**2014 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM**
**AUTONOMOUS GROUND SYSTEMS (AGS) TECHNICAL SESSION**
**AUGUST 12-14, 2014 - NOVI, MICHIGAN**

# ENABLING SEMI-AUTONOMOUS BEHAVIORS WITH A SINGLE CAMERA

**Camille Monnier**
**Andrey Ost**
**Stan German**
Charles River Analytics
Cambridge, MA

## ABSTRACT

*Semi-autonomous behaviors, such as leader-following and "point-and-go" navigation, have the potential to significantly increase the value of squad-level UGVs by freeing operators to perform other tasks. A variety of technologies have been designed in recent years to enable such semi-autonomous behaviors on board mobile robots; however, most current solutions use custom payloads comprising sensors such as stereo cameras, LIDAR, GPS, or active transmitters. While effective, these approaches tend to be restricted to UGV platforms capable of supporting the payload's space, weight, and power (SWaP), and may be cost-prohibitive to large-scale deployment. Charles River has developed a system that enables both leader-following and "point-and-go" navigation behaviors using only a single monocular camera. The system allows a user to control a mobile robot by leading the way and issuing commands through arm/hand gestures, and is capable of following an operator both on foot and aboard a vehicle. The operator may equally direct the robot via a lightweight interface, by simply indicating an object of interest or destination in the robot's camera view.*

## INTRODUCTION

Teleoperation remains the dominant form of control for UGVs, even for comparatively simple tasks such as travelling between locations. This need for continuous "heads-down" operation of a UGV places an undesirable burden, both physical and cognitive, on human operators of these systems – as a result, squad-level unmanned systems are currently used as on-demand tools that temporarily provide a new situation-specific capability (e.g., reconnaissance or explosive ordnance disposal (EOD)) at the cost of decreased mobility and local situational awareness. The development of semi-autonomous behaviors, such as leader-following and "point-and-go" navigation, represents an important step toward reducing the cognitive loads associated with operating a UGV, and freeing operators to perform other tasks. A variety of technologies have been designed in recent years to enable such semi-autonomous behaviors on board mobile robots; however, most current solutions use custom payloads comprising sensors such as stereo cameras, LIDAR, GPS, or active transmitters. Although effective, this approach limits the portability of such technology to UGV platforms capable of supporting the payload's space, weight, and power (SWaP), and may be cost-prohibitive to large-scale deployment. We present a Monocular Unmanned Leader-Follower (MULE-F) system that enables both leader-following and "point-and-go" navigation behaviors using only a single monocular camera. The system allows a user to control a mobile robot by leading the way and issuing commands through arm/hand gestures, and differentiates between the leader and nearby pedestrians. The system is capable of following an operator both on foot and aboard a vehicle. The operator may equally direct the robot via a lightweight interface, by simply indicating an object of interest or destination in the robot's camera view. We have evaluated the system's capabilities on publicly available benchmark datasets, as well as in representative scenarios captured using small and medium-sized unmanned ground vehicles.

## DESIGN

The MULE-F system is a Robot Operating System (ROS) based software system that may be installed as a stand-alone capability as part of a lightweight appliqué, or on an existing

onboard computer. Figure 1 illustrates the system's major components. The system enables a human operator to issue commands via hand and arm gestures, and may by equally controlled via a wireless operator control unit (OCU). The system requires only a single camera and small computer. Figure 2 illustrates a typical hardware configuration on board a small unmanned ground vehicle (SUGV).

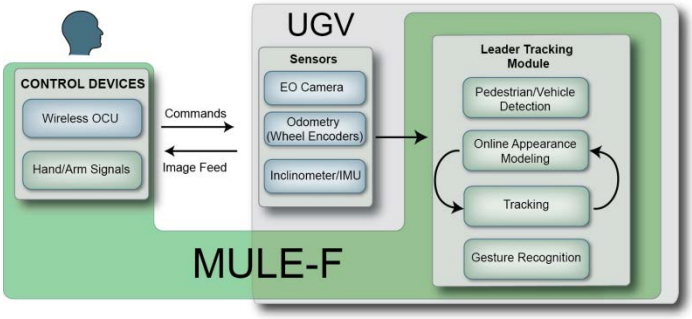


**Figure 1: MULE-F Architecture**

**Figure 2: Vision-based autonomy appliqué components, consisting of an Allied Vision Technology Mako camera and Dynamixel servo (left), and a compact, fanless IntensePC computer (center). System installed on a Dragon Runner SUGV (right).**

## LEADER-FOLLOWING

Charles River Analytics has developed a compact system to autonomously track and follow a designated dismounted operator [1,2]. The system enables a UGV to autonomously follow an operator on foot, using a lightweight video camera and requiring no modifications to the leader's equipment (in particular, no special clothing, markers or transmitters are required). The system is designed to require only a single color camera, with the intent to maximize portability across existing UGVs. The core software capabilities are designed with an emphasis on reduced computational complexity to minimize the payload's size, weight, power and cost (SWAP-C) impact. The system is comprised of three core modules: a *pedestrian and vehicle detection module* that determines the locations of all visible pedestrians and vehicles in the UGV camera's field of view; an *appearance-learning and tracking module* that maintains a lock on the leader and differentiates between the leader and nearby pedestrians and vehicles; and a *gesture recognition module* that enables natural control of the vehicle. The system integrates efficient object detection software with kinematic tracking and online appearance learning techniques to reliably track a leader in complex outdoor environments.

Our approach to pedestrian and vehicle detection is based on a sliding-window approach, in which a previously trained classifier is evaluated at multiple locations and scales in an image. At each location in the search space, which typically comprises ~50,000-100,000 windows in a 640x480 image, a battery of heterogeneous features including edge, contrast, and intensity distributions are computed and processed by an efficient boosted cascade classifier. Resulting detections are collected and processed by a kinematics and appearance-based tracking algorithm that performs data association and state estimation. The trackers were evaluated on publicly-available benchmark datasets for pedestrian and vehicle detection, exceeding 90% detection accuracy at a $10^{-1}$ false positives per image (FPPI) on the INRIA pedestrian dataset [3] and TME Motorway tracking dataset [4]. Spurious false positives and missed detections are resolved through the tracking framework, which is based on a particle filter implementation [5]. Figure 3 illustrates typical outputs for the pedestrian and vehicle tracker on several datasets, including the Performance Evaluation of Tracking and Surveillance (PETS) 2009 dataset [6], ETH Pedestrians [7], and TME Motorway.

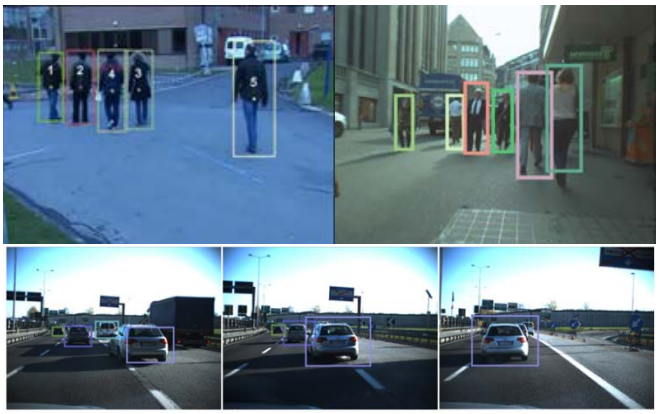

**Figure 3: Tracking outputs on benchmark datasets. Clockwise from the top-left: PETS 2009; ETH Pedestrians; TME Motorway.**

The end-to-end tracking system operates at 20Hz, enabled by efficient sharing and re-use of features across detection, tracking, and gesture recognition modules. The leader-following system has been integrated and tested on multiple mobile robots, including Dragon Runner, TALON, and Segway RMP platforms. Figure 4 illustrates an outdoor test

in which MULE-F follows a human leader in a 1700m loop in a busy Boston public park.

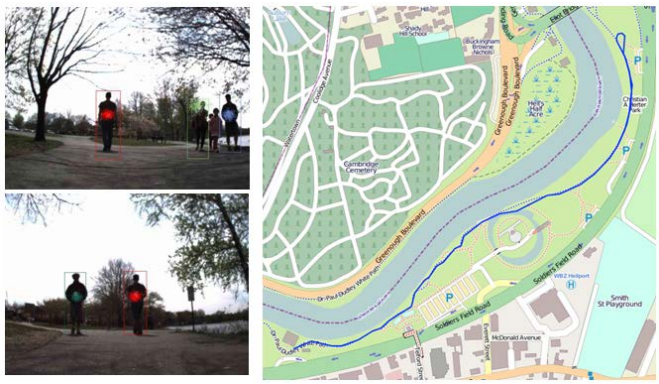

**Figure 4: GPS track and camera views of a system test in a Boston public park. The environment contains a large variety of challenging terrain features including trees and hills, varying lighting, and passers-by.**

### *Gesture Recognition*

Non-verbal communication remains an important component of squad-level interaction, and provides a natural method of signaling intent without the use of specialized equipment. The MULE-F gesture-recognition module is designed to recognize hand and arm signals using only camera input, enabling a user to issue commands to the robot in a rapid and natural way. The system supports leader self-designation, as well as a lexicon of five gestures enabling various degrees of direct motion control, including "start," "stop," "turn left/right," and "move forward". The core software consists of a multi-class classification framework in which $N$ classifiers are trained, such that each classifier is tuned to distinguish between a specific gesture and all other gestures/non-gestures in the dataset. In order to recognize a new example, we find the maximum likelihood gesture model that corresponds to an image: $argmax\ p(\mathrm{I}|\Theta_i)$, where $\Theta_i$ is the model for the $i^{\mathrm{th}}$ gesture and $I$ is the input image. Each gesture classifier consists of a support vector machine (SVM) trained on an extensive dataset of gestures performed by multiple individuals, as well as non-gestural data. For efficiency, the same image features computed for detection are re-used for gesture recognition. As with the pedestrian and vehicle tracking module, detected gestures are temporally filtered to eliminate false positives caused by noise. In our evaluations, the gesture recognition system recognized 98.7% of gestures in a dataset of over 1300 annotated instances.

At initialization, the gesture recognition module analyzes each tracked pedestrian for a "follow me" gesture, consisting of a raised arm. The first pedestrian to issue the command is selected as the leader, which may be manually confirmed via the OCU for added safety. Once the leader has been selected, the system processes gestures only for the track identified as the leader. Figure 5 illustrates the tracking and gesture recognition process, visualized via the developer interface.

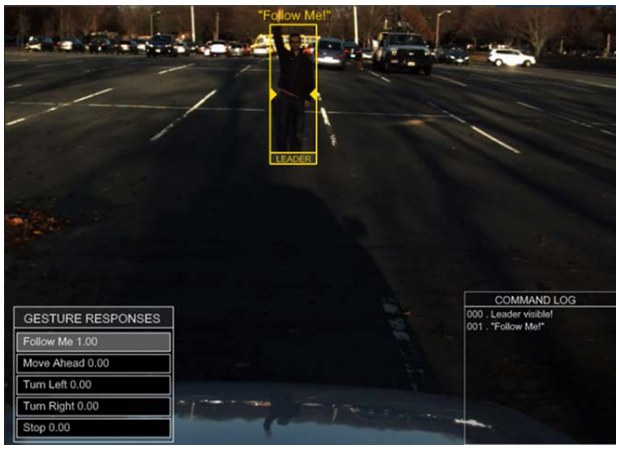


**Figure 5: Visualization of the gesture recognition module, via the development and testing interface. The list of enabled gestures is displayed in the bottom-left corner, while a log of recently issued commands is shown in the bottom-right corner. The leader is highlighted in yellow.**

## POINT-AND-GO NAVIGATION

Robots are often used to investigate potentially dangerous situations; however, remote tele-operation can be cumbersome and fatiguing, and may distract an operator from important developments in his or her immediate surroundings. A technology such as point-and-go navigation, in which an operator may simply select an object of interest in the robot's camera view as a navigation goal, would significantly reduce the need for heads-down operation during simple tasks such as travelling down a road. The MULE-F system enables an operator to specify an object of interest as a destination by simply selecting it on the screen of the OCU. Alternatively, as the system is ROS-based, a third-party interface may issue the same commands using a properly structured ROS message. Figure 6 illustrates usage of the point-and-go navigation component.

### *Destination Tracking*

The point-and-go navigation software is based on an implementation of Median Flow, applied to a densely sampled grid of points within the user-selected region of interest. Points within the region are tracked from frame to frame of video using the Kanade-Lucas-Tomasi (KLT)

approach to optical flow [8]. Median Flow then estimates the quality of each point prediction and filters out the outliers. Point quality is based on normalized cross-correlation (NCC), a commonly used template matching technique, and forward-backward (FB) error [9], which is the Euclidean distance between the initial point trajectory and its trajectory after performing tracking forward and backward in time. Because the filtered points are not independent, they can be treated as part of a larger unit. Calculating the median scale and position displacement over all filtered points then gives us a reliable estimate of the bounding box displacement.

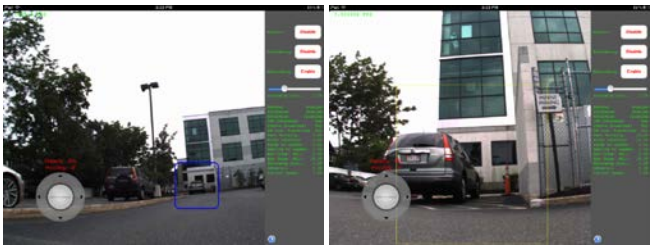

**Figure 6: Point-and-go navigation. The operator may select a visible object and issue a "drive to" command (left) using pinch-to-zoom or a pointing device. The system will then drive autonomously and signal to the operator when the destination has been reached (right).**

To recover from tracker failures due to occlusions, which may be caused by moving people or vehicles in the scene, or by platform motion during obstacle avoidance, the system reacquires tracked objects using a trained appearance model. Figure 7 illustrates a scenario in which moving objects (a pedestrian and vehicle) temporarily occlude the UGV's view of the selected destination. In order to maintain an appearance model robust to changes in viewpoint and background, we continuously update an online classifier. Classifier updates exploit a set of *structural constraints* generated by the Median Flow tracker: if the tracker exhibits high confidence, than we can be relatively certain that detector confidences near the track should be positive and those far away negative. These constraints allow us to guide the learning process by augmenting the training data with only the most difficult examples (i.e. bootstrapping), those that violate the structural constraints. To account for variation in lighting and exposure conditions, we apply NCC between the tracked patch and the initialized patch, to estimate tracker confidence. If confidence is low, we avoid updating the appearance model for that frame to prevent training on incorrectly labeled data. In either case, the detector processes each frame using a scanning window and outputs a maximum-confidence detection. If this detection confidence is greater than the tracker confidence, then the tracker is reinitialized with the new detection patch.

We use a random fern-based classifier due to its computational efficiency and ease of iterative updates, in conjunction with the edge, color, and contrast features produced by the pedestrian and vehicle tracking module.

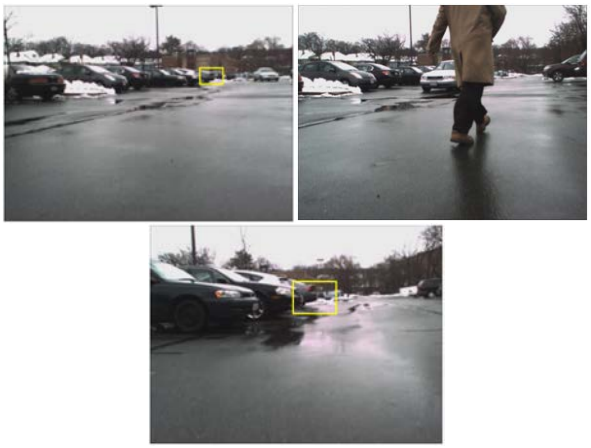

**Figure 7: Occlusion handling during point-and-go navigation. The system is capable of re-acquiring an initially selected goal despite multiple occlusions by moving objects and change in perspective.**

### *Passive Ranging*

Ranging presents a particular challenge for single-camera systems, which must infer distance to an object based on either known properties of the target or environment, or by analyzing an object's change in appearance over time. The MULE-F point-and-go navigation module enables long-distance driving and autonomous stopping by estimating time-to-arrival as a function of the rate of change of the tracked object's angular resolution. The approach is computationally efficient, and functions in the absence of odometry, as measurements are made entirely based on visual information. **Figure 8** illustrates the geometry of the problem.

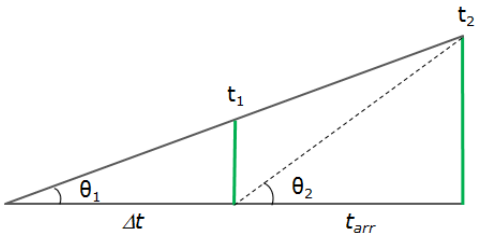


**Figure 8: Time-to-arrival geometry for a visually-tracked object. Time to arrival $t_{arr}$ may be inferred directly given a minimum of two angular observations at different times $t_1$ and $t_2$.**

Given subsequent angular measurements of a fixed-height (or width) object, the object's range may be estimated as a function of time:

$$t_{arr} = dt \frac{\tan(\theta_2)}{\tan(\theta_1)-\tan(\theta_2)} \quad (1)$$

Following (1), absolute metric distance may be calculated given reliable odometry or a constant velocity assumption.

## CONCLUSIONS

We present a lightweight, single-camera system capable of providing semi-autonomous behaviors including leader-following, gesture recognition, convoying, and point-and-go navigation. While sensors such as LIDAR and GPS provide valuable data for obstacle avoidance and navigation, the use of these technologies is restricted to certain types of platforms, and is not appropriate for all applications. In this work, we establish that useful semi-autonomous behaviors may be achieved using only a single camera and lightweight computer, enabling deployment on a wide variety of platforms at minimal cost. We anticipate that continued development in this area will enable the deployment of squad-level robotic teammates capable of acting as semi-autonomous mules, relays, scouts, and perimeter security.

## REFERENCES

[1] Monnier, C., Ostapchenko, A., & German, S., "A Monocular Leader-follower System for Small Mobile Robots". SPIE. Baltimore, MD (2012).

[2] Monnier, C., Ostapchenko, A., & German, S., "Robust leader tracking from an unmanned ground vehicle". SPIE, Baltimore, MD (2013)

[3] Dalal, N. & Triggs, B., "Histograms of oriented gradients for human detection". *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. (2005).

[4] Caraffi, C., Vojíř, T., Trefný, J., Šochman, J., & Matas, J., "Toyota Motor Europe (TME) Motorway Dataset", 2012

[5] Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., & Van Gool, L., "Online Multi-Person Tracking-by-Detection from a Single, Uncalibrated Camera". PAMI. (2010).

[6] Ferryman, J., Shahrokni, A., "PETS2009: Dataset and Challenge", Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009

[7] Ess, A., Leibe, B., Schindler, K. and van Gool, L., "A Mobile Vision System for Robust Multi-Person Tracking". CVPR (2008).

[8] Tomasi, C. & Kanade, T., "Detection and Tracking Point Features", Carnegie Mellon University Tech.Report . (1991).

[9] Kalal, Z., Mikolajczyk, K., & Matas, J., "Forward-backward error: Automatic detection of tracking failures", *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2756-2759. IEEE. (2010).

[10] Ozuysal, M., Calonder, M., Lepetit, V. and Fua, P., "Fast keypoint recognition using random ferns", PAMI (2010)