Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE DÉFENSE

Development of a Vision-Based Robotic Follower Vehicle

J. Giesbrecht

Defence R&D Canada

Technical Report

DRDC Suffield TR 2009-026

February 2009

Canada

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **FEB 2009**	2. REPORT TYPE	3. DATES COVERED
4. TITLE AND SUBTITLE **Development of a Vision-Based Robotic Follower Vehicle**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence R&D Canada - Suffield,PO Box 4000, Station Main,Medicine Hat, AB,Canada T1A 8K6, ,**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**84**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Development of a Vision-Based Robotic Follower Vehicle

J. Giesbrecht
Defence R&D Canada – Suffield

Defence R&D Canada – Suffield

Technical Report

DRDC Suffield TR 2009-026

February 2009

Principal Author

Original signed by J. Giesbrecht

J. Giesbrecht

Approved by

Original signed by D. Hanna

D. Hanna
Head/AISS

Approved for release by

Original signed by Dr. P.A. D'Agostino

Dr. P. A. D'Agostino
Head/Document Review Panel

Abstract

This report provides an overview of the development of a vision-based leader/follower robotic vehicle at Defence R&D Canada – Suffield, with the eventual goal of autonomous convoying for military logitics. The experimental system uses a pan/tilt/zoom camera to track a lead vehicle or human, estimating the leader’s path and following it autonomously. This vision-based approach frees the system from reliance on GPS, radios, and active sensing equipment necessary for current leader/follower systems. Included in this report are the details of the computer vision, camera control, and vehicle control algorithms, as well as the results of field trials of the camera tracking system. Finally, it reports on experiments with the complete follower system following other vehicles and even dismounted humans.

Résumé

Ce rapport est une vue d’ensemble de la mise au point, à R&D pour la défense Canada – Suffield, d’un véhicule robotisé muni d’un système prédécesseur / suiveur à vision artificielle. Le but est d’aboutir éventuellement à des systèmes d’escorte autonome en matière de logistique militaire. Ce système expérimental utilise une caméra dotée de fonctions de pivotement horizontal et d’inclinaison verticale ainsi que d’un zoom pour retracer le trajet d’un véhicule ou d’un humain prédécesseur et d’estimer ce trajet pour être en mesure de suivre le prédécesseur de manière autonome. Cette méthode basée sur la vision artificielle libère le système qui ne dépend plus d’un GPS, de communication radio ni des systèmes de télédétection actifs qui sont nécessaires aux systèmes actuels prédécesseur / suiveur. Les détails des algorithmes de vision artificielle, de commande de la caméra et de commande des véhicules sont inclus dans ce rapport ainsi que les résultats des expériences sur le terrain du système de suivi à l’aide de caméras. On y documente enfin les expériences sur le système complet de suivi d’autres véhicules et même d’humains se déplaçant à pied.

This page intentionally left blank.

Executive summary

Development of a Vision-Based Robotic Follower Vehicle

J. Giesbrecht; DRDC Suffield TR 2009-026; Defence R&D Canada – Suffield; February 2009.

Background: In modern military conflicts, support vehicles and their drivers have become much more vulnerable than in the past to roadside bombs and Improvised Explosive Devices (IEDs). If convoys could be composed at least partly of unmanned vehicles, fewer soldiers would need to be put at risk. Current robotic leader/follower vehicles rely on GPS, radio communications, and active sensing to relay positional information from a leader to a follower vehicle. In addition to being easily detected by a savvy enemy, these systems are also vulnerable to our own electronic countermeasures against IED attacks. Therefore, a leader/follower system inspired by human driving is being developed which relies on computer vision and a pan/tilt/zoom camera to track and estimate the path taken by a lead vehicle or human. This approach requires no special equipment on the leader vehicle, no reliance on GPS, and only minimal sensing and computing on the follower vehicle.

Principal Results: This work developed a number of significant sub-components to enable robotic vision-based leader/follower:

- A set of computer vision algorithms capable of being trained at run-time to not only track an arbitrary leader vehicle, but also to estimate its position in the world.
- A set of pan/tilt/zoom control algorithms which can maintain the camera's attention on the leader vehicle despite the motion of both the leader and follower vehicles.
- A vehicle control algorithm to allow the robot to drive the leader's estimated path.

Technical details of each of the above are given. In addition, the result of one field demonstration following a lead vehicle at speeds of up to 10km/h, over a distance of 7km are shown. Further tests with the robotic vehicle following a dismounted human are also presented.

Significance of Results: This project accomplished one of the first demonstrations of vision-based vehicle following anywhere in the world. It is especially significant that an arbitrary leader can be chosen at run-time, and was successfully shown to follow commercial trucks, other robotic vehicles and even humans. The research shows great promise and potential long term pay-off in protecting Canadian soldiers in the battlefields of the future.

Future Plans: In order to become practically effective, the leader follower system needs to be improved. Firstly, the reliability of the computer vision system will be increased by the addition of more sensing modalities, such as extra cameras, a wider variety of computer vision algorithms, and infrared capabilities. Adaptive data filtering and vehicle control schemes will increase the driving speed of the following system. As a long term goal, the leader/follower system will be implemented on a military logistic vehicle to demonstrate its usefullness.

Sommaire

Development of a Vision-Based Robotic Follower Vehicle

J. Giesbrecht ; DRDC Suffield TR 2009-026 ; R & D pour la défense Canada – Suffield ; février 2009.

Contexte : Les véhicules logistiques et leurs chauffeurs sont beaucoup plus vulnérables aux bombes et aux dispositifs explosifs de circonstance (IED) durant les conflits militaires modernes que par le passé. Moins de soldats seraient en danger si les convois pouvaient se composer au moins en partie de véhicules sans équipage. Les véhicules robotisés actuels qu'ils soient le prédécesseur ou le suiveur dépendent des GPS, des communications radio et des systèmes de télédétection actifs pour relayer l'information d'un véhicule prédécesseur à un véhicule suiveur. Non seulement, ces systèmes peuvent-ils être facilement détectés par un ennemi dangereux mais ils peuvent succomber à nos propres contremesures électroniques des attaques IED. Un système de véhicule prédécesseur/ suiveur, inspiré de la conduite humaine, est en voie de mise au point et consiste à obtenir la vision artificielle d'une caméra dotée de fonctions de pivotement horizontal et d'inclinaison verticale ainsi que d'un zoom pour retracer et estimer le trajet d'un véhicule ou d'un humain prédécesseur. Cette méthode ne requiert pas d'équipement spécial sur le prédécesseur, ne dépend pas d'un GPS mais seulement d'un système minimum computationnel et de télédétection actif sur le véhicule suiveur.

Résultats principaux : Ces travaux ont abouti à la mise au point d'un certain nombre de sous-éléments importants mettant en service les véhicules prédécesseur / suiveur fonctionnant sur le principe d'une vision artificielle dont :

– un ensemble d'algorithmes de vision artificielle capable d'être exercée au moment de l'exécution pour non seulement retracer un véhicule prédécesseur arbitraire mais aussi pour estimer sa position géographique ;
– un ensemble d'algorithmes de commande des fonctions de pivotement horizontal et d'inclinaison verticale ainsi que de zoom en mesure de maintenir l'attention de la caméra sur le véhicule prédécesseur bien que les deux véhicules (prédécesseur et suiveur) soient en motion et
– un algorithme de contrôle de véhicule permettant à un robot de prendre le chemin estimé du véhicule prédécesseur.

Les détails techniques de chacun des sous-éléments décrits ci-dessus y sont inclus. De plus, on y produit les résultats d'une démonstration sur le terrain d'un véhicule prédécesseur à des vitesses allant jusqu'à 10 km/h sur une distance de 7 km. On y présente aussi les essais ultérieurs sur les véhicules robotisés qui font le suivi d'un humain se déplaçant à pied.

Portée des résultats : Ce projet a accompli l'une des premières démonstrations de suivi effectuées par un véhicule à vision artificielle et capable d'accomplir ce suivi dans le monde entier. Il est particulièrement important qu'on puisse choisir un véhicule prédécesseur arbitrairement au moment de l'exécution et on a réussi à démontrer la capacité à suivre des camions utilitaires, d'autres véhicules robotisés et même des humains. La recherche indique

qu'il existe un potentiel de rentabilité à long terme très prometteur dans le domaine de la protection des soldats canadiens sur les champs de bataille du futur.

Perspectives d'avenir : Il faut améliorer le système prédécesseur / suiveur pour qu'il soit efficient au niveau pratique. Il faut d'abord augmenter la fiabilité de la vision artificielle en ajoutant des modalités de détection telles que des caméras supplémentaires, une plus grande variété d'algorithmes de vision artificielle et des capacités infrarouge. Des schémas adaptifs de filtrage de données et de commande de véhicules prédécesseur / suiveur augmenteront la vitesse du système suiveur. Un but à long terme est de démontrer l'utilité de ce système en l'implémentant sur des véhicules de logistique militaire.

This page intentionally left blank.

Table of contents

List of figures

This page intentionally left blank.

1 Introduction

Robotic vehicles have the potential to save lives. In the current asymmetric military environment in which battlefields are no longer characterized by fronts, but rather guerrilla style warfare, logistics vehicles and their drivers have become much more vulnerable than in the past. Roadside bombs and Improvised Explosive Devices (IEDs) have become a favorite weapon of insurgents. If logistics vehicles were able to follow their leader vehicles autonomously, the drivers could focus on situational awareness and defence, or perhaps be removed from the vehicle entirely to ride in a safer, more hardened vehicle.

Similarly, the dismounted soldier in any armed conflict can never carry as much equipment and supplies as required. A personal mule follower robot could provide the carrying capacity for critical supplies that could make the difference in a combat situation, providing that extra box of ammunition or that extra ration.

To create autonomous systems like these, current robotic convoying and leader/follower mules rely on the transmission of waypoint coordinates between the leader and follower, requiring computer, radio and GPS equipment on both the leader and follower units [1, 2, 3]. If the robot were able to follow its leader using only a camera it would reduce the cost and complexity of such systems, and allow the follower robot to naturally follow any specified object, be it vehicle or human.

1.1 Pan/Tilt/Zoom Tracking for UGVs

This work was undertaken within the DRDC Intelligent Logistics Advanced Research Project (ARP). The goal is to develop a vision based robotic leader/follower system, with the eventual goal of autonomous convoying of large logistics vehicles. Two main subcomponents are required to make this possible:

1. A pan/tilt/zoom camera system capable of recognizing the leader and continually estimating the leader's position despite motion of the follower vehicle.

2. A follower control algorithm and a robotic vehicle capable of driving the leader's path autonomously.

The robotic leader/follower application creates a number of challenging requirements for a visual tracking system. Firstly, it is highly desirable that the system has the ability to be trained on a leader target at run-time, so that any vehicle can be used as a leader. Secondly, varying vehicle speeds and convoy configurations require that the system function over a wide range of distances. Thirdly, the motion of both the follower and leader vehicles over rough terrain requires fast dynamic response of the camera pan/tilt control. Finally, the vision system must be robust enough to always maintain the leader in the field of view, or the follower robot will become lost.

1.2 Vision-Based Following

This work has been inspired by the way a human driver would follow a lead vehicle. Using the leader's colour and textural features, a human continually locates the leader in his/her field of view. Using the size and relationship of the target's features, the human estimates the distance to the leader, remembering the path that it took. Finally, using the neck and eyes, the human can follow the leader's current trajectory while driving the path that the leader took some time before, following at an arbitrary distance behind. This approach is shown graphically in Figure 1.

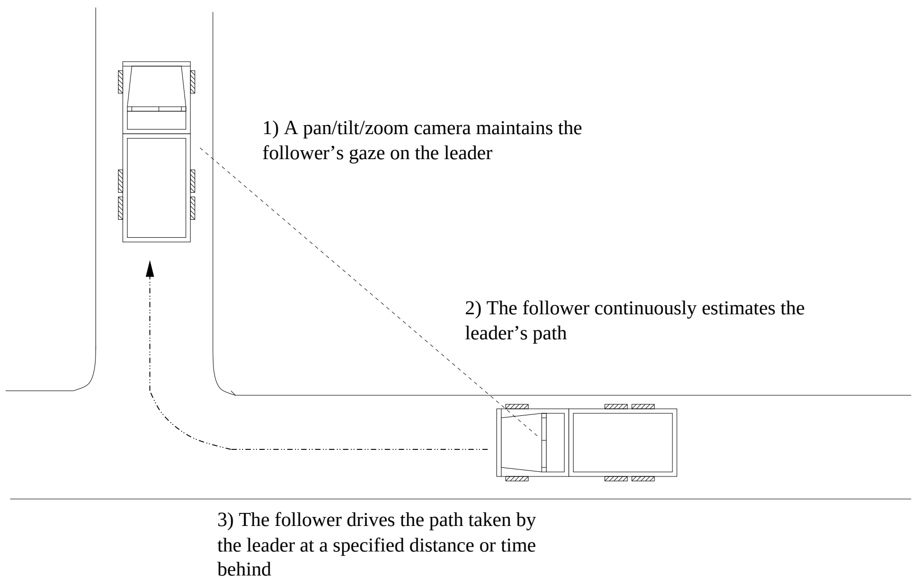


Figure 1: *The vision-based approach to robotic following.*

In order to accomplish the leader/follower task, the use of a visual tracking system confers a number of distinct advantages over other technologies, such as laser range finders, GPS, sonar, radar, etc.:

- No hardware or software is required on the leader.
- The hardware on the follower vehicle is relatively inexpensive.
- In a military context, no active sensing or radio communications are required which could alert the enemy to the convoy's presence.
- It makes it easier to choose a leader at run-time rather than having it pre-programmed.

However, there are a number of challenges imposed by choosing a vision-based approach:

- Obscurations such as mud, dust and intervening obstacles can cause the follower to completely lose the leader (Figure 2).
- Vibrations from vehicle motion can blur video images, resulting in erroneous data.

- Computational delay in visual systems can make controlling a camera to follow a moving target difficult.

(a) Original Image (b) Poor Lighting (c) Dust (d) Turning

Figure 2: *Some of the difficulties for an autonomous convoying vision system.*

1.3 Hardware

1.3.1 Vehicle

The test platform for the pan/tilt/zoom tracking system is the Multi-Agent Tactical Sentry (MATS) vehicle [4], shown in Figure 3. It was developed at Defence R&D Canada – Suffield, and is currently in use with the Canadian Forces (CF). It is a tele-operated UGV, meaning that a user controls it from a remote control station using a joystick, video feed, and a map display of the vehicle's position.

Figure 3: *The Multi-Agent Tactical Sentry (MATS) robotic vehicle.*

This vehicle is being used in this project as a test platform, but it could potentially function as a soldier mule robot. It is also intended that the tracking system will later be transferred to a small personal robot, as well as a large logistics truck for autonomous convoying.

1.3.2 Camera

The camera system used is the DI-5000 Camera from ICX Technologies [5], shown in Figure 4. It has a 25 times zoom lens (2.4 to 60mm), resulting in a horizontal field of view of 45 degrees to 2 degrees. The video is output in NTSC format, resulting in an image resolution of 640 x 480 pixels. It also includes an infrared camera, which may be useful for future work in this area.

Figure 4: *The DI-5000 pan/tilt/zoom camera.*

The tracking system is by no means intended to be specific to the camera used for these tests. It is hoped that the robustness of the control system will enable it to be used on any pan/tilt/zoom camera, with the adjustment of a few parameters.

1.3.3 Software and Computing

The follower vehicle architecture is shown in Figure 5. The general process is as follows:

1. Image processing algorithms locate the leader in the image stream, estimating its range and bearing in world coordinates.

2. The camera control algorithm sends RS-232 commands to adjust the pan/tilt/zoom of the camera to maintain the leader in its field of view.

3. A vehicle control algorithm smoothes the vision range and bearing, generating vehicle speed and steering commands to follow the leader's path. [1]

4. The desired speed and velocity commands are sent via RS-232 to the MATS vehicle. The vehicle's Ancaeus control system[2] uses the vehicle actuators to execute the desired commands.

[1] This algorithm was developed by researchers at the University of Toronto and tested at Defence R&D Canada – Suffield.

[2] The Ancaeus architecture was developed at Defence R&D Canada – Suffield, and has been used to tele-operate a wide variety of robotic vehicles.

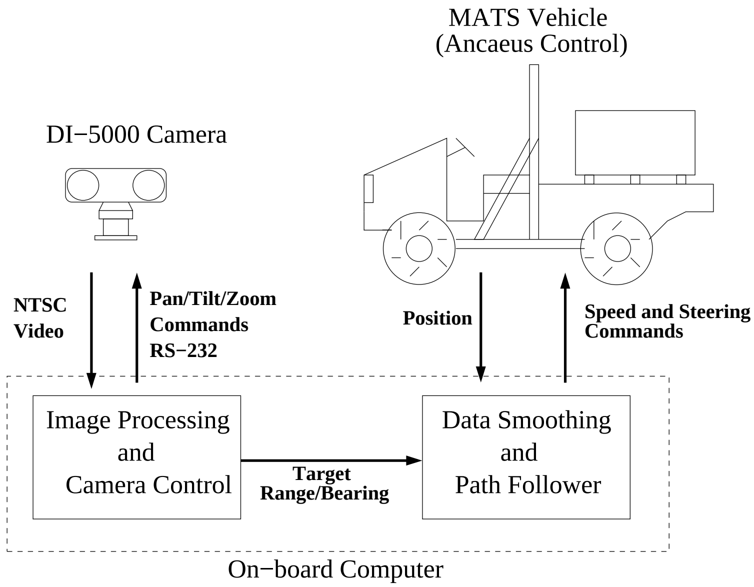


Figure 5: *The leader/follower control architecture.*

The complete camera tracking system was reported as a Master's Thesis published through the University of Calgary [6], to which readers are referred for a more complete description.

Processing power for the vision, estimation and control software is provided by a Dual Xeon 3.6GHz computer, running Fedora Core 3 Linux. Video from the camera is captured by an Osprey 440 frame grabber at 30Hz, and digitized at 640x480 resolution. A number of software libraries were used to speed up the development process:

- Trolltech Qt [7] - A library for developing Graphical User Interfaces and multi-threaded programs.
- Intel OpenCV [8] - A computer vision library for a wide variety of tasks, such as displaying and converting images, etc.
- Evolution Robotics ViPR [9] - An image recognition library based upon the SIFT algorithm.

1.4 Objectives and Contributions

The primary goal of this research is to follow a leader vehicle or human using only visual information. In, particular, in developing the camera tracking system, three scientific contributions were produced:

1. A vision-based vehicle tracking camera system which is trainable at run-time.
2. A novel zoom control algorithm suitable for the pan/tilt/zoom tracking problem.

3. A demonstration of a complete pan/tilt/zoom tracking system, enabling leader path estimation from a moving follower vehicle.

4. A path tracker which follows the lead vehicle at a set following time, rather than a set following distance. This allows the follower to slow down for corners and obstacles as the leader does.

The remainder of this document is organized as follows: Section 2 provides a literature review of previous leader/follower systems and camera tracking algorithms. Section 3 outlines the computer vision algorithms employed in this work, including colour tracking and object recognition. Section 4 details the control scheme for the pan, tilt and zoom of the camera tracking system, while Section 6 provides results of actual leader/follower experiments conducted on the DRDC Suffield Experimental Proving Ground. Finally, conclusions and future work can be found in Sections 7 and 8.

2 Background

This section is a literature review of the many sub-components of this project: leader/follower control, computer vision, and camera control. Because the work is a combination of many fields of study which are themselves each quite involved, the review is quite broad. For coherence, it has been broken down into the following sections: Section 2.1 covers current systems for autonomous convoying and robotic leader/follower with special attention paid to the sensing methods employed; Section 2.2 reviews methods of recognizing and tracking objects using computer vision; finally, Section 2.3 reviews current pan/tilt/zoom tracking systems, and the control algorithms employed.

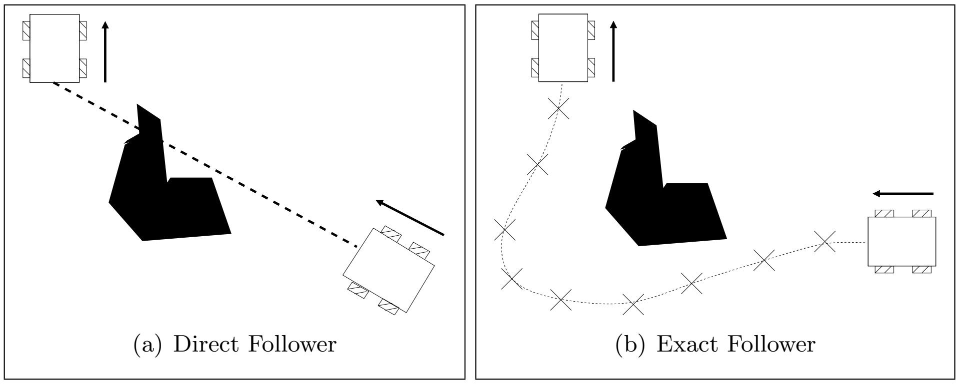
2.1 Robotic Leader/Follower

Military convoying is one potentially useful application of autonomous robotics, reducing the human involvement in transporting goods such as ammunition and water across potentially dangerous terrain. The ability of an autonomous vehicle to follow a leader is key to this task. The role of a leader/follower vehicle in military applications is discussed at length by the United States National Research council, in their book *Technology Development for Unmanned Ground Vehicles* [10]. In it, they describe three basic roles:

> Missions appropriate for follower UGV capabilities include (1) serving as a soldier's "mule" to carry weapons, ammunition, and other items cross-country behind dismounted soldiers; (2) operating as a logistics resupply vehicles to follow a leader vehicle in road-traversing convoy mode; and (3) accomplishing logistics resupply cross-country (including poor roads and paths) following a leader vehicle by an interval of minutes to hours.

In the past, a number of robotic follower systems have been designed within both military and non-military contexts to fulfill these roles. Within these roles, we can define two basic types of leader/follower systems [11]: (1) a perceptive follower which uses a sensing system to follow its leader, often in a relative coordinate system; and (2) a delayed follower which receives a path, usually in global coordinates, to follow the leader at some arbitrary time later.

We can also define direct and exact following (Figure 6). In direct following (Figure 6(a)), the robot pursues the current target position, without accumulating or matching the leader's path. This will cause the follower to cut corners, and may be unsuitable for environments cluttered with obstacles. In contrast, an exact follower (Figure 6(b)) will accumulate and attempt to match the leader's trajectory. This is more appropriate for most environments, but requires more complexity in the follower system. Perceptive followers can be either direct [12, 13, 14] or exact [15, 16], while delayed followers are always exact [1, 11, 17]. The goal of this project is to create a perceptive exact follower system, relatively rare in the literature.



(a) Direct Follower (b) Exact Follower

Figure 6: *Direct vs. exact following.*

2.1.1 Perceptive Following

Perceptive following is distinguished by immediate, sensor based following of the leader. A wide variety of sensors have been employed to accomplish the perceptive following task, each with their own advantages and disadvantages.

Laser range finders, such as the SICK or Velodyne scanners (Figure 7(a)), use the time of flight of laser beams to measure distance to objects. They send out multiple beams of light in a swath, and can provide a profile of the back of a leader vehicle. This method was used in [15] for automobiles and in [11] for an armoured military vehicle. Although accurate, lasers are expensive, range limited, power hungry, and vulnerable to dust, fog, etc.

Automotive back-up radar sensors, such as the Delphi Forewarn system [18] (Figure 7(b)) are much more robust to environmental conditions, but do not provide the fidelity of laser sensing, and have limited range. However, newer automotive adaptive cruise control radars operating in the 76 GHz, with a range of up to 200 meters were used for the United States TARDEC Robotic Follower ATD and CAST programs [19]. It should be noted that active sensors such as laser and radar have the potential down-side of alerting enemies to the vehicle's presence on the battlefield.

Stereo vision cameras, which are a passive sensing method, use the disparity between features in a pair of physically separated cameras to produce range measurements, much like the human eyes. One example is the commercially available Point Grey Bumblebee ([20], Figure 7(c)). These cameras can provide accurate measurements of the leader's position, and even its pose, but have a limited range given a reasonable baseline between cameras. Therefore they are normally used for leader/follower on smaller robots, such as by Kubinger[21].

Monocular vision, another passive method, uses a single camera to estimate the range and bearing to the leader vehicle. This has the advantages of being inexpensive and able to work over long ranges with the use of a zoom lens. Smith used this to create a large person-following robot for surveying in rough terrain [12]. Nguyen et. al. created a person following robot for carrying military supplies using monocular vision [22]. Juberts et. al.

used a square planar target for on-highway following [23]. Benhimane et. al. [24] used monocular vision for a on-road car-following system without the use of special targets. The RACCOON system was developed to follow another vehicles tail-lights at night [25]. It has also been applied to numerous smaller robots [16, 26, 27, 28, 29, 30, 31, 32]. The author of this report has also done a preliminary study on a small indoor robot [13]. Unfortunately, the data retrieved using monocular vision is often noisy if travelling at high speeds or over rough terrain, and can be prone to complete failure in finding the leader. This report will show that these difficulties can be overcome.

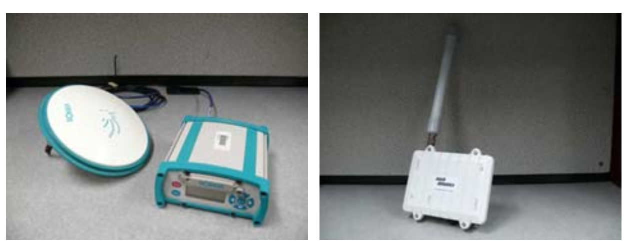
One interesting alternative to remote sensing is a tethered cable between the leader and follower vehicles. It measures the length of cable spooled out and the angle between the cable and the bumper to determine the range and bearing to a leader vehicle. To the author's knowledge, this has only been reported on the "Autonomous Solutions, Inc." website [33] and not in any scientific report.

(a) Laser Scanners

(b) Radar

(c) Stereo Vision

(d) Monocular Vision



(e) GPS and Radio

Figure 7: *Sensors used for autonomous leader/follower.*

Using these different sensing modalities, perceptive leader/follower systems have been developed for robots which fly [34, 35], swim [36, 37, 31, 38, 14, 39], or roll across the ground. Many of them have been developed to follow people using monocular vision, and fulfill the personal mule role [22, 30, 40].

More specific to this project, a number of ground robots have also used pan/tilt/zoom cameras to accomplish the follower task. This can be accomplished using either pre-defined markers on the leader [41, 42], or by using the appearance of the leader itself [24, 35, 43].

2.1.2 Delayed Following

The delayed follower task involves following a leader under non-line-of-sight conditions. The leader accumulates a path as a set of waypoints in a global coordinate system, and transmits it to the follower some arbitrary time later. This provides an exact trajectory for the follower vehicle but requires a method to transmit the trajectory to the follower, usually involving data radios. Furthermore, this requires GPS on both the leader and follower vehicles, whereas the perceptive systems described above only require equipment on the follower.

One such system, developed by the French company Thales Airborne Systems, was concerned with both perceptive and delayed convoying [11], and was demonstrated on-road and off-road using military vehicles. Some automotive highway systems can be found in [2, 3]. The author of this report has also done experiments at low speeds using this method [17].

The current gold standard are the convoying systems produced for the United States TARDEC Robotic Follower and Convoy Active Safety Technologies (CAST) [1] projects. The robotic convoy trucks, shown in Figure 8, use GPS waypoint following to achieve speeds of at least 65 km/hr. They also include a number of other sensing modalities including vision and radar.

Figure 8: *The US TARDEC Convoy Active Safety Technologies (CAST) vehicles [1].*

One intriguing notion is to combine the idea of a perceptive and delayed follower. The TARDEC systems [1] uses laser scanners to pre-record a 3-D terrain map of leader's route. This map is then passed to the follower, which attempts to adjust its position to match the 3-D features its own laser senses. Although this allows for GPS free operation, the downside to this approach is the complexity of the equipment required on both the leader and follower.

Another intriguing possibility is to pre-record a video of the path the leader has taken, and transmit it to a delayed follower vehicle [44]. This vehicle uses the sequence of images to control its own steering and velocity to match the leader's path. This was demonstrated on-road on a small electric car.

The literature also contains a number of works examining the problem of controlling an entire convoy of vehicles, using a variety of control methodologies [11, 45, 15, 32, 29, 46].

2.2 Monocular Visual Recognition and Tracking

This section investigates currently available technology for locating a target and estimating its 3-D position in a stream of images from a single camera. There are 3 main challenges to the monocular tracking process: (1) identify the correct target, (2) correctly locate its position and orientation in the image, and (3) estimate its 3-D position or pose. There exists a huge body of work in this field, which can be grouped into 3 basic approaches:

1. Fiducial Based - A special marker is placed on the target which uniquely identifies it and provides information about its pose relative to the camera. This is the most accurate method, but may fail if the target is even partially occluded.

2. Model Based - The tracker builds up a model of the 3-D structure of the object to be tracked, and uses this to estimate its current pose. This method can handle partial occlusions, but may be affected by camera calibration or model inaccuracies.

3. Appearance Based - The tracker uses the visual features of the object to follow it and estimate its pose, using sensing modalities such as colour and texture. This method is the most flexible, but the robustness and accuracy depend on the tracker's ability to detect and track distinctive image features, which may not always be possible.

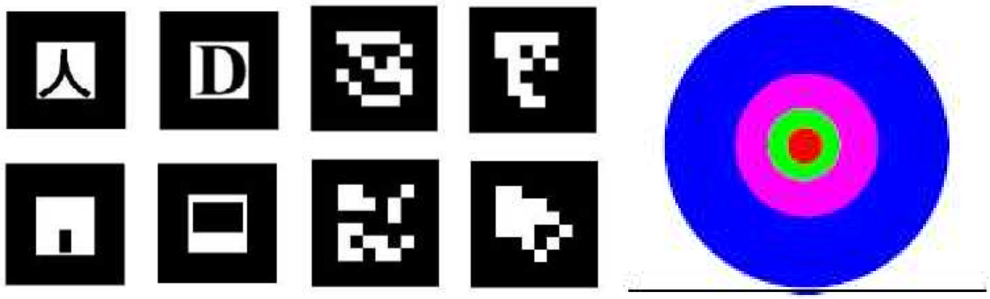
Within these groups, trackers can be classified as either image independent (operates on only one image) or recursive (requires a sequence of images and previous knowledge to track).

A good summary of visual tracking techniques can be found in [47], which goes well beyond the scope of this report.

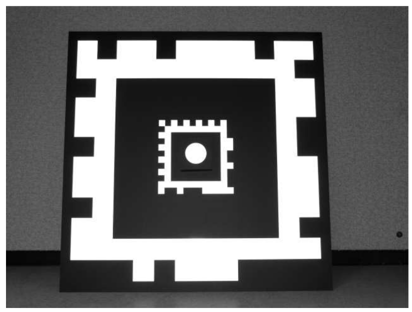
2.2.1 Fiducial Based

One pragmatic approach to solving the monocular-vision tracking problem is to place a special target marker, or fiducial, on the object. These distinctively shaped and coloured markers make it easier not only to find the target with simple computer vision algorithms, but also to obtain its position, or even its full pose (depending on the marker used). This is done by comparing the known geometry of the fiducial with its perceived geometry. Examples include the ARToolkit [48](Figure 9(a)), ARTag[49] (Figure 9(b)), concentric rings (Figure 9(c),[50]) and the Space Vision Marker System [51](9(d)).

Fiducials have been used on a number of small robots for the robot following task [32, 42, 52, 35, 53, 26, 16]. In a more fully developed use of a fiducial target, the Toyota company used



(a) ARToolkit (b) ARTag (c) Multi-Ring



(d) SVMS

Figure 9: *Fiducials used for augmented reality and space flight applications.*

infrared LED lights to communicate information and determine distance for on-highway platooning [54]. For the robot following task, Smith [12] used the sequence of white dots shown in Figure 10, on a coloured planar board to produce full pose information.

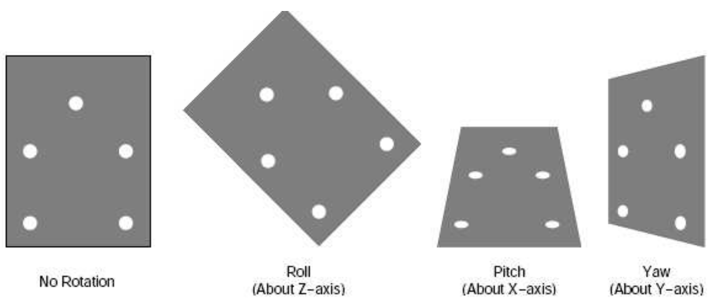


Figure 10: *A sophisticated fiducial system and how it can be used to obtain pose information (taken from [12]).*

Unfortunately, for any system using fiducials, the recognition system must be tied closely to the specific fiducial used, and cannot easily be adapted to another target. Also, the recognition system will be susceptible to image noise and partial occlusion of the target. For these reasons other tracking methods are pursued in this project.

2.2.2 Model Based

When tracking, it is often beneficial to use knowledge of the 3-D structure of the target, both to help recognize it, and also to estimate its pose. This model may be a CAD model, a set of planar parts, or a more generic model such as a rectangle, etc.

Many model based approaches detect the edges of the object to track it [55, 56, 57, 58]. Other model based approaches use visual features such as corners to create a model, and overlay them on a skeleton or mesh frame [59, 60] (Figure 11(a)).Some have applied model based tracking to vehicles [61, 62, 63, 64] (Figure 11(b)). Unfortunately, the model based methods are complex and seem to lack the robustness of other sensing means, and may become confused by background clutter, even though they provide good pose information about the target.

(a) Face Model (b) Vehicle Model

Figure 11: *Two examples of model based tracking (taken from [61] and [59]).*

2.2.3 Appearance Based

A more direct way of tracking an object is by recognizing it directly using its appearance, such as its colour, texture, etc. This is generally more robust but may make it more difficult to extract pose information. Furthermore, depending on the appearance features used, this can be computationally intensive. We can generally separate the appearance-based methods into those that use (1) colour, (2) image templates, or (3) small feature points on the object.

2.2.3.1 Colour

Identifying an object by colour is simple, intuitive, and robust to partial occlusion and scale changes. For these reasons, a number of works have used colour for leader/follower robots [26, 16, 36, 65]. Additionally, the Robocup tournament has driven countless implementations [66, 67, 68], and some of this software is available in open source, such as CMUVision [69], or as commercial software, such as Cognachrome [70]. Other colour "blobfinding" leader/follower robots can be found in [71, 72, 29, 73, 31, 22, 30, 40].

These algorithms can be computationally efficient. However, as with all colour based recognition, they are fragile to changes in illumination. If an object of similar colour enters the field of view, it may be mistaken for the desired target. These methods do not work well on patterned, textured objects, where no colour dominates.

Figure 12: An example of a spherical target and the resultant blob (taken from [66]).

To track multi-coloured objects, rather than using just one colour to track the target, some implementations extract statistical information of the target, and then compare it with those extracted from different sub-regions of subsequent images to detect a match [40, 71, 74, 75, 76, 77, 78]. These methods are better than tracking a single colour, but again will fail if the target is too close in colour to the background or another object in the image.

Adaptive recognition using machine learning has also been used to try and make colour recognition practical, such as the CAMSHIFT algorithm [79] and many others [22, 68, 80, 30, 73, 66, 76]. From preliminary experiments at DRDC, the downside to adaptive methods is that over long term tracking, if the parameters are not tuned correctly, the system may learn away from the intended target on to a different target. Furthermore, trackers like these which work recursively on a sequence of images may be prone to catastrophic failure if the target becomes obscured.

2.2.3.2 Templates and Correspondence

Recognizing a generic object directly using some sort of trained template of its appearance is a more flexible way to accomplish target recognition. These methods use pattern recognition techniques, rather than being based on colour, and can be trained to recognize a variety of objects. However, with flexibility comes a marked increase in computation time. This has been an active area of research for many years [81, 82, 83]. Benhimane et. al. achieved vehicle platooning with an electric car by minimizing the sum of squared difference between an image template of the leader and the current image[24]. Templates are good because they are simple, but not robust to occlusions. Robustness to lighting and scale changes requires a number of pre-trained templates.

2.2.3.3 Feature Points

Most objects contain small features, or interest points, which can be used to track them. If you record the relationships of a number of these features to each other, you can use them to recognize an object and determine its pose. These methods are easier to make robust to illumination changes and partial occlusion, but tend to be more computationally complex.

Examples include the Harris detector [84], Lucas-Kanade Tracker [85], and the Scale Invariant Feature Transform (SIFT) [86]. The SIFT algorithm in particular is extremely robust to lighting variations, partial occlusion, scale and orientation. Numerous papers have indicated its superiority to other features [87, 88, 89]. Also, it has been shown to be effective for robotic object tracking [90, 91, 13]. An image and the extracted SIFT keypoints are shown in Figure 13. Drawbacks include long computation time and a requirement for high resolution, blur-free images.

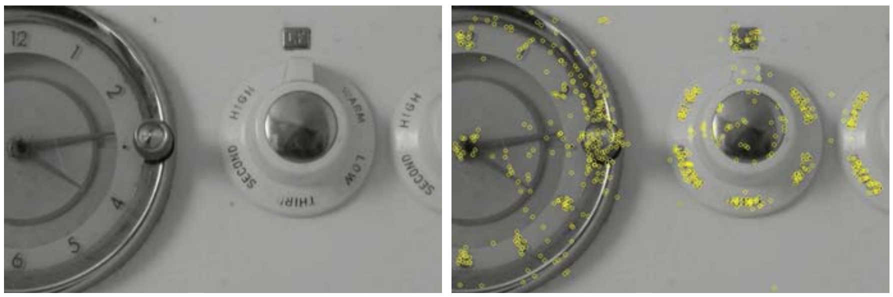


Figure 13: *A sample image and the recognized keypoints found using the SIFT algorithm.*

2.2.4 Other Methods

Many camera tracking systems used for moving objects use the motion of the target itself to identify it (optical flow). This is common in surveillance applications for tracking people or vehicles. It has the benefit of providing easy and fast target recognition without any model of the subject, but is limited by being unable to distinguish the target object from any other moving object [92, 93, 28, 94, 95, 96]. This has been used on mobile robots to track and follow people [97, 40].

Optical flow methods have a number of drawbacks which preclude their use in UGV applications. One is that the system must be able to account for the vehicle and camera motions before it can find the target. These methods also require a motion within the image for detection. A vehicle driving at a matching speed directly in front of the robot will not be found by an optical flow tracker. Furthermore, most optical flow methods depend on brightness constancy to operate.

Another visual feature that can be used to track an object is a contour, which is a deformable curve, or snake which represents an outline of a target. Examples can be found in [73, 98, 99] (Figure 14) These methods are also good for tracking the silhouette of an object, which may be useful under poor lighting conditions [100].

2.2.4.1 Combining Features

Of the methods of monocular tracking given above, none of them are cure-all solutions. Each has strengths and weaknesses. Fiducials are the fastest, most robust and most accurate, but require placing objects on the target. Edges, contours and 3-D models are complex and can fail with cluttered backgrounds. Also, they are not easy to adapt to new targets. Colour

Figure 14: *Person detection using edge detection and contours (taken from [73])*

.

tracking is simple and fast, robust to occlusions and scaling, but fragile under changing lighting conditions and provides poor localization information. Template matching and feature point methods require textured objects, and fail under fast motion or low resolution. Image templates are also susceptible to partial occlusions. Feature point methods are robust and good for appearance based geometric localization, but can be computationally intensive and are not suited for plain objects.

For these reasons, many researchers have attempted to use multiple visual cues to create better visual tracking systems. Schlegel uses colour histograms and contours [73]. Lee combines SIFT features with 3-D lines [90]. Bellato uses colour vision and laser range finding [101]. Guo uses colour regions, corner features, and lines to detect vehicles [102], while Xiong uses colour and line features for the same application [103].

2.3 Pan/Tilt/Zoom Control

This section will review the approaches in the literature for controlling a camera's pan,tilt and zoom motions from visual data. Some of these works use monocular vision, some use stereo, but the difficulties are the same: noisy and delayed visual data is used to control the camera's motion in real-time, aiming for the utmost responsiveness while ensuring the system remains stable and on-target.

There are three critical aspects to the control of visual systems which make it more difficult than a standard control system design. The first is that if the controller is not good enough to keep the tracking error within a certain bound (the field of view of the camera), then the system fails entirely. This requirement means that the dynamic performance of the controller is important.

The second critical aspect is the delay inherent in the visual processing feedback loop. This delay pushes closed loop controllers towards the stability limit, making the desired dynamic performance more difficult to attain. Contrary to most control systems, improving sensing with the addition of another or a better vision algorithm can sometimes serve to make the control system worse, because it extends the processing time and creates more negative effects on stability. The system delay and the trade-offs between sensing and processing time must be carefully managed. An exhaustive analysis of delay issues can be found in [104].

A third critical aspect is the intertwining of control and sensing fidelity. A controller which is overly reactive may blur images, hindering some image processing algorithms. Furthermore, when controlling zoom cameras, the target must be kept large enough in the image to get good image processing, without being zoomed in so far that any small motion of the target takes it out of the field of view.

Most of the earlier works in this field focused on using vision to control manipulator robots (typically referred to as visual servoing). Examples include [105, 106]. A series of works by Corke [107, 108, 109] provide much guidance for the control of robots using visual data. Malis provides a survey of this field in [110].

2.3.1 Pan and Tilt Control

There are also a number of controllers designed specifically for pan/tilt cameras. One of the most thorough works is a series of papers by Wavering and Fiala, using the TRICLOPS camera [58, 111, 112]. Another well analyzed set of works is by Barreto et. al. [113, 114, 115].

There are a number of other papers available on visual tracking, but none seem as rigorous or informative as the previous three sets of works by Corke, Wavering et. al, and Barreto et. al.. Wu et. al. [65] use a simple Proportional, Integral, Derivate (PID) tracker, while Oh et. al. [116] and Daniilidis et. al. [117] used Linear Quadratic Gaussian (LQR) controllers. Papanikolopoulos et. al. [105] used both a PI and a LQG controller. Naeem et. al. apply both LQG control [118] and model predictive control [39] to track cables for an underwater robot. Hong et. al. [53] use a two stage method. In the initial training step, the target dynamics are learned while tracking is accomplished using a PI regulator. In the second stage these dynamics are used with a more elegant controller. This seems fragile if the target dynamics are not stable.

Saccadic tracking, a biologically inspired method in which the control is done in a two-step process is used in [96, 28]. Like a human eye, the controller does a fast step motion to the center of the target when it is in the edge of the scene (saccade), and a slow careful motion to track when it is in the center. However, when applied to visual servoing, these methods are somewhat unsophisticated and underanalyzed. It would seem that the same could be accomplished using a properly tuned standard controller.

2.3.2 Zoom Control

Control of the pan/tilt angles can be modelled as a regulation problem, driving the angles to the target in the image to zero. However, the control of zoom (focal length) for a lens is not so simple. The complicating issue is the choice of setpoint for the proper zoom level (Figure 15). A long focal length (close in zoom) provides a good view of the target for the computer vision algorithm, but makes it difficult to maintain it in the field of view. Conversely, zooming out makes it easier for the control system to reject disturbances, but makes it more difficult for the vision system to recognize the target and estimate its distance.

In most works, zoom control is entirely decoupled from the pan/tilt control, or not used at all.

(a) Short Focal Length: The target is easy to keep in field of view, but difficult to recognize and estimate its distance.

(b) Long Focal Length: the target is easy to locate and recognize, but difficult to maintain in the camera field of view.

Figure 15: *The fundamental task in zoom control: adjust the focal length to attain the highest possible resolution, while ensuring the target will not leave the field of view.*

Some researchers have addressed this problem by having cameras at two different focal lengths [43]. This means having one camera with a wider field of view (and lower resolution) to ensure contact with the target, and a second camera with a narrower field of view for high resolution image processing. This is referred to as foveal vision, and has a strong analogue in human sight.

The key issue to zoom tracking is deciding on a metric to set the proper zoom level. One simple method is to pick an appropriate size of target in the image, and then use that size as a static reference setpoint for the focal length controller [119, 120, 121, 122].

Using a single image-size setpoint does not address the fundamental goal: to zoom as close to the target as possible while ensuring the target will stay in the field of view. To do this properly requires a dynamic measure of how much the target moves in the image and a measure of the controller's ability to compensate for its motion. The system is able to zoom in much further on a target which is not subject to fast, random motions than for one that is. Some examples of controllers which do this include [123, 124, 125].

3 Image-Based Leader Recognition

This section describes the algorithms used to develop a visual tracking system for a robotic leader/follower system using pan/tilt/zoom. This system identifies a target object in the sequence of images from a colour video camera. It also reports the number of pixels to the center of the target from the center of the image, and an estimated distance to the target. For this task a visual tracking system has a number of specific requirements. Firstly, it is desireable that it can be trainable at run-time, so that a user can pick the leader vehicle it wishes to follow. Also it needs to be effective over a range of distances, or object scales, so that the changing space between the leader and follower will not cause the vision system to break down. Thirdly, it should be as tolerant as possible to noise from motion blur, dust, etc., so that leader vehicle can still be identified while both the leader and follower are in motion. Accuracy in terms of distance measurements is crucial, as the follower robot will be planning its path based upon the data from the vision system. Finally, it is important to have a fast update rate with a minimum of delay, as both can directly affect the dynamic performance of the control system for the pan/tilt/zoom mechanisms of the camera.

With these requirements in mind, this work presents a system with two visual cues:

1. A colour tracker working in the HSV image space.

2. An object recognition tracker using the Scale Invariant Feature Transform (SIFT).

It has been shown that multiple visual cues can be an effective solution to tracking real world objects [126]. The two visual cues chosen were selected due to their individual characteristics and complementary nature. The colour tracker is computationally faster, and more immune to image noise than the SIFT tracker, but does not work well for multi-colored or textured objects, and is somewhat vulnerable to lighting changes. The SIFT tracker is more immune to lighting changes, and can accurately determine distance, even if the object is partially obscured, but is computationally slower, and vulnerable to image noise and motion blur.

In addition to the complementary nature of the two algorithms chosen, military vehicles have two useful characteristics which justify this algorithm selection: (1) they are homogeneous in colour which is beneficial for the HSV tracker, and (2) they are highly textured in appearance, which is beneficial for the SIFT tracker. This is also true of many non-military vehicles as well. Typical military armored vehicles are shown in Figure 16.

The remainder of this section will go into detail on the Graphical User Interface developed for the colour tracking system, as well as technical descriptions and results for the two algorithms employed.

3.1 User Interface and Training

The first step in the visual tracking process is training the system on the target object. For this purpose, a special Graphical User Interface (GUI) has been developed which interfaces to the pan/tilt/zoom mechanisms of the DI-5000 camera, and also to the image capture

(a) Leopard Tank

(b) LAV III

(c) RG-31 Nyala

(d) HLVW

Figure 16: *The homogeneous colours and intricate textures typical of Canadian Forces vehicles.*

hardware on the PC to display the view from the camera. The basic layout is shown in Figure 17.

The left side of the GUI contains all the basic controls for the camera, such as pan, tilt, zoom, focus, etc. On the right side, there are two buttons indicated as “Train SIFT” and “Train Colour”. When the user wishes to train the visual tracker, he/she presses one of the training buttons, which pops up a view of the camera’s current video feed. The user then uses the camera controls to center the image on the leader vehicle, and zoom to an appropriate size. When satisfied, the user draws a box around the vehicle to be followed. Once the system has grabbed the portion of the image the user selected, it prompts the user for the distance to the object, which it will use in future estimates of the target’s range. The training window is shown on the right side of Figure 17. Once the system has been trained for either/both the colour tracker and the SIFT tracker, the user clicks the “Autonomous Control” button to enable the system to begin panning, tilting and zooming to keep the target centered in its view.

Once trained, each of the colour and SIFT trackers are capable of reporting a number of data items for each image, shown visually in Figure 18:

Phi (ϕ) The horizontal angle in radians to the center of the target, from the center axis of the camera.

Psi (ψ) The vertical angle in radians to the center of the target, from the center axis of

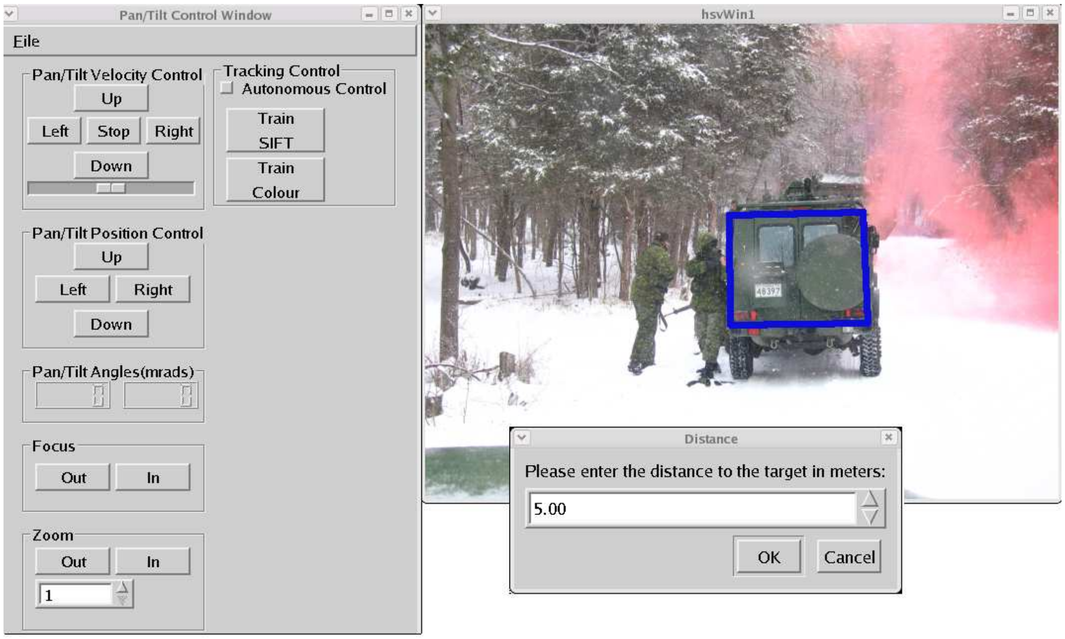


Figure 17: *The GUI used to train the visual tracker.*

the camera.

Range (r) The estimated distance in meters to the target.

Details of how each algorithm recognizes the target and determines its properties during and after training will be given in the next two sections.

3.2 Colour Tracking

The colour tracker used is similar to the one presented in [127], but was developed independantly by the author based upon standard computer vision practices using the OpenCV software library.

The first step to initiate tracking is training the system by using the “Train Colour” button in the GUI shown in Figure 17. It prompts the user to draw a box around a portion of the displayed camera video containing only the colour the user wishes to track. During this stage, the user can pan, tilt and zoom the camera as necessary to provide the best view of the target for training. Once selected, the tracking software takes the pixels within the user-drawn box, and creates histograms for hue, saturation and value (Figure 19). The highest peaks in the histograms are then used to determine appropriate colour to be used by the tracking system. Parameters used to determine how selective the software is in determining matching colours can be adjusted by the user at run-time to emphasize one target characteristic (H,S or V) over another. Also at training time, the target width and height are determined from the image using the distance as measured by the user.

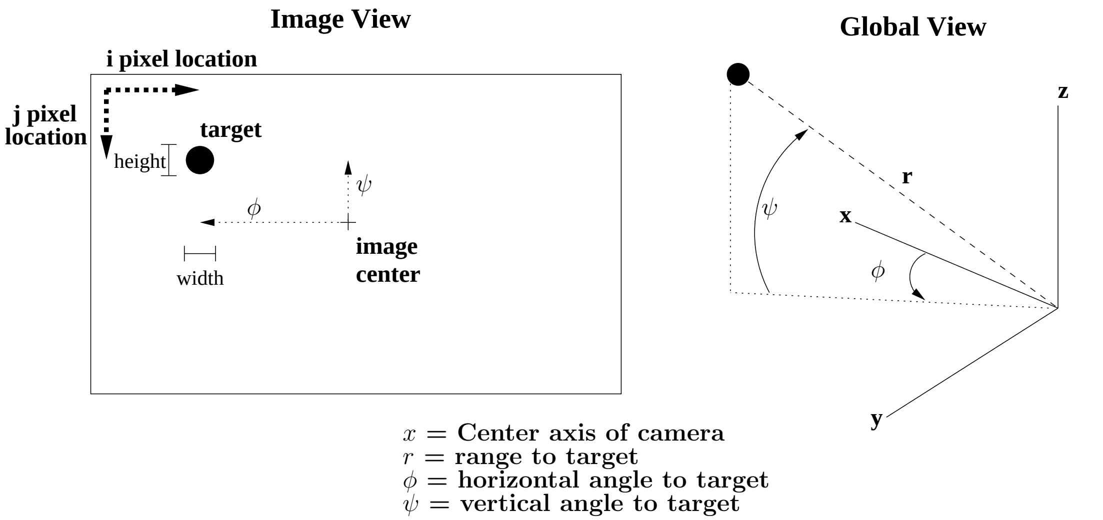


Figure 18: *The coordinate system used by the visual trackers.*

(a) Training Image (b) Hue (c) Saturation (d) Value

Figure 19: *HSV histograms for training the colour system.*

After the training stage, the colour vision algorithm will be used to find the target in each subsequent 640x480 image received by the camera. The algorithm is summarized in Figure 20. A sequence of illustrating images is shown in Figure 21, and will be referred to during this section. The implementation of this algorithm was done in C++ using the OpenCV image library [8]. The goal is to match pixels in each image to the leader's colour in the HSV colour space (shown in Figure 22). The largest matching area in the image is considered our target, and its bounding box is used to estimate the position and range to the object being tracked.

Algorithm for colour tracking:

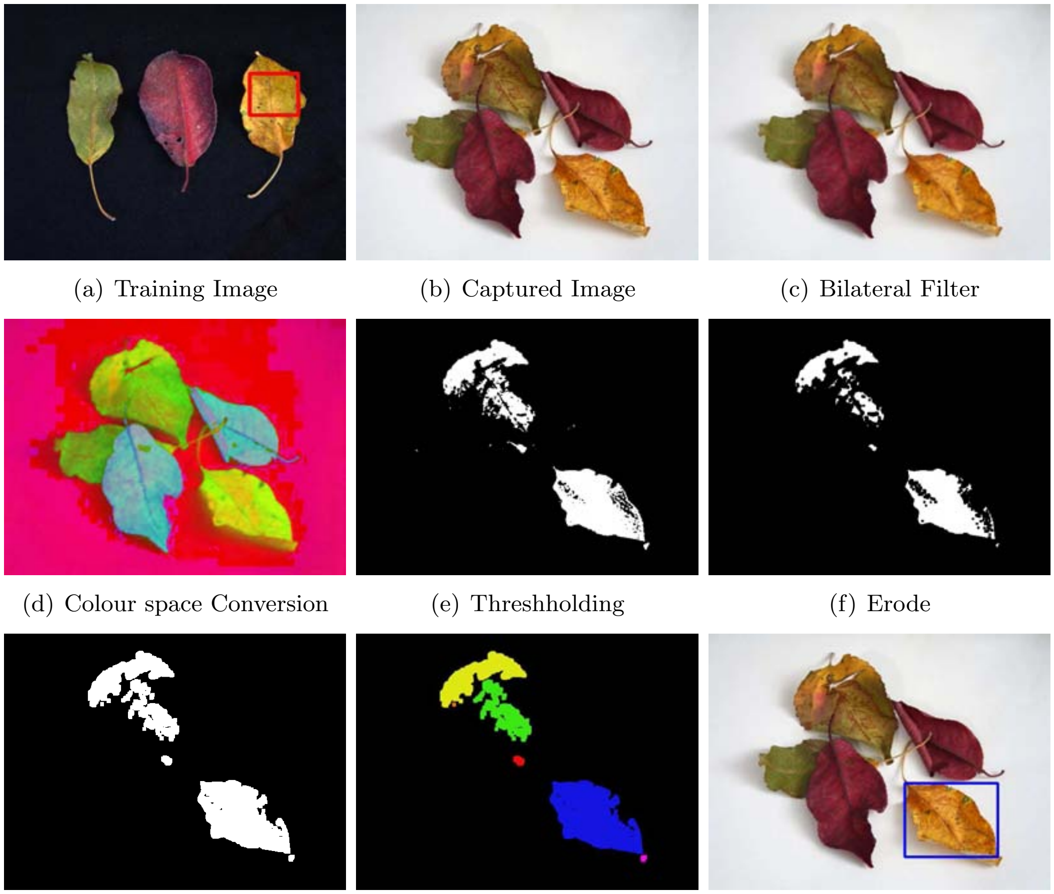
1. Capture an NTSC video image using the on-board framegrabber (Figure 21(b)).
2. Smooth the image using a bi-lateral filter to remove noise and sharpen edges (Figure 21(c)).
3. Convert the image to the HSV colour space to improve immunity to lighting changes (Figure 21(d)).
4. Create a binary target image by picking through all the pixels in the image, selecting those which are within the bounds for each of the hue, saturation and value threshholds (Figure 21(e)).
5. Erode and dilate the image to remove noise and reconnect segmented areas (Figure 21(f) and Figure 21(g)).
6. Flood fill the remaining areas to identify the size of target areas (Figure 21(h)).
7. Of the areas connected by the flood fill, select the region with the largest number of pixels, retrieving its characteristics to report (Figure 21(i)).

Figure 20: *Algorithm for simple colour tracking*

There is one obvious flaw to this tracking algorithm: If a larger secondary object of the same colour comes into the field of view the colour tracker will select the larger target incorrectly. However, the assumption is that the zoom tracking algorithm will keep the field of view small enough to prevent this. Furthermore, it is assumed that the SIFT object recognition algorithm (described in Section 3.3), will correctly identify the original target under these circumstances. This, in concert with a properly tuned filter, should keep the camera trained on the correct object.

The method of determining target distance is based upon the pinhole camera model [128]. It uses the current focal length of the camera to convert image dimensions to real world dimensions, as shown in Figure 23. The pinhole model states that the ratio of a target size (h) to its image size (h'), is the ratio of its distance away (d) to the focal length of the camera (f):

$$\frac{h}{h'} = \frac{d}{f} \tag{1}$$



(a) Training Image (b) Captured Image (c) Bilateral Filter

(d) Colour space Conversion (e) Threshholding (f) Erode

(g) Dilate (h) Flood Fill (i) Final Target

Figure 21: *The simple colour tracking algorithm.*

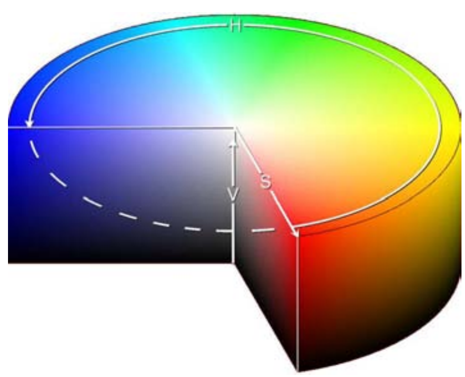


Figure 22: *The HSV colour space used for tracking.*

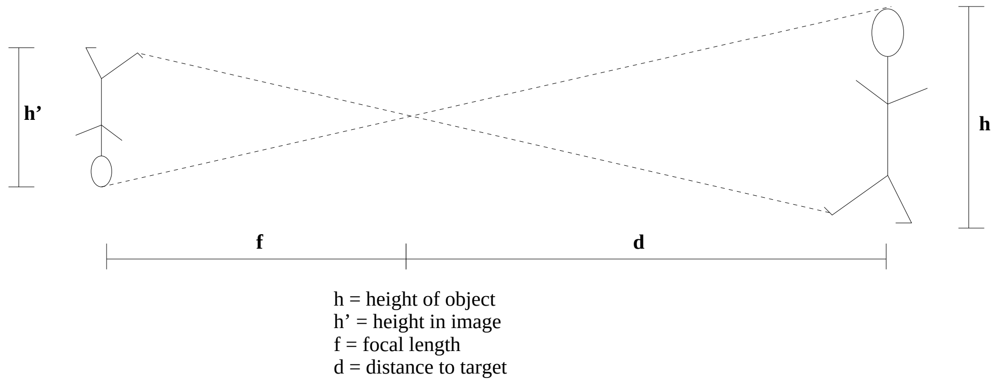


Figure 23: *The pinhole model used to determine the distance to a target for the colour tracker.*

Put another way, if a target is twice as far away, it will appear half as large in the image for a given focal length. For distance estimation, if its true height and the focal length of our camera are known, the size of its image is measured to get the distance. The equation holds equally true for the width of an object, so the distance can be found two different ways: by a distance due to height (d_h) and a distance due to width (d_w):

$$d_h = \frac{h \cdot f}{h'} \tag{2}$$

$$d_w = \frac{w \cdot f}{w'} \tag{3}$$

Note that $d_h = d_w = d$ if all measurements are accurate.

Tracking the target using the pan/tilt motions of the camera also requires the horizontal and vertical angles to the center point target, ϕ and ψ, in radians. Referring to Figure 24, it can be shown that they are found by counting the number of horizontal and vertical pixels to the center of the target (i,j), multiplying by a magnification factor (M_i,M_j), and using the following equations:

$$\phi = \tan^{-1}\left(\frac{i \cdot M_i}{f}\right) \tag{4}$$

$$\psi = \tan^{-1}\left(\frac{j \cdot M_j}{f}\right) \tag{5}$$

There is a major flaw in this method of distance estimation. If a 2-D target object is not parallel to the image plane in either pitch or yaw, the corresponding height or width becomes smaller due to foreshortening, and the values of d_h and d_w respectively will increase. Therefore, the final estimated distance to the target is chosen from the lesser of the two target distances, so that the algorithm will still report an accurate distance if only pitch or yaw occurs. If both occur simultaneously, the distance to the target will be overestimated.

When tracking a 3-D object it can actually appear larger when it yaws (i.e. when a truck turns a corner, you can see the side as well as the back). For this reason, when tracking a

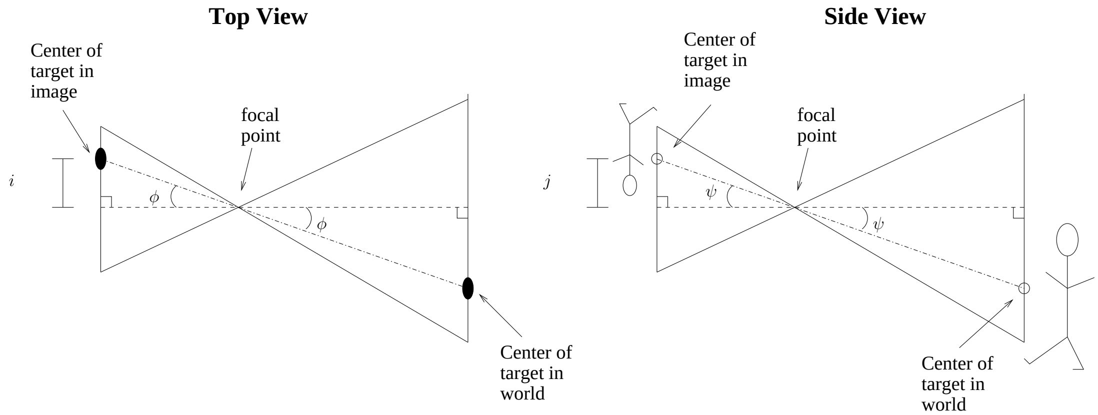


Figure 24: *Determining the angles to the target.*

leader vehicle, only the height of the target is used to maintain consistency around corners. This assumes small leader vehicle pitch relative to the follower vehicle.

In the above description of the colour tracking system two key points were left out: 1) the selection of appropriate threshholds for the H, S and V values, and 2) the determination of the height and width of the targets. Both of these are determined during the leader training stage using the GUI shown in Figure 17.

If the colour tracker follows only a set of specific targets, it is possible to pre-define colour threshholds and target size. However, it is beneficial to train when running the program to be able to pick a target at run-time. Also, the colour characteristics of the target under the current lighting conditions will be used, which makes for more reliable tracking.

3.3 SIFT Tracking

The Scale Invariant Feature Tracking (SIFT) algorithm for object recognition is a different type of tracker from the colour tracker just described. It relies on finding a large number of small, distinct feature points on an object, based on intensity rather than colour. The position relationships of these small features are used not only recognize an object, but also to determine its distance and orientation. The software used in this project relied upon the object recognition libraries included with the Evolution Robotics ERSP toolkit[9]. The library extracts feature points from a training image and compares these feature points to those extracted from successive camera images. For a planar leader object, only one training image is necessary. For 3-D objects, training images from different views makes the algorithm more reliable in tracking the leader. The algorithm detects unique features in an image of an object by analyzing the texture of a small window of pixels. Up to 1,000 feature points are extracted from an image, each consisting of the feature's location and a texture description. A small portion of these features, filtered for uniqueness make up a model database for that object.

This method has a number of desirable characteristics for real world applications. It is un-

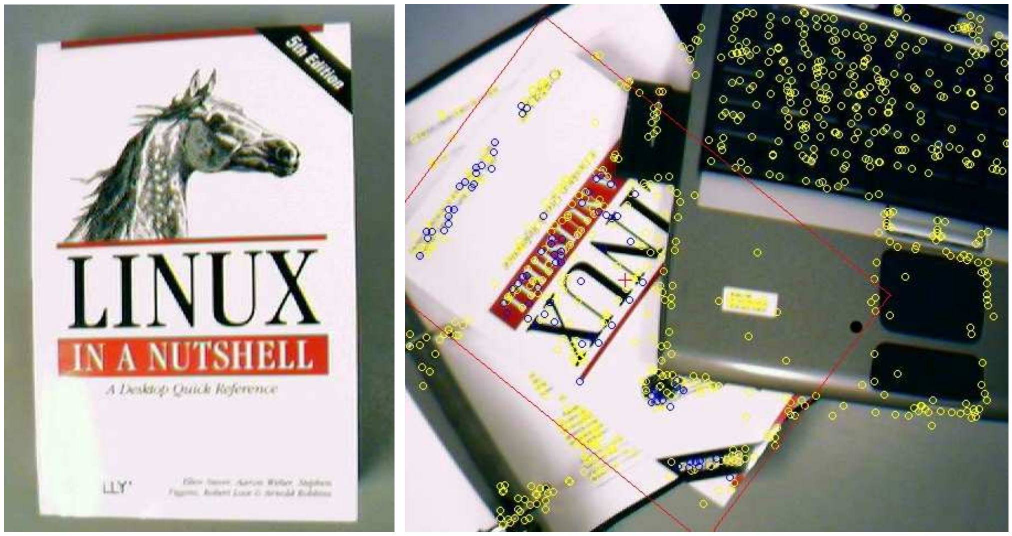


Figure 25: *A training image and the recognized image in a cluttered scene, with the red box indicating the position of the recognized object. Feature points are shown as yellow circles.*

affected by moderate changes in scale, rotation and translation. It also has some immunity to changes in lighting, and can be used on low cost, low resolution cameras. Finally, the algorithm will typically recognize objects with 50% to 90% occlusion [13]. It specializes in planar, textured objects, but also works well with 3-D objects having slightly curved components. A model image and the subsequent recognized image are shown in Figure 25.

This tracking method is trained by the same GUI as the colour tracker, shown in Figure 17. The user presses the "Train SIFT" button, at which time he can pan/tilt/zoom as necessary to draw a box around the target object. The software then crops the target out of the image, and passes that portion to the SIFT libraries as the training image. In subsequent images, the libraries provide a distance to the object, a bounding box which surrounds it in the image, and a target center. The other properties required by the rest of the system (ϕ, ψ) are calculated using the pinhole model in the same way as for the colour algorithm.

4 Camera Pan/Tilt/Zoom Control

The work in this section aims to design an appropriate controller for the combined vision and camera system to maintain the camera centered on the target at an appropriate zoom level. This is known as fixation in biological terms. The first section will examine control for the pan and tilt degrees of freedom. The second section will examine the development of an appropriate control structure for obtaining optimum focal length of the zoom lens. Reader's are referred to [6] for a more complete description of the control design.

4.1 Pan/Tilt Control

The DI-5000 camera is controlled by serial RS-232 velocity commands from the computer. The controller PC obtains images from the NTSC stream using a framegrabber, and analyzes them to obtain the relative location of the target from the center of the camera's field of view. For this particular system, the dominant feature for design consideration is the delay introduced by the PC framegrabber and image processing. There is also a delay associated with the serial command input to the camera, even though the motors themselves are fairly responsive. The goal will be to analyze the effects of this delay and to design a control structure capable of dealing with it.

A picture of the control loop for the pan degree of freedom is shown in Figure 26. It consists of the following components: a pan/tilt camera system (R(z)), a controller (D(z)), and a visual feedback system (V(z)). The motion of the target object is modelled as a disturbance (W(z)). The goal of the control design will be to reject the disturbance such that the relative position between the center of the image and the center of the target (Y(z)) remains at zero.

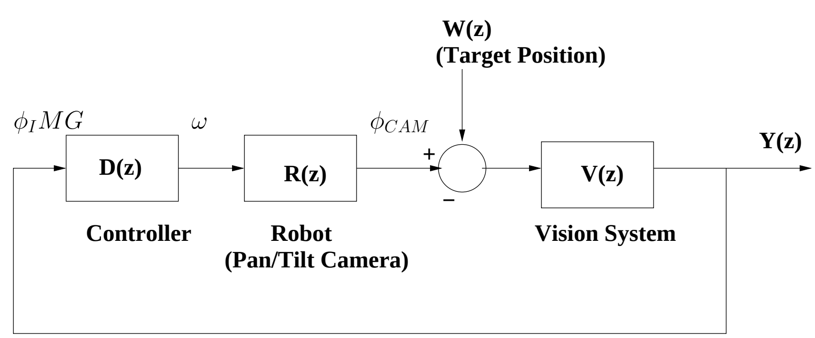


Figure 26: *Control loop for camera system. W(z) is the unknown target position, and Y(z) is the target's position in the image, which is regulated to zero (the center of the image). ω is the pan or tilt velocity sent to the camera.*

Because the target system is completely unknown and the goal is to regulate the angles to zero, there is no need to estimate the target position in real world coordinates for control purposes. Rather, the control system is based on image-based servoing using image coordinates. This is a regulation problem in control system terminology, in which the goal is to drive the angles to the target in the image to zero (ϕ,ψ). Target motion is considered a perturbation, and performance is evaluated by rejection of that perturbation [109]. For this control problem, the pan and tilt degrees of freedom are completely decoupled, treating the regulation problems for ϕ and ψ as independent control loops [113].

The general control design approach is as follows, and is detailed in the sections below:

1. Find a reasonable dynamic model of the pan/tilt system using Least-Squares Parameter Estimation.
2. Design an LQG controller using MATLAB tools and simulations.
3. Implement the controllers on the camera and test for performance.

In order to design an appropriate controller for the camera system, an accurate system model is required. In this case, because the internal components and control software within the camera were proprietary and unknown, recursive least-squares (RLS) parameter estimation was used to determine a linear dynamic model. This algorithm can be found in many sources, including [129, 130]. The algorithm for RLS parameter estimation was implemented in Matlab and applied to the DI-5000 camera to estimate a transfer function for the pan and tilt mechanisms.

Once a suitable dynamic model of the pan/tilt camera was obtained, control design and simulations using MATLAB were used to analyze the effect of delay and create an effective controller. Linear Quadratic control is an "optimal" method based on a state-space model of the plant to be controlled [131, 132]. It provides an optimized control signal derived from the internal states of the camera and a user-defined quadratic cost function for performance. It attempts to drive all the internal states of the plant to zero (Linear Quadratic Regulation). If an estimate of these internal states is provided by a Kalman Filter, then the algorithm takes the name Linear Quadratic Gaussian (LQG) control. LQG provides the designer with a method to control the importance of control gains and system response, while being somewhat immune to sensor and plant model noise. However, performance is somewhat dependant on the accuracy of the plant model.

The LQG controller works with a fixed feedback gain K, and a Kalman Filter gain L. The internal states of the system are estimated as:

$$\hat{x}[n] = A\hat{x}[n-1] + Bu[n-1] + L(y_v[n] - C\hat{x}[n-1]) \tag{6}$$

where $\hat{x}[n]$ is the estimated state vector at time n, $u[n]$ is the control signal, and $y_v[n]$ is the noisy sensor measurement (i.e. the ϕ and ψ measurements obtained from the computer vision algorithms). A, B, and C constitute the dynamic state-space model, as obtained from the least squares parameter estimation. From there, a control signal is generated by:

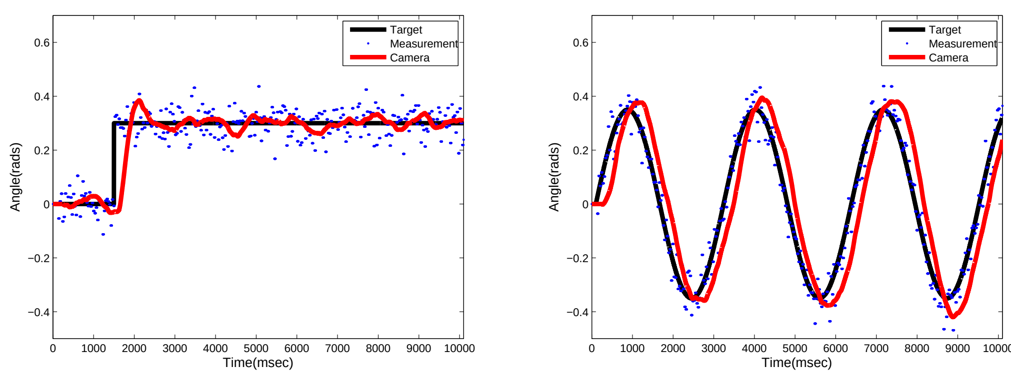
$$u[n] = -K\hat{x}[n] \tag{7}$$

This control structure is used separately for each of the pan and tilt degrees of freedom of the camera (i.e. there is a separate state vector x, Kalman Filter, and controller for each of ϕ and ψ). For the pan degree of freedom, the estimated state $\hat{x}[n]$ would contain the horizontal position and velocity of the target (ϕ,$\dot{\phi}$) while the control signal $u[n]$ would be the pan velocity sent to the camera. For the tilt of the camera, the estimated state $\hat{x}[n]$ would contain the vertical position and velocity of the target (ψ,$\dot{\psi}$) while the control signal $u[n]$ would be the tilt velocity sent to the camera.

One benefit of LQG control is that it is possible to explicitly include delay states in the discrete state-space model to approximate delays in the real camera system. The estimator and LQR controller are then designed for the expanded state system. The phase lag can be almost entirely eliminated using this approach (as compared with standard Proportional-Integral control). Details can be found in [6].

A simulation of the LQG control algorithm with moderate sensor measurement noise is shown in Figure 27. The performance depends on the specific tuning of the control and filter gains. In general, it was found that the LQG design was relatively easy to tune for good performance without becoming unstable.

In a scenario with sensor noise, the designer needs to balance the control and filter gains between eliminating phase lag, preventing overshoot and eliminating noise. If the filter is made slower such that its noise rejection is good, the step and sine responses will be slow. If the control system is tuned to be "fast" so that it eliminates the phase lag, it will be susceptible to noise.



(a) Step Response

(b) Sine Response

Figure 27: *Simulated response of the camera LQG controller (with sensor noise).*

4.2 Zoom Control

Zoom control is used in this system to choose the appropriate focal length to maintain a constant target image size regardless of its distance from the camera. This is important not only for reliable object recognition, but also for accurate distance measurements. Unfortunately, one particular difficulty with visual control loops is that if the target moves out of the field of view the system can't sense it at all and complete failure ensues. Any zoom control algorithm must keep the field of view as small as possible on the target while ensuring that the field of view is large enough that the target is not lost under noisy conditions. Previous approaches to zoom control were outlined in Section 2.3.2. Most algorithms choose an ideal target size in the image, and use the error from this ideal to servo the zoom mechanism [120, 121]. Some other methods use the noise measure from a Kalman Filter to decide the camera zoom level. However, this leaves little room for error when the tracking becomes noisy, and will not adapt to changing tracking conditions. Other algorithms zoom out when the target nears the edge of the field of view and the risk of losing tracking is increased [119].

To ensure that the target never nears the edges of the field of view, a three part zoom controller is presented [3]:

ρ - Proportional control to maintain ideal target image size if no tracking error is present.

τ - A quick reacting component which zooms the camera out instantly in response to a quick motion of the target towards the edge of the image. This gain is based solely on the current horizontal and vertical tracking errors.

σ - A smoothing component which zooms the camera out in response to longer periods of target displacement from the image center. This computes a moving average of the horizontal and vertical tracking errors over a specified number of previous images (moving average window).

A mathematical description of the zoom controller is given, using the image coordinates and variables[4].

$$f_k = f_{k-1} + \rho(\xi - \sqrt{h'_{pixels} \cdot w'_{pixels}}) + \tau(|\phi_k| + |\psi_k|) + \sigma \cdot a_k \tag{8}$$

$$\text{where} \qquad a_k = \frac{1}{n} \sum_{i=k-n+1}^{k} (|\phi_i| + |\psi_i|) \tag{9}$$

[3] This algorithm is entirely of the author's design, and is not based on any previous work in the literature.

[4] The image variables and coordinates are described more fully in Section 3.

where the variables are:

$f_k =$ camera focal length command at time step k
$\rho =$ gain for proportional target size control
$\xi =$ ideal target image size in pixels (usually chosen from size of training image)
$h'_{pixels} =$ height of target in image (pixels)
$w'_{pixels} =$ width of target in image (pixels)
$\sigma =$ gain for the average tracking error over a window of time
$a_k =$ a moving average of the tracking errors
$\tau =$ gain for current tracking error
$\phi =$ horizontal tracking error (radians)
$\psi =$ vertical tracking error (radians)
$n =$ number of previous images to average the tracking error (moving average window)

Although this mathematical description seems complex, the concept of the algorithm is quite simple: zoom the camera to the ideal target size under ideal conditions, and zoom out under less than ideal conditions. Equation 8 is now broken down into its three parts for better explanation. The first component zooms the camera to the ideal target size:

$$f_k = f_{k-1} + \rho(\xi - \sqrt{h'_{pixels} \cdot w'_{pixels}})$$

This is accomplished by changing the focal length (f_k) in reaction to the difference between an ideal target size (ξ) and the current measured target size ($\sqrt{h'_{pixels} \cdot w'_{pixels}}$). The square root is used to ensure that this control term doesn't grow to the squared power as the target size increases, but rather linearly. The user can tune the proportional gain ρ for desired response (normally a positive value).

The second term of the algorithm zooms the camera out in response to the tracking error, which is the target distance from the center of the image:

$$f_k = f_{k-1} + \tau(|\phi_k| + |\psi_k|)$$

By setting the τ tuning parameter negative, the focal length will be decrease (zoom out) based on the current Manhattan distance from the center of the image to the target ($|\phi_k| + |\psi_k|$).

The third component uses a moving average to keep the camera zoomed out during sustained periods of tracking error. This could also be viewed as a smoothing component:

$$f_k = f_{k-1} + \sigma \cdot a_k$$

$$\text{where} \qquad a_k = \frac{1}{n} \sum_{i=k-n+1}^{k} (|\phi_i| + |\psi_i|)$$

This term uses the sum of the tracking errors ($|\phi_i| + |\psi_i|$) over a pre-defined window of previous images (n). The gain σ, which is again normally negative, causes the focal length to decrease (zoom out) over periods of poor camera tracking.

Using these three terms and properly tuning ρ, σ and τ, will balance the algorithm's desire to zoom in to the ideal image size with the desire to zoom out due to tracking error. The downside to this approach is that these parameters must be tuned manually for good zoom performance in the face of erroneous sensor measurments. This proved relatively straightforward and intuitive for this camera system. Overall, despite being simple to implement and tune, and requiring no model, the controller proved remarkably effective. It was not overly sensitive to tuning parameters, and was stable over a wide range of tuning parameters.

5 Follower Vehicle Control

The goal of the vehicle controller is to use the leader's range and bearing from the vision system to force the vehicle to follow the leader's path. The controller was designed to follow not the leader's current position, but rather the path the leader has taken delayed by an arbitrary following time (typically 10 seconds or so). This means that the leader and follower would travel closer together when the leader was driving slowly (such as around corners), and further apart when travelling at higher speeds.

In order to accomplish this while taking advantage of the existing MATS architecture, a two level controller was designed. The outer loop, developed by researchers at the University of Toronto, smoothes the leader range and bearing values from the pan/tilt/zoom tracking system, and generates the speed and steering angle required to follow the leader's delayed path. The inner loop, which was originally developed for tele-operated control of the MATS vehicles, adjusts the steering wheel and gas pedal of the vehicle to generate the speeds and steering angles requested. This two-level architecture shown graphically in Figure 28. The key to using the two-stage control architecture is that the inner loop must operate much faster than the outer loop.

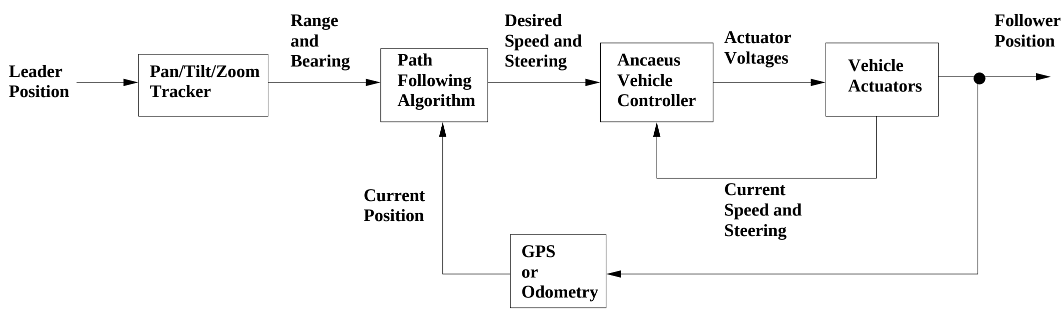


Figure 28: *Two-level vehicle control architecture.*

.

5.1 Path Following (Outer Loop) Control

The goal is to track the planar trajectory of the leader vehicle delayed by a constant time, τ. Specifically, if (x, y) is the position of the follower and (x_0, y_0) is the position of the leader, it attempts to make $(x(t), y(t))$ track $(x_0(t-\tau), y_0(t-\tau))$. For simplicity, the delayed leader position is defined as $(x_d(t), y_d(t))$.

The position of the leader vehicle, (x_0, y_0), is determined using the camera vision data relative to the current follower position, as described in Section 3. The current follower position (x, y) and heading (θ) are determined by using either the vehicle's on-board GPS, or by using vehicle dead-reckoning. For most UGV applications, dead-reckoning is not accurate enough to be practical over an extended period of time. However, in this instance, absolute positional accuracy is not necessary. The dead reckoning only needs to be accurate enough to navigate the vehicle over the distance between the leader and follower vehicles. Therefore, both GPS and dead-reckoning were successfully tested in this work.

The controller itself uses a bicycle model for a kinematic model of the vehicle, which is commonly used in vehicle following applications. It is given by:

$$\dot{x} = v_c \cos\theta$$
$$\dot{y} = v_c \sin\theta$$
$$\dot{\theta} = \frac{v_c}{d}\tan\gamma_c,$$

where (x, y) is the rear axle position, θ is the heading, d is the distance between the front and rear axles, v_c is the commanded speed, and γ_c is the commanded steering. One set of control laws that has worked well in practice is

$$v_c = v_d + k_{p,e_1} e_1, \quad k_{p,e_1} > 0$$
$$\gamma_c = k_{p,e_2} e_2 + k_{p,e_\theta} e_\theta, \quad k_{p,e_2}, k_{p,e_\theta} > 0,$$

where v_d is the speed of the delayed leader, e_1 is the longitudinal error, e_2 is the lateral error, and e_θ is the heading error. These are defined as follows:

$$e_1 = (x_d - x)\cos\theta + (y_d - y)\sin\theta$$
$$e_2 = -(x_d - x)\sin\theta + (y_d - y)\cos\theta$$
$$e_\theta = \theta_d - \theta.$$

These quantities are identified in Figure 29.

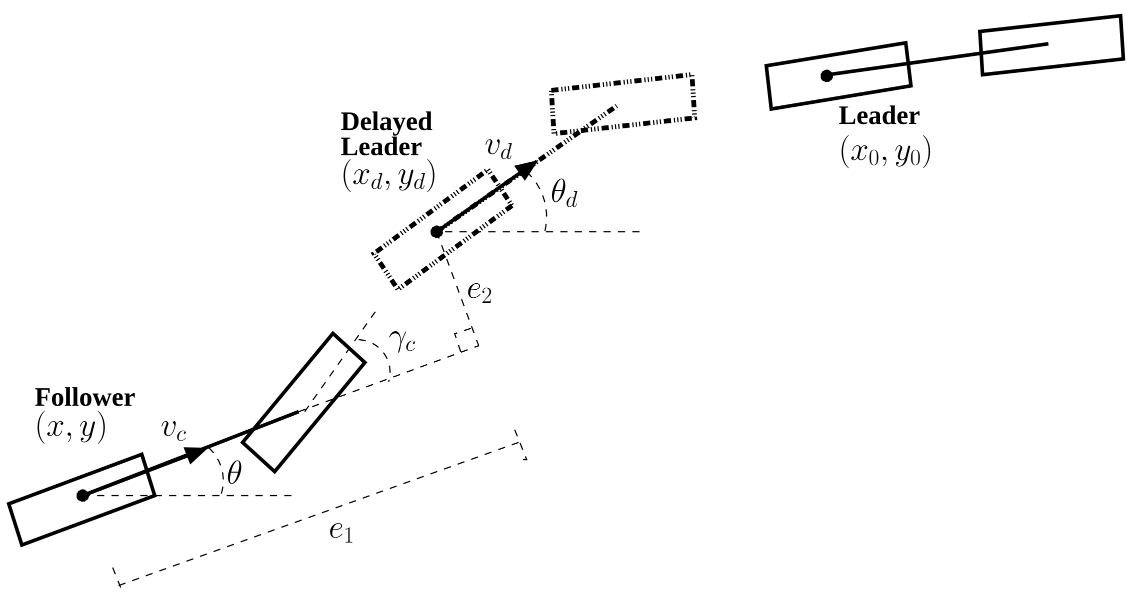


Figure 29: *Trajectory tracking of delayed leader*

After some initial field trials, it was observed that the follower was turning late, which caused a large initial lateral error. As such, a lookahead option was added. With a lookahead time defined, the controller picks a point ahead of the delayed leader's point to compute the lateral and heading errors that are used for feedback. The longitudinal error is not changed. This is similar to Gehrig and Stein's Control Using Trajectory (CUT) algorithm [133], except the lookahead point is based on a constant time not distance. The lookahead point is defined as

$$(x_l(t), y_l(t)) := (x_0(t - \tau + l), y_0(t - \tau + l)), \quad l < \tau,$$

where l is the constant lookahead time. Hence, the lateral and heading errors are now computed by

$$\begin{aligned} e_2 &= -(x_l - x)\sin\theta + (y_l - y)\cos\theta \\ e_\theta &= \theta_l - \theta. \end{aligned}$$

As input to the control law, the measurements of x_0 and y_0 came from the camera system as described earlier.

From our control laws, it is obvious that we need measurements or estimates of $x_d, y_d, \theta_d, x_l, y_l, \theta_l, v_d$. The details of how to obtain these quantities (from the follower's onboard sensor measurements) will be discussed in a future paper.

5.2 Ancaeus (Inner Loop) Control

The inner loop controller operates within the Ancaeus system on-board the MATS robotic vehicle. It was initially designed to allow a user at a remote control station to tele-operate the MATS vehicle, using a joystick or keyboard commands to control the speed and steering of the vehicle. The Ancaeus system is also capable of many other functions, such as operating the vehicle camera, horn, brakes, etc. and collecting data from all the on-board sensors for location, heading, fuel level, temperature, etc. The Ancaeus system consists of a vehicle communications protocol and a number of electronic hardware modules retrofitted onto vehicles. These modules include a Vehicle Control Processor Module (VCPM), Navigation Module, Audio/Visual Module and a Communications Module.

Under normal function, the human user interfaces to the remote ground station, which translates keyboard and joystick commands into the appropriate Ancaeus commands. These are sent over a wireless radio link to the vehicle, which puts them into action. In this case, for autonomous operation, the computer on-board the MATS vehicle sends Ancaeus commands to the vehicle via an RS-232 interface.[5]

The two most important commands sent by the path tracking algorithm are the *set speed*, and *set steering* commands (i.e. v_c and γ_c). The Ancaeus Vehicle Control Processor Module (VCPM) accepts these commands and manipulates actuators on the steering wheel and gas pedal of the vehicle to force the appropriate action. Proportional-Integral-Derivative (PID) loops take the input speed in km/h, or steering radius in centimeters, and convert them to

[5] Note that this is the same computer which was running the camera pan/tilt/zoom tracker and the path following algorithm.

voltages which drive the actuators themselves. At the same time, the Navigation Module on the MATS vehicle reports the current vehicle position and heading as obtained using either the on-board GPS, or using dead reckoning from vehicle odometry. This is used as feedback for the outer loop path following algorithm.

6 Results

This section will document test results using the overall integrated system in a leader/follower scenario. For more detailed results of the individual camera vision and control components, the reader is referred to [6].

6.1 Camera System Results

The camera recognition system described earlier is capable of locating an object to sub-degree angular accuracy and to within a few centimeters for range. However, it is important to verify that the vision algorithm developed is suited to the autonomous convoying task.

An outdoor test using real vehicles is presented where the MATS robotic vehicle was driven behind a commercial half-ton truck while recording video from the on-board camera (without pan/tilt tracking or zoom). Speeds were varied betwen 0 and 24 km/h, and following distance between 8 and 38 meters. The positions of both vehicles were recorded by a differential GPS system accurate to 2cm. Both the SIFT and colour trackers were trained at a distance of 15.5 meters.

Some representative images from the video are shown in Figure 30, at distances of 10 meters, 15.5 meters, 25 meters, and 37 meters. The left picture at each distance is the video image. The second image is the binary image of colour segmentation. The third image contains the target found by the colour tracker in the blue bounding box. The fourth image is the result of SIFT tracking. SIFT features are shown as yellow circles, and those features identified on the target are shown as blue features. The target location is given by the red bounding box. The colour tracker worked over the complete range of distances between 10 meters and 37 meters, but the SIFT tracker only worked up to a distance of 25 meters, as it requires a minimum amount of resolution in the target to find features.

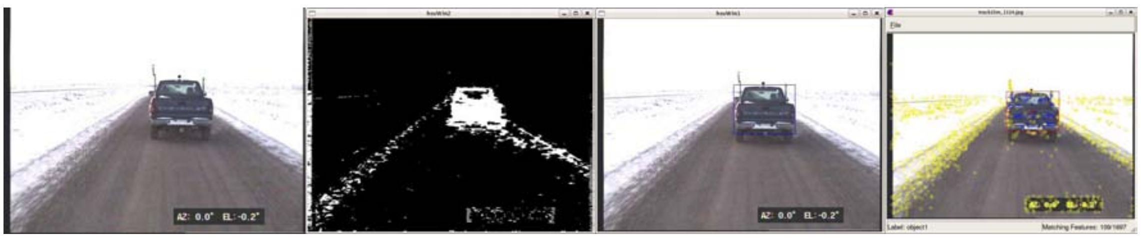
The distance between the leader and follower vehicle estimated by the colour tracker is shown in Figure 31[6]. The colour tracker was successful for the full range of distances, but it tended to overestimate the distance when the truck was further away. One promising result is that the colour tracker never failed to find the truck when it was in the image. Figure 32 shows one image of the truck taken during a turning operation, when it was barely visible. This is useful, as only part of the truck needs to be in the field of view for the system to generate a command for the pan/tilt camera.

The results from the SIFT tracker are shown in Figure 33. The SIFT algorithm did not find the truck in 523 out of 2501 images which contained at least part of it. This was due to either a lack of resolution of the target, or becauase of obscuring dust, etc. In Figure 33 it can be seen that SIFT fails completely at long distances, such as at image numbers 1000 and 2100. A typical image from these distances can be seen in Figure 30(d), where the SIFT algorithm could not find enough features on the truck to make a recognition.

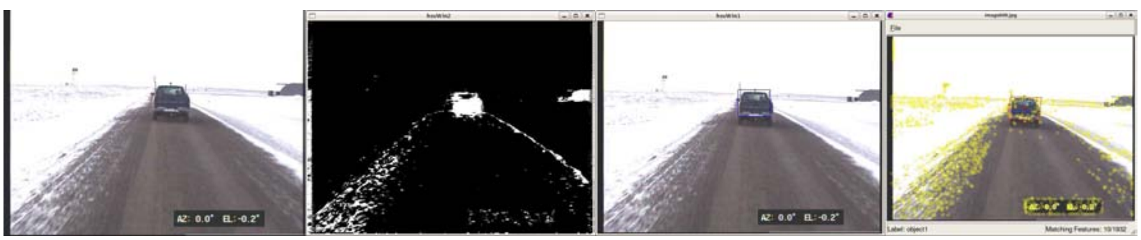
[6]It is important to note that these results did not use the camera zoom mechanism.

Video Image **Colour Segmentation** **Colour Target** **SIFT Target**

(a) 10 meters



(b) 15.5 meters



(c) 25 meters

(d) 37 meters

Figure 30: *The colour and SIFT trackers in operation.*

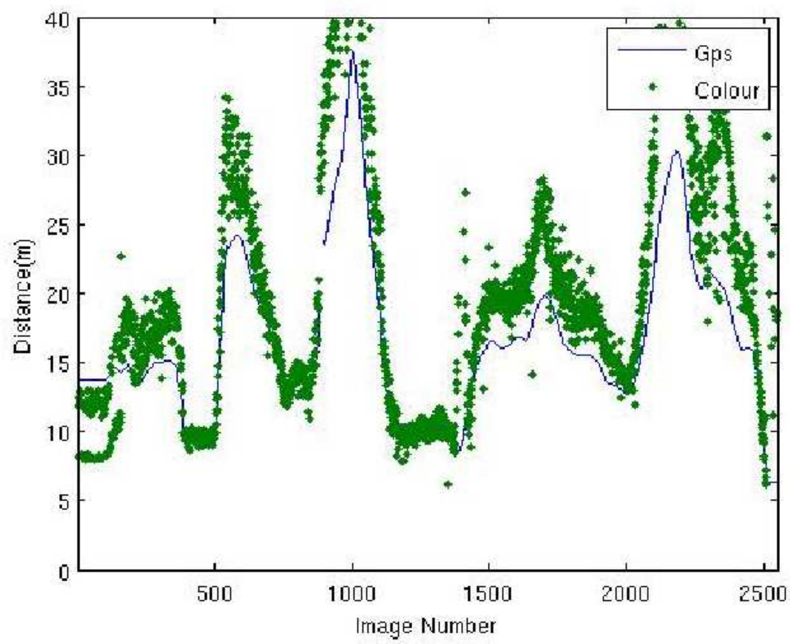


Figure 31: *Estimated and ground-truth distance for the leader/follower test using the colour tracker.*

Figure 32: *The colour tracker working under extreme circumstances.*

However, when the truck image is constrained to a reasonable size, the SIFT tracker finds the truck reliably, and the distance estimation is less noisy than for the colour tracker.

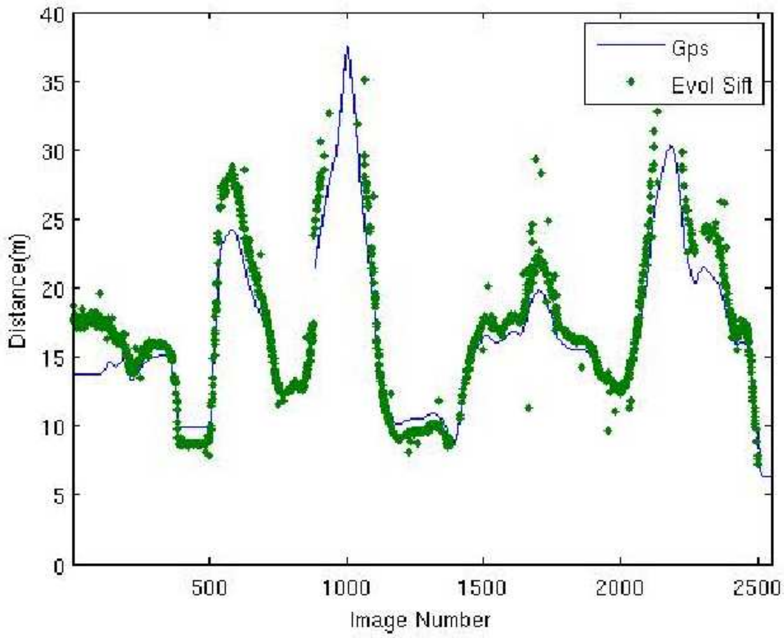


Figure 33: *The results of SIFT tracking in the truck video.*

From these results, it can be concluded that zoom is an important component for robust and accurate tracking. Both the SIFT and colour trackers were more consistent and accurate when the size of the target object was constrained. However, both have enough scale invariance that they won't be fragile.

Table 1 shows the relative strengths and weaknesses of the two image processing methods developed, as determined anecdotally during the course of testing.

Results of the pan/tilt/zoom response of the camera to moving objects can be found in [6].

Table 1: *The colour and SIFT based trackers compared*

Colour Tracker	SIFT Tracker
Computationally fast	Computationally slow
Easy to understand	Complex
Insensitive to image noise and motion blur	Susceptible to noise and motion blur
Degrades gracefully	Fails suddenly
Can remain focused on object despite changes to viewing angle and scale	Sensitive to viewing angle and somewhat to scale
Vulnerable to lighting changes	**Robust to lighting changes**
Inaccurate for distance if partially obscured or change in lighting	**Accurate for distance if partially obscured or change in lighting**
Inaccurate for distance if target yaws and tilts simultaneously	**Accurate for distance if target yaws and tilts simultaneously**
Can by confused by similarly coloured objects	**Robust to similarly coloured objects**
Poor for non-homogeneous or segmented targets	**Excellent for non-homogeneous or segmented targets**

6.2 Vehicle Following Results

This section presents results of leader/follower experiments that were undertaken on the DRDC Suffield Experimental Proving Ground in November of 2008. MATS vehicles were used in both the leader and follower roles, as shown in Figure 34. The leader vehicle was human-driven, while the follower vehicle operated autonomously. The differentially-corrected GPS systems on the MATS vehicle provided ground-truth data for the experiments.

(a) Straightaway (b) Turning

Figure 34: *Following a vehicle leader.*

All of the follower's subsystems were active during this test: the vision system continuously estimated the leader's position in the camera field of view, the pan/tilt controller kept the camera centered on the target, while the zoom algorithm maintained an appropriate focal length. The path following algorithm smoothed the vision data and generated steering and speed commands for the vehicle to maintain a fixed following time.

The leader and follower vehicles were driven around 1.3 km loop on a section of gravel road on the DRDC - Suffield Experimental Proving Ground. A visual plot of the leader's path is shown in Figure 35. The GPS ground-truth path taken by the leader vehicle is shown in blue, while the follower's path is indicated in red[7]. An enlarged view of a portion of the track is shown in Figure 36.

Figure 37 shows the ground truth distance between the two vehicles as well as the distance estimated by the camera system[8]. To analyze the accuracy of the path tracking system, two measurements of error are presented:

[7] This image was generated using Google Earth, which can create image offsets for some geographic locations. The actual path taken by the leader vehicle was actually more or less in the middle of the roads shown in the image, even though it doesn't appear that way.

[8] Data for bearing error (angular accuracy) is not included due to the inaccuracy of measuring vehicle heading during turns using GPS. However, given that the leader path estimated during the straight road sections is in the middle of the GPS track, it can be qualitatively stated that the bearing error is small relative to the range error.

Figure 35: *Leader and paths during the leader/follower test.*

.

Figure 36: *Leader and paths during the leader/follower test (enlarged).*

.

- Longitudinal Error - The distance along the path between the follower's current location and the desired location (i.e. the leader's delayed postion).
- Lateral Error - The distance perpendicular to the path between the follower's position and the delayed leader's position.

The results for these measurements are shown in Figure 38. For these experiments, the lack of an intertial heading sensor on the MATS vehicle meant that the following algorithm required GPS to determine heading. In examining Figure 38, spikes can be seen in the lateral and longitudial path tracking error, corresponding to the hairpin turn on each lap of the experiment. In addition to being a tight turn with lots of deceleration and acceleration, another effect is present. As the follower vehicle turns, the latency in the heading measurement from GPS caused large errors in the follower's estimate of the leader's position as the leader accelerated down the straightaway.

Tests were successfully conducted using wheel odometry only, but poor calibration resulted in an offset from the leader's path. Integration of an intertial heading sensor to free the vehicle from reliance on GPS is underway, and results are expected to outperform the GPS-based results.

Tracking moving objects outdoors from a moving platform is not an easy task. The changing lighting conditions, especially on a bright sunny day made the target data from the vision system less accurate, especially for the colour HSV tracker. However, while traversing this rough road with many turns the camera system did not lose tracking on the leader vehicle. Filtering the vision data was important, but the zoom algorithm which zooms

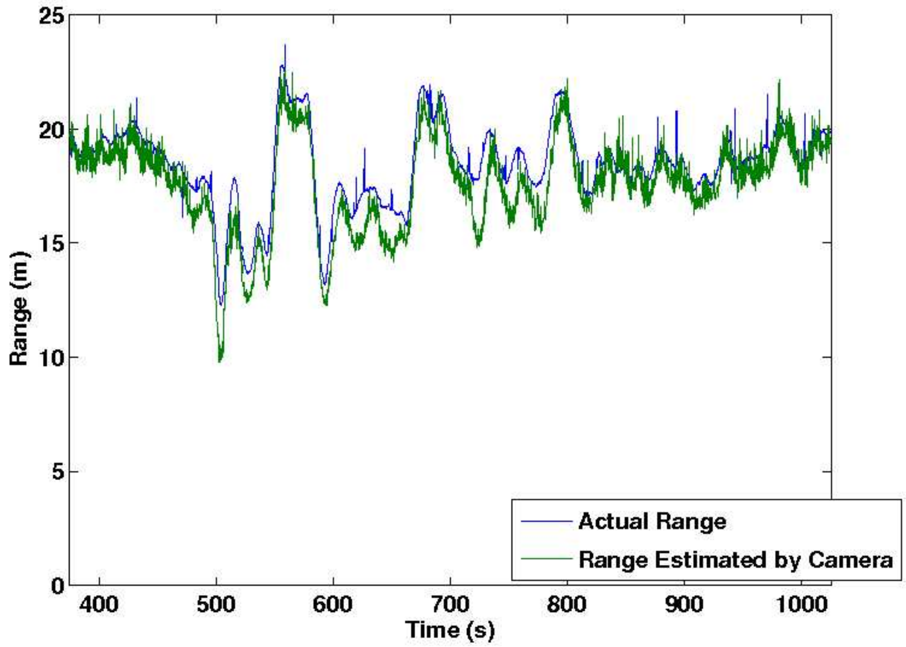


Figure 37: *Estimated and actual distance between the leader and follower vehicles for one loop during the leader/follower test.*

.

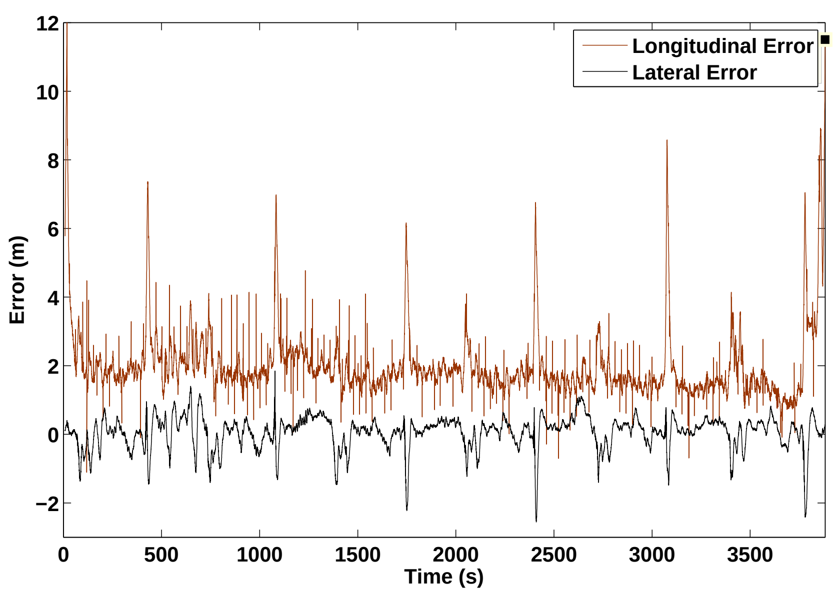


Figure 38: *Lateral and longitudinal path errors.*

out in response to poor tracking performance was essential in maintaining tracking during certain rough portions of the route. Unfortunately, this zooming out action means that the accuracy of the distance estimation decreases dramatically.

For this reason, the roughness of the road had two negative effects on the experiment: 1) limiting the top speed the follower could travel and still track the target, and 2) increasing the error in the position estimates (especially for range). The washboard sections of the road would induce camera vibrations, creating motion blur in the images. Also, bumps in the road would cause the vehicle to roll and pitch, making tracking difficult and forcing the camera to zoom out.

The most difficult portion of the route was the near-hairpin turn, at the west end of the route. A close-up of the leader and follower tracks for this portion are shown in Figure 39. At this point, the camera's view of the lead vehicle is extremely distorted from the straight behind view used for training, and the lighting on the vehicle changed dramatically. This can also be seen in the peaks for longitudinal and lateral tracking error in Figure 38, at time 490, 1050, etc.

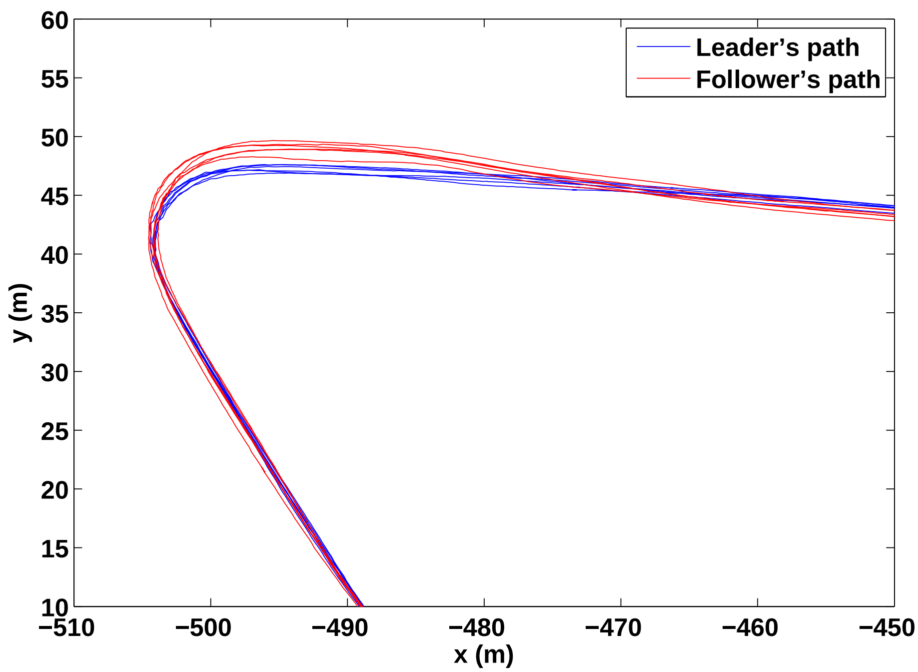


Figure 39: *Leader and follower paths during the leader/follower test (sharp turn).*

.

A statistical summary of the data from the leader/follower test is shown in Table 2.

Table 2: Statistical data for the complete leader/follower test.

Path Data

Loop Distance	1.3 km
Approx. Num. Loops	5.6
Total Distance	7.3 km
Mean Leader/Follower Separation	17.68 m
Max. Leader/Follower Separation	23.71 m
Min. Leader/Follower Separation	10.46 m
Mean Follower Speed	7.6 km/h
Max. Follower Speed	10.2 km/h

Visual Data

Mean Distance Measurement Error	0.72 m
Std. Dev. of Distance Measurement Error	0.62 m
Max. Distance Measurement Error	2.42 m

Follower Data

Mean Lateral Path Following Error	0.36 m
Std. Dev. of Lateral Path Following Error	0.48 m
Max. Lateral Path Following Error	2.56 m
Mean Longitudinal Path Following Error	2.02 m
Std. Dev. of Longitudinal Path Following Error	1.09 m
Max. Longitudinal Path Following Error	12.00 m

6.3 Person Following Results

Preliminary tests were also undertaken to assess the practicality of using this system as a "mule robot" to carry supplies for a dismounted soldier. The colour tracking system was trained on the human leader, and the vehicle followed his path, as shown in Figure 40(a). A view from the follower's camera is shown in Figure 40(b).

(a) Leader and Follower

(b) Camera View from Follower

Figure 40: *Following a human leader.*

The leader's path and the follower's path are shown in Figure 41, although no statistical analysis is presented (leader is in blue, follower is in red). One major difficulty is that a human is capable of many maneuvers that a vehicle is not, such as sharp u-turns, sideways motion, gap crossing, etc. Such a system would require the leader human to remain aware of the "mule's" limitations when moving. Safety is also be a major concern. A radar or laser obstacle detector and a remote kill switch could be installed to prevent the vehicle from harming its human leader or itself.

Figure 41: *Leader and follower paths using a human leader.*

7 Conclusions

This project successfully demonstrated a complete vision-based leader/follower system for an Unmanned Ground Vehicle. This is an inherently difficult problem due to the noise and processing delays incurred while using computer vision from a moving platform. In creating this system, the following original scientific contributions were produced:

1. A visual tracking system for which the user can choose the target at run-time.
2. A method of controlling the focal length of a camera for successfull pan/tilt/zoom tracking.
3. A demonstration of accurately estimating a leader's path from a moving follower vehicle using a commercial off-the-shelf camera.

The goals were accomplished through the successful integration of a number of component pieces. The computer vision system uses colour and SIFT based target tracking. It not only identifies and locates a target in an image stream but also estimates its 3-D position. A control system based on Linear Quadratic Gaussian control manages the delays prevalent in the visual processing loop to provide responsive tracking of the pan/tilt camera, while a separate zoom algorithm ensures that the leader never leaves the follower's field of view. Finally, a path following system enables the robotic vehicle to drive the leader's path with a pre-defined time delay.

There were a number of lessons learned in the accomplishment of this project. Firstly, using multiple visual cues allows the strengths of one to offset the weaknesses in the other. For example, colour tracking is not ideal for outdoor tracking due to problems with lighting changes and similarly coloured objects. However, having a vision algorithm with a fast update rate is essential, and it therefore provides a nice complement to the SIFT algorithm, even though it is not entirely effective on its own. Secondly, for visual tracking from a moving platform it is necessary to have a zoom algorithm which can regulate the size of the target image, enabling arbitrary distances between the leader and follower vehicles. Such an algorithm must have a method to zoom out in response to poor tracking so that target motions faster than the ability of the camera to track won't cause catastrophic failure.

8 Future Work

The leader/follower system demonstrated in this project is not yet field-ready, and a number of improvements need to be made before it can be used by the Canadian Forces.

Firstly, the immunity of the vision system to changing lighting conditions must be improved. One approach would be the introduction of other sensing modalities (i.e. visual cues), such as using colour histograms over the object, rather than one colour alone. This would provide better tracking of multi-coloured objects. A shape or silhouette tracker using contours could improve invariance to outdoor lighting conditions. A learning algorithm could be produced where successful SIFT recognitions could be used to train colour system to adapt to changing conditions. The implementation of a foveal vision system is also planned, such that if the zoom camera loses tracking a wider angle camera could re-fixate the zoom camera on the target. For night-time operations, the investigation of recognizing targets in IR images (Figure 42) should be investigated.

Figure 42: *An infrared image of an armoured vehicle travelling on a road.*

Secondly, the speeds the vehicles were driving are not operationally relevant, and will need to be increased to be practically useful. The limiting factors were the inaccuracy of the follower's heading measurement using GPS or odometry, and the inaccuracy of the vision data over long distances. Both of these factors meant that when the leader was than 25m away, its path as estimated by the follower became noisy and erratic. Three solutions will be implemented to improve this:

1. A heading gyro will be added on the follower to provide a heading reading which is more responsive than the GPS or odometry heading estimate.
2. Increased data filtering of the vision data when travelling at higher speeds to smooth the estimated leader's path on straightaways, while still maintaining accuracy around corners at slower speeds.
3. A maximum allowed following distance will be implemented such that the follower will follow at a set time behind the leader at slow speeds, and at a set distance behind the leader at higher speeds.

Thirdly, the safety systems on the follower vehicle must improved. Tele-operated control or remote kill will be necessary to ensure the robotic vehicle does not cause damage in the case of system failure. Obstacle sensors should be added to ensure that any humans or vehicles moving between the leader and follower are not struck, and vision-based road following algorithms could allow the vehicle to stay more safely in the center of the road under conditions of noisy data.

Finally, the system needs to be tested on actual logistics vehicles (trucks) in realistic scenarios to evalutate the practicality of vision-based convoying in the long term.

This page intentionally left blank.

References

[1] Jaczkowski, J. (2003), Robotic Follower Experimentation Results, *Proceedings SPIE Unmanned Ground Technology*, Vol. 5083.

[2] Kato, S., Tsugawa, S., Tokuda, K., Matsui, T., and Fujii, H. (2002), Vehicle Control Algorithms for Cooperative Driving With Automated Vehicles and Intervehicle Communications, *IEEE Transactions on Intelligent Transporation Systems*, 3(3).

[3] Seto, Y. and Inoue, H. (1999), Development of platoon driving in AHS, *JSAE Review*, 20, 93–99.

[4] Giesbrecht, J., Penzes, S., and Fairbrother, B. (2008), The Multi-Agent Tactical Sentry, (DRDC Suffield TR 2008-148) Defense R&D Canada – Suffield, Medicine Hat, Alberta.

[5] DefendIR (online), ICX Technologies. http://www.icxt.com/products/icx-surveillance/thermal-imaging/defendir/. (Access date: 19 Feb. 2009).

[6] Giesbrecht, J. (2009), A Pan/Tilt/Zoom Tracking System for Unmanned Ground Vehicles, Master's thesis, University of Calgary, Calgary, Alberta, Canada.

[7] Qt Product Website (online), Trolltech Inc. http://trolltech.com/products/qt/. (Access date: 19 Feb. 2009).

[8] OpenCV Website (online), Intel Corporation. http://www.intel.com/research/mrl/research/opencv/. (Access date: 17 Mar. 2008).

[9] Munich, M., Pirjanian, P., Bernardo, E., Goncalves, L., Karlsson, N., and Lowe, D. (2006), SIFT-ing Through Features with ViPR, *IEEE Robotics and Automation*, 13(3), 72–77.

[10] (2002), Technology Development for Army Unmanned Ground Vehicles, Committee On Army Unmanned Ground Vehicle Technology, National Academies Press.

[11] Vasseur, L., Lecointe, O., Dento, J., Cherfaoui, N., Marion, V., and Morillon, J. (2004), Leader-follower function for autonomous military convoys, *Proceedings SPIE Unmanned Ground Technology Conference*, Vol. 5422.

[12] Smith, J. (2001), A Tracking System for a Seismic Surveying Mobile Robot, Master's thesis, University of Alberta, Edmonton, Alberta, Canada.

[13] Giesbrecht, J. (2006), Leader/Follower Behaviour Using the SIFT Algorithm for Object Recognition, (DRDC Suffield TM 2006-108) Defence R&D Canada – Suffield.

[14] Sattar, J., Giguere, P., Dudek, G., and Prahacs, C. (2005), A Visual Servoing System for and Aquatic Swimming Robot, In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[15] Ng, T. C. and Guzman, J. (2005), Autonomous Vehicle-Following Systems: A Virtual Trailer Link Model, *Proceedings of IEEE/RSJ Conference on Intelligent Robots and Systems*, pp. 3057– 3062.

[16] Chiem, S. and Cervera, E. (2004), Vision-Based Robot Formations with Bezier Trajectories, In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[17] Giesbrecht, J., Mackay, D., Collier, J., and Verret, S. (2005), Path Tracking for Unmanned Ground Vehicle Navigation, (DRDC Suffield TM 2005-224) Defence R&D Canada – Suffield.

[18] Dual-beam Radar Back-up Aid Website (online), Delphi Corp. http://delphi.com/manufacturers/cv/safety/active/dualbeam/. (Access date: 19 Feb. 2009).

[19] Jaczkowski, J. J. (2003), Robotic Follower Experimentation Results, *Proceedings 3rd Annual Intelligent Vehicle Systems Symposium V*, 5083(311).

[20] Bumblebee Stereo Vision Camera Website (online), Point Grey Research. http://www.ptgrey.com/products/bbxb3/index.asp. (Access date: 19 Feb 2009).

[21] Kubinger, W., Borbely, S., Hemetsberger, H., and Isaacs, R. (2005), Platform for evalutation of embedded computer vision, *Proceedings 13th European Signal Processing Conference*, pp. 130–133.

[22] Nguyen, H., Kogut, G., Barua, R., and Burmeister, A. (2004), A Segway RMP-based Robotic Transport System, *Proceedings of SPIE Mobile Robots XVII*, 5609, 244–255.

[23] Juberts, M., Murphy, K., Nashman, M., Scheiderman, H., Scott, H., and Szabo, S. (1993), Development And Test Results for a Vision-Based Approach to AVCS, *Proceedings ISATA meeting on Advanced Transport Telematics / Intelligent Vehicle Highway Systems*.

[24] Benhimane, S., Malis, E., Rives, P., and Azinheira, J. (2005), Vision-based control for car platooning using homography decomposition, *Proceedings of the IEEE International Conference on Robotics and Automation*.

[25] Sukthankar, R. (1993), Racoon: A Real Time Autonomous Car Chaser Operating Optimally at Night, *Proceedings of the IEEE Intelligent Vehicles and Systems*.

[26] Parker, L., Kannan, B., Tang, F., and Bailey, M. (2004), Tightly-Coupled Navigation Assistance in Heterogenous Multi-Robot Teams, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[27] Das, A., Fierro, R., Kumar, V., Southall, J., Spletzer, J., and Taylor, C. (2001), Real-time vision based control of a nonholonomic mobile robot, In *Proceedings IEEE Int. Conference on Robotics and Automation*, pp. 1714–1719.

[28] Balkenius, C. and Kopp, L. (1996), Visual Tracking and Target Selection for Mobile Robots, *Proceedings of First Euromicro Workshop on Advanced Mobile Robot*, pp. 166–171.

[29] Michaud, F., Lepage, P., Frenette, P., Letourneau, D., and Gaubert, N. (2006), Coordinated Maneuvering of Automated Vehicles in Platoons, *IEEE Transactions on Intelligent Transportation Systems*, 7(4), 437–447.

[30] Kania, R. and Rose, M. D. (2005), Autonomous Robotic Following Using Vision Based Techniques, Technical Report Army Research Development and Engineering Command, Visual Perception Lab.

[31] Atkins, E., Lennon, J., and Peasco, R. (2002), Vision-based Following for Cooperative Astronaut-Robot Operations, *Proceedings IEEE Aerospace Conference*.

[32] Mariottini, G., Pappas, G., Prattichizzo, D., and Daniilidis, K. (2005), Vision-based Localization of Leader-Follower Formations, *Proceedings of IEEE Conference On Decision Making and Control*.

[33] Guidline Website (online), Autonomous Solutions Inc. http://www.autonomoussolutions.com/products/guideline.php. (Access date: 19 Feb. 2009).

[34] Rafi, F., Khan, S., Shafiq, K., and Shah, M. (2006), Autonomous Target Following by Unmanned Aerial Vehicles, *Proceedings of SPIE Unmanned Systems Technology VII*, Vol. 6230.

[35] Chitrakaran, V., Dawson, D., Kannan, H., and Feemster, M. (2006), Vision-Based Tracking for Unmanned Aerial Vehicles, Technical Report Clemson University Department of Electrical and Computer Engineering.

[36] Kia, C. and Arshad, M. R. (2005), Robotics Vision-based Heuristic Reasoning for Underwater Target Tracking and Navigation, *International Journal of Advanced Robotic Systems*, 2(3), 245–250.

[37] Zhou, J. and Clark, C. (2006), Autonomous fish tracking by ROV using Monocular Camera, In *Proceedings Canadian Conference on Computer and Robot Vision*, pp. 68–68.

[38] Okamoto, A. (2003), Robust Control of a Platoon of Autonomous Underwater Vehicles, Master's thesis, University of Idaho.

[39] Naeem, W., Sutton, R., and Ahmad, S. (2004), Pure Pursuit Guidance and Model Predictive Control of an Autonomous Underwater Vehicle for Cable/Pipeline Tracking, *IMarEST Journal of Marine Science and Environment*, (1), 15–25. Part C.

[40] Chen, Z. and Birchfield, S. (2006), Autonomous Person Following Using A Vision-Based Mobile Robot, Technical Report National Institute for Medical Informatics.

[41] Cervera, E., Martinet, P., and Renaud, P. (2004), Towards a reliable vision-based mobile robot formation control, *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 4, 3176– 3181.

[42] Orqueda, O. and Fierro, R. (2006), A Vision-based Nonlinear Decentralized Controller for Unmanned Vehicles, *Proceedings IEEE International Conference on Robotics and Automation.*

[43] Clady, X., Collange, F., Jurie, F., and Martinet, P. (2001), Object tracking with a pan-tilt-zoom camera: application to car driving assistance, In *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1653 – 1658.

[44] Royer, E., Bom, J., Dhome, M., Thuilot, B., Lhuillier, M., and Marmoiton, F. (2005), Outdoor autonomous navigation using monocular vision, In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems.*

[45] Belkhouche, F. and Belkhouche, B. (2005), Modeling and controlling a robotic convoy using guidance law strategies, *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 35(4), 813–825.

[46] Swaroop, D. (1997), String Stability of Interconnected Systems: An Application to Platooning in Automated Highway Systems, (California PATH Program Research Report UCV-ITS-PRR-97-14) University of California, Berkeley.

[47] Lepetit, V. and Fua, P. (2005), Monocular Model-Based 3D Tracking of Rigid Objects A Survey, *Computer Graphics and Vision*, 1(1), 1–89.

[48] Kato, H. and Billinghurst, M. (1999), Marker tracking and hmd calibration for a video-based augmented reality conferencing system, In *Proceedings of the 2nd International Workshop on Augmented Reality.*

[49] Fiala, M. (2004), ARTAg Revision1, A Fiducial Marker System Using Digital Techniques, (Technical Report NRC/ERB-1117) National Research Council of Canada.

[50] Cho, Y. and Neumann, U. (1998), Multi-ring Color Fiducial Systems for Scalable Fiducial Tracking Augmented Reality, In *Proceedings of IEEE 1998 Virtual Reality Annual International Symposium (VRAIS '98).*

[51] Bondy, M., Krishnasamy, R., Crymble, D., and Jasiobedzki, P. (2007), Space Vision Marker System, In *Proceedings of AIAA SPACE Conference.*

[52] Carceroni, R., Harman, C., Eveland, C., and Brown, C. (1998), Design and Evaluation of a System for Vision-Based Vehicle Convoying, (Technical Report TR-678) University of Rochester.

[53] Hong, P., Sahli, H., Colon, E., and Baudoin, Y. (2001), Visual Servoing for Robot Navigation, In *Third International Conference on Climbing and Walking Robots.*

[54] Mio, M., Tachibana, A., Aoki, K., and Nishida, M. (1995), Platoon System based on optical inter-vehicle communication, *Intelligent Transport Systems World Congress*, 1, 1272–1277.

[55] Armstrong, M. and Zisserman, A. (1995), Robust object tracking, In *Proceedings Asian Conference on Computer Vision*.

[56] Gennery, D. (1992), Visual tracking of known three-dimensional objects, *International Journal of Computer Vision*, 7, 243–270.

[57] Preisig, P. and Kragic, D. (2006), Robust Statistics for 3D Object Tracking, In *Proceedings IEEE InternationalRobotics and Automation*.

[58] Wavering, A. J., Schneiderman, H., and Fiala, J. C. (1995), High-Performance Tracking with TRICLOPS, In *Asian Conference on Computer Vision*.

[59] Vacchetti, L. and Lepetit, V. (2004), Stable Real-Time 3D Tracking Using Online and Offline Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10), 1385–1391.

[60] Merad, D., Didier, J., and L, M. S. (2006), Tracking 3D free form object in video sequence, In *Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision*.

[61] Lou, J., Tan, T., Hu, W., Yang, H., and Maybank, S. (2005), 3-D Model-based Vehicle Tracking, *IEEE Transactions on Image Processing*, 14(10), 1561–1569.

[62] Juza, M., Marik, K., Rojicek, J., and Stluka, P. (2006), 3D Template-Based Single Camera Multiple Object Tracking, In *Computer Vision Winter Workshop*, Czech Pattern Recognition Society.

[63] Roller, D., Daniilidis, K., and Nagel, H. (1993), Model-based object tracking in monocular image sequences of road traffic scenes, *International Journal of Computer Vision*, 10(3), 257–281.

[64] Broggi, A., Bertozzi, M., Fascioli, A., Bianco, C., and Piazza, A. (2000), Visual Perception of Obstacles and Vehicles for Platooning, *IEEE Transactions on Intelligent Transportation Systems*, 1(3), 164–176.

[65] Wu, H., Chen, Q., Oike, H., Hua, C., Wada, T., and Kato, T. (2005), High Performance Object Tracking System Using Active Cameras, *Proceedings of the IEEE International Conference on Computer Vision*, 37(213-223).

[66] Li, B., Hu, H., and Spacek, L. (2003), An Adaptive Color Segmentation Algorithm for Sony Legged Robots, In *Proceedings of the 21st IASTED International Conference on Applied Informatics*.

[67] Fasola, J. and Veloso, M. (2006), Real-Time Object Detection Using Segmented and Grayscale Images, In *Proceedings IEEE International Conference on Robotics and Automation*.

[68] Browning, B. and Veloso, M. (2005), Real-time, adaptive color-based robot vision, In *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 3871– 3876.

[69] CMUVision Website (online), Carnegie Mellon University. http://www.cs.cmu.edu/ jbruce/cmvision/. (Access date: 19 Feb 2009).

[70] Cognachrome Website (online), Newton Labs Inc. http://www.newtonlabs.com/cognachrome/. (Access date: 2 Nov 2008).

[71] Sattar, J., Giguere, P., Dudek, G., and Prahacs, C. (2005), A Visual Servoing System for and Aquatic Swimming Robot, In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1483– 1488.

[72] Tsalatsanis, A., Valavanis, K., and Yalcin, A. (2007), Vision Based Target Tracking and Collision Avoidance for Mobile Robots, *Journal of Intelligent and Robotic Systems*, 48(2), 285–304.

[73] Schlegel, C., Illmann, J., Jaberg, H., Schuster, M., and Worz, R. (1998), Vision based person tracking with a mobile robot, In *Proceedings Ninth British Machine Vision Conference.*

[74] Elgammal, A., Duraiswami, R., and Davis, L. (2001), Efficient nonparametric adaptive color modeling using fast gauss transform, In *Proc. IEEE Conf. Computer Vision and Pattern Recognition.*

[75] Bibby, C. and Reid, I. (2005), Visual Tracking at Sea, In *Proceedings of IEEE International Conference on Robotics and Automation.*

[76] Comaniciu, D., Ramesh, V., and Meer, P. (2000), Real-Time Tracking of Non-Rigid Objects using Mean Shift, In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*

[77] Nummiaro, K., Koller-Meier, E., and Gool, L. V. (2002), A color-based particle filter, In *International Workshop on Generative-Model-Based Vision.*

[78] Muselet, D., Macaire, L., and Postaire, J. (2005), Color Histograms Adapted to Query-Target Images for Object Recognition across Illumination Changes, *EURASIP Journal of Applied Signal Processing*, 14, 2164–2172.

[79] Bradski, G. R. (1998), Computer Vision Face Tracking For Use in a Perceptual User Interface, *Intel Technology Journal*, 1(Q2).

[80] Lu, X. and Zhang, H. (2006), Color Classification Using Adaptive Dichromatic Model, In *Proceedings IEEE International Conference on Robotics and Automation.*

[81] Leow, W. and Miikkulainen, R. (1997), Visual schemas in neural networks for object recognition, *Connection Science*, 9, 161–200.

[82] Schneiderman, H. (2004), Learning a Restricted Bayesian Network for Object Detection, In *IEEE Conference on Computer Vision and Pattern Recognition.*

[83] Malis, E. and Benhimane, S. (2005), A unified approach to visual tracking and servoing, *Robotics and Autonomous Systems*, 52, 39–52.

[84] Zhang, Z., Deriche, R., Faugeras, O. D., and Luong, Q.-T. (1995), A Robust Technique for Matching two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry, *Artificial Intelligence*, 78(1-2), 87–119.

[85] Tomasi, C. and Kanade, T. (1991), Detection and Tracking of Point Features, (Technical Report CMU-CS-91-132) Carnegie-Mellon University.

[86] Lowe, D. (2004), Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60(2), 91–110.

[87] Mikolajczyk, K. and Schmidt, C. (2003), A performance evaluation of local descriptors, In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*.

[88] Moreels, P. and Perona, P. (2007), Evaluation of features detectors and descriptors based on 3D objects, *International Journal of Computer Vision*, 73(3), 263–284.

[89] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2005), A comparison of affine region detectors, *International Journal of Computer Vision*, 1(65), 43–72.

[90] Lee, S., Kim, E., and Park, Y. (2006), 3D Object Recognition Using multiple features for robotic manipulation, In *Proceedings IEEE International Conference on Robotics and Automation*.

[91] A. Shademan, J. F. (2004), Using scale-invariant feature points in visual servoing, In *Proceedings of the SPIE Optomechatronic Sensors, Actuators, and Control*, pp. 63–70.

[92] Kang, S., Paikab, J., Koschana, A., Abidia, B., and Abidi, M. A. (2003), Real-time video tracking using PTZ cameras, In *Proceedings of SPIE 6th International Conference on Quality Control by Artificial Vision*.

[93] Barreto, J. P., Batista, J., Peixoto, P., and Araujo, H. (2002), Integrating Vision and Control to Achieve High Perfomance Active Tracking, Technical Report ISR/DEEC – University of Coimbra, Coimbra, Portugal.

[94] Shiao, Y.-S. (2001), Design and Implementation of Real-Time Tracking System Based on Vision Servo Control, *Tamkang Journal of Science and Engineering*, 4(1), 45–58.

[95] Coombs, D. and Brown, C. (1993), Real-Time Binocular Smooth-Pursuit, *International Journal of Computer Vision*, 11(2), 147–165.

[96] Takeuchi, Y., Wang, Z.-F., Ohnishi, N., and Sugie, N. (1995), In *Proceedings of Second Asian Conference on Computer Vision*.

[97] Blackburn, M. and Nguyen, H. (1995), Vision Based Autonomous Robot Navigation: Motion Segmentation, In *Proc. International Symposium on Automotive Technology and Automation Conference on Robotics, Motion, and Machine Vision in the Automotive Industries*.

[98] Olson, C. and Huttenlocher, D. (1997), Automatic Target Recognition by Matching Oriented Edge Pixels, *IEEE Transactions on Image Processing*, 6(1), 103–113.

[99] Ferrari, V., Fevrier, L., Jurie, F., and Schmid, C. (2006), Groups of Adjacent Contour Segments for Object Detection, (Technical Report 5980) INRIA Technical Report.

[100] Belongie, S., Malik, J., and Puzich, J. (2001), Shape matching and object recognition using shape contexts, (Technical Report CSD00 -1128) UC Berkeley.

[101] Bellotto, N. and Hu, H. (2006), Vision and laser data fusion for tracking people with a mobile robot, In *IEEE International Conference On Robotics and Biomimetics*.

[102] Yanlin, G., Sawhney, H., Kumar, R., and Hsu, S. (2007), Robust object matching for persistent tracking with heterogeneous features, *IEEE Transactions on Pattern Analysis Machine Intelligence*, 29(5), 824–839.

[103] Xiong, T. and Debrunner, C. (2004), Stochastic Car Tracking with Line and Color Based Features, *IEEE Transactions on Intelligent Transportation Systems*, 5(4).

[104] Sharkey, P. and Murray, D. (1996), Delay versus performance of visually guided systems, *IEEE Proc.-Control Theory and Application*, 143(5), 436–447.

[105] Papanikolopoulos, M. and Pradeep, K. (1993), Visual Tracking of a Moving Target by a Camera Mounted on a Robot: A Combination of Control and Vision, *IEEE Transactions on Robotics and Automation*, 9, 14–35.

[106] Hosoda, K., Sakamoto, K., and Asada, M. (1995), Trajectory generation for obstacle avoidance of uncalibrated stereo visual servoing without 3d reconstruction, In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*.

[107] Corke, P. (1993), Experiments in high-performance robotic visual servoing, In *Proceedings of International Symposium on Experimental Robotics*.

[108] Corke, P. and Good, M. (1993), Controller design for high-performance visual servoing, In *Proceedings of IFAC World Congress*.

[109] Corke, P. and Good, M. (1996), Dynamic effects in visual closed-loop system, *IEEE Transactions on Robotics and Automation*, 12(5), 671–683.

[110] Malis, E. (2002), Survey of vision-based robot control, In *European Naval Ship Design, Captain Computer IV Forum*.

[111] Wavering, A. and Lumia, R. (1993), Predictive Visual Tracking, In *Proceedings of SPIE Intelligent Robots and Compter Vision XII*.

[112] Fiala, J. and Wavering, A. (1994), On Visual Pursuit Systems, Technical Report NIST Internal Report 5513.

[113] Barreto, J., Batista, J., Peixoto, P., and Araujo, H. (2002), Integrating Vision and Control to Achieve High Perfomance Active Tracking, Technical Report ISR/DEEC – U. of Coimbra.

[114] Barreto, J., Peixoto, P., Batista, J., and Araujo, H. (1999), Robust Vision for Vision-Based Control of Motion, Ch. Evaluation of the robustness of visual behaviors through performance characterization, IEEE Press.

[115] Barreto, J., Batista, J., and Araujo, H. (2000), Model Predictive Control to Improve Visual Control of Motion Applications in Active Tracking of Moving Targets, In *Proceedings International Conference on Pattern Recognition*.

[116] Oh, P. (2002), A Biologically Inspired Approach to Visually Servoed Tracking, In *Proceedings CISST Motion Tracking and Object Recognition Conference*.

[117] Daniilidis, K., Krauss, C., Hansen, M., and Sommer, G. (1998), Real-time tracking of moving objects with an active camera, *Real-Time Imaging*, 4, 3–20.

[118] Naeem, W., Sutton, R., and Ahmad, S. (2003), LQG/LTR Control of an Autonomous Underwater Vehicle using a Hybrid Guidance Law, In *Proc 6th of IFAC Conference on Manoeuvring and Control of Marine Craft*.

[119] Hatano, K. and Hashimoto, K. (2003), Image-Based Visual Servo Using Zoom Mechanism, *Proceedings of SICE Annual Conference*.

[120] Nelson, E. (2006), Zoom Techniques for Achieving Scale Invariant Object Tracking in Real-Time Active Vision Systems, Ph.D. thesis, Rochester Institute of Technology.

[121] Fayman, J., Sudarsky, O., Rivlin, E., and Rudzsky, M. (2001), Zoom tracking and its applications, *Machine Vision and Applications*, 13(1), 25–37.

[122] Mirmehdi, M., Palmer, P., and Kittler, J. (1997), Towards optimal zoom for automatic target recognition, In *Proceedings of the Tenth Scandinavian Conference in Image Analysis*.

[123] Shah, H. and Morrell, D. (2004), An Adaptive Zoom Algorithm For Tracking Targets Using Pan-Tilt-Zoom Cameras, In *Proceedings of ICASSP*.

[124] Denzler, J., Zobel, M., and Niemann, H. (2003), Information Theoretic Focal Length Selection for Real-Time Active 3-D Object Tracking, In *Proceedings Ninth IEEE International Conference on Computer Vision*.

[125] Tordoff, B. and Murray, D. (2007), A method of reactive zoom control from uncertainty in tracking, *Computer Vision and Image Understanding*, 105(2), 131–144.

[126] Soto, A. (2002), A Probabilistic Approach for the Adaptive Integration of Multiple Visual Cues Using an Agent Framework, Ph.D. thesis, Carnegie Mellon University.

[127] Batavia, P. and Singh, S. (2001), Obstacle Detection Using Adaptive Color Segmentation and Color Stereo Homography, *Proceedings of the IEEE International Conference on Robotics and Automation*.

[128] Forsyth, D. and Ponce, J. (2003), Computer Vision: A Modern Approach, Prentice Hall.

[129] Ljung, L. and Soderstrom, T. (1983), Theory and Practice of Recursive Estimation, Cambridge, MA, USA: MIT Press.

[130] Goodwin, G. and Sin, K. (1984), Adaptive Filtering, Prediction and Control, Englewood Cliffs, NJ, USA: Prentice-Hall.

[131] Alvarez, L. and Nickerson, J. (1989), Application of optimal control theory to the design of the NASA/JPL 70-meter antenna servos, Technical Report Jet Propulsion Labratory.

[132] Grimble, M. and Hearns, G. (1998), LQG Controllers for Unstable Systems with Transport Delays, In *Proceedings of the 37th IEEE Conference on Decision and Control*.

[133] Gehrig, S. and Stein, F. (1998), A Trajectory-Based Approach for the Lateral Control of Car Following Systems, *IEEE Intl. Conf. On Systems, Man and Cybernetics*, pp. 3596–3601.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Suffield PO Box 4000, Station Main, Medicine Hat, AB, Canada T1A 8K6	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Development of a Vision-Based Robotic Follower Vehicle

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.)

Giesbrecht, J.

5. DATE OF PUBLICATION (Month and year of publication of document.) February 2009	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.) 80	6b. NO. OF REFS (Total cited in document.) 133

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Suffield
PO Box 4000, Station Main, Medicine Hat, AB, Canada T1A 8K6

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Suffield TR 2009-026	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
( ) Defence departments and defence contractors; further distribution only as approved
( ) Defence departments and Canadian defence contractors; further distribution only as approved
( ) Government departments and agencies; further distribution only as approved
( ) Defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11)) is possible, a wider announcement audience may be selected.)

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

This report provides an overview of the development of a vision-based leader/follower robotic vehicle at Defence R&D Canada – Suffield, with the eventual goal of autonomous convoying for military logitics. The experimental system uses a pan/tilt/zoom camera to track a lead vehicle or human, estimating the leader's path and following it autonomously. This vision-based approach frees the system from reliance on GPS, radios, and active sensing equipment necessary for current leader/follower systems. Included in this report are the details of the computer vision, camera control, and vehicle control algorithms, as well as the results of field trials of the camera tracking system. Finally, it reports on experiments with the complete follower system following other vehicles and even dismounted humans.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

robotics
navigation
convoying
control
Canadian Forces

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca