an area ripe for exploration. Robust data fusion methods are essential to ensuring that autonomous vehicles can respond effectively to real-time environmental inputs.

Future Research Directions

The future of vision-based target tracking in autonomous vehicles holds great promise. As technology advances, the following areas are expected to witness significant progress:

1. **Improved Algorithms:** Developing algorithms that can handle complex and dynamic scenarios with high computational efficiency is a pressing need. Future research should focus on improving the adaptability and precision of vision-based tracking systems across different vehicle types and environments.

2. **Enhanced Data Fusion Techniques:** Continued exploration into sophisticated data fusion methodologies will enable more accurate and reliable target tracking. Techniques that seamlessly integrate diverse sensor data while accounting for uncertainties will be pivotal.

3. **Real-Time Processing and Decision-Making:** Advancements in real-time processing capabilities will allow autonomous vehicles to interpret their surroundings and make decisions faster and more effectively. Leveraging the latest hardware innovations, such as GPUs and dedicated AI accelerators, will be key to achieving this.

4. **Interdisciplinary Applications:** Beyond transportation, vision-based target tracking can be extended to other domains, such as robotics, defense, and environmental monitoring. Investigating cross-disciplinary applications can open new pathways for innovation.

5. **Autonomous Collaboration:** Exploring systems that allow multiple autonomous vehicles to collaborate in tracking and navigation tasks could lead to breakthroughs in fleet management and coordinated operations.

Vision-based target tracking represents a pivotal component in the evolution of autonomous vehicles. With the continued development of robust algorithms and data fusion techniques, it is anticipated that autonomous vehicles will achieve a higher degree of environmental awareness, enabling them to sense their surroundings and track objects of interest with greater accuracy and efficiency. In the near future, advancements in this field will pave the way for autonomous systems that operate seamlessly and intelligently, meeting the demands of diverse and complex applications.

References

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 compliance 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, visionbased 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 selfsimilar 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; 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.; Gonza1ez-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.