2012 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
ROBOTIC SYSTEMS (RS) MINI-SYMPOSIUM
AUGUST 14-16, MICHIGAN

COMBAT ID – COMBAT IDENTIFICATION SYSTEM

Garbis Salgian, Zsolt Kira, Raia Hadsell, Han-Pang Chiu, Xun Zhou, Bing-Bing Chai, Supun Samarasekera
SRI International, Princeton, NJ
Bernard Theisen, Jeffery Ramsey
U.S. Army TARDEC, Warren, MI

ABSTRACT

This paper will document the development of the Combat Identification (CombatID) System. The CombatID System was designed to create a platform agnostic payload that could be attached to any fielded Unmanned Ground Vehicle (UGV) to assist the Soldier in contingency basing operations. This paper will describe the approach taken to develop the system, providing a detailed description of the system, including sample results for individual modules. This paper will also provide insight on the evaluation of CombatID system's performance.

INTRODUCTION

The Combat Identification (CombatID) System was designed to create a platform agnostic payload that could be attached to any fielded Unmanned Ground Vehicle (UGV) to assist the Soldier is contingency basing operations. First the robotic payload would need to meet the threshold requirement of reliable detection of Friendly Force and Foe personnel within a 60 meter radius, 180 degrees around the robot. We developed, tested and demonstrated a fully integrated hardware and software solution running on two robot systems and three additional Friendly Force entities. The proposed solution is designed to run through multiple classes of robot systems starting from Small UGV's through large tactical or combat vehicles.

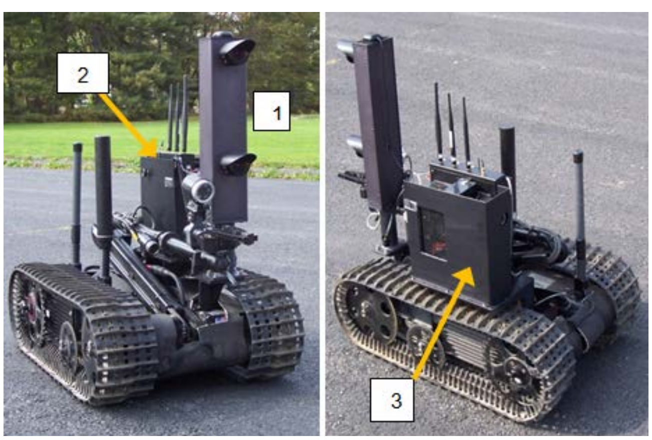

Figure 1: CombatID system installed on a TALON robot.

Figure 1 shows the system installed on a TALON robot. The main components are: (1) the sensor head; (2) the RF-GPS unit and (3) on-board processing platform. The robot sensor system (1+2 in Figure 1) consists of a stereo camera, a MEMs IMU, a GPS receiver, an RF-ranging unit, and a dual-band mesh radio. All processing is done on-board on a small form-factor COTS computer (3) equipped with a low-power GPU card. Each of the friendly (blue-force) soldiers is provided with an identical RF-GPS unit consisting of a lightweight GPS receiver, an RF-ranging unit and a dual-band mesh radio.

This paper will describe the approach taken to develop the system, providing a detailed description of the system, including sample results for individual modules. This paper will provide insight on the evaluation of CombatID system's performance.

SYSTEM DESCRIPTION

The system implements a layered approach to localization, detection, communication and classification. The navigation module on each robot fuses stereo-vision-based visual odometry with readings from the IMU and any available GPS readings to produce a robust localization solution and maintain it even in GPS-denied environments. To further improve robustness, multiple robots equipped with this system can use RF ranging to improve their relative positions. This step is accomplished with a distributed Kalman filter architecture.

Once the robot's geo-location has been established, the friend-foe tracking module integrates this information, along with positions of other robots, video-based detections, and readings from sensors worn by blue-force soldiers to establish identities of the blue-force entities. The soldier-worn RF-GPS units broadcast their positions and range information to the robots over an ad-hoc wireless mesh that does not require any additional infrastructure. The identities and locations received via radio are then correlated with the output of the person and vehicle detectors that operate on stereo images to identify the blue-force entities.

The block diagram of the system is presented in Figure 2. The major subsystems are described in more detail in the following sections.

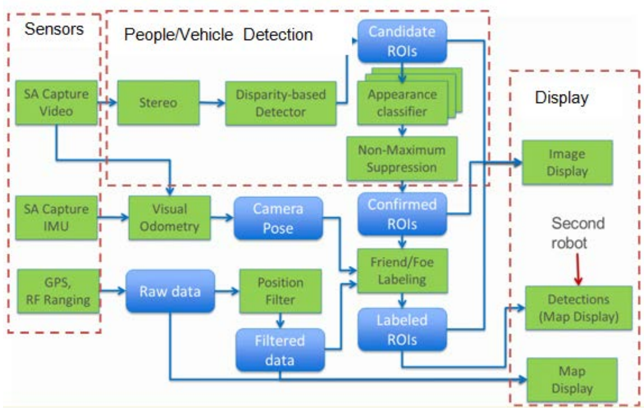

Figure 2: Block diagram for CombatID system.

Sensors

The sensors used for the CombatID system (listed on the left side of Figure 2) are: stereo camera pair, IMU, GPS and Radio Frequency (RF) Ranging.

The stereo camera pair provides range information over a wide Field Of View. We use two 5MegaPixel cameras with a baseline of 27.5cm. The vertical stereo arrangement enables longer baselines with a small footprint, which is important for small robots. One stereo head is equipped with fisheye lenses, which results in 180degrees coverage. For longer range detections, the second head is equipped with 80degrees lenses. This enables reliable detections beyond 60m in front of the camera.

The IMU is rigidly mounted inside the stereo head enclosure and it is used in conjunction with the visual odometry module to compute the current pose of the camera.

The GPS receiver and RF-ranging unit are mounted in a separate enclosure together with a dual-band mesh radio. The RF/GPS unit mounted on the robots is identical to the one carried by friendly soldiers. Each unit broadcasts over the ad-hoc mesh network its current position from GPS and the relative range to the other units obtained from the RF-ranging radio.

People and vehicle detection

The detection of objects in the scene uses only the robot's on-board video. This approach combines 3D shape cues based on stereo vision with appearance cues to classify both stationary and moving persons and vehicles. While pedestrian detection from video has been the focus of significant work in the computer vision literature, most of the previous work addresses only the case where people are relatively close to the camera (less than 40m), with a person being at least 50 pixels tall in the image. We are addressing people detection at long ranges (60m and beyond) using a combination of high-resolution sensors and a novel appearance classification design. The approach is briefly described in the following subsections. For more details, please see Reference 3.

The appearance classifier consists of a cascade of three convolutional neural networks. The first network is trained to classify pedestrians using both appearance and disparity information, without having to hand-design additional features to leverage stereo depth information. A second classifier is specifically designed for long-range detections in order to increase recall and decrease false positives. Finally, we cascade a third, lower-resolution classifier that is faster with the higher-resolution classifier in order to speed up classification.

Stereo Detection

The stereo camera pairs are calibrated to obtain both intrinsic and extrinsic parameters. The fisheye cameras were calibrated using the Matlab toolbox described in Reference 1. The extrinsic parameters were estimated using synchronized images of a checkerboard pattern in both cameras and each fisheye image was then mapped onto a cylindrical image. The second (80 degrees) camera pair was calibrated using the Bouget Matlab calibration toolkit.

We compute stereo disparity maps using a fast CUDA implementation. The resulting disparity image is then used to find vertical structures in the scene that could potentially correspond to pedestrians or vehicles. The image is discretized into a fixed number of patches and we estimate the ground plane for each patch from the disparity information. Next, we create a mask of all disparity pixels that are above this estimated ground and use connected components to group these above-ground pixels together to provide a final estimate of all objects that are vertical. An ROI (region of interest) in the image space is created from the bounding box of each component.

Classification using Convolutional Networks

The ROIs produced by the stereo detector are fed into classifiers that determine whether the ROI corresponds to a pedestrian or vehicle. Previous appearance-based classifiers have achieved some success for cases where there are more than 50 pixels on a pedestrian, but they typically degrade heavily beyond this range. We work with smaller number of pixels per target, either from the fish-eye lenses or from the 80 degree lens at long ranges (pedestrians up to 100 m). Table 1 summarizes the statistics of the data set.

CombatID – Combat Identification System.

Attribute	Fisheye	80 Degree
Frames	24,572	14,533
Detections	99,522	64,204
Mean Height (pixels)	87.78	88.37
% l.t. 50 pixels	32.1	35.6
% l.t. 40	14.4	11.6

Table 1: Pedestrian detection data set.

For classification we use convolutional neural networks, a deep learning method that has achieved success for tasks such as object recognition. This approach supports the use of multiple modalities, in our case Electro-Optical (EO) images and stereo disparity images.

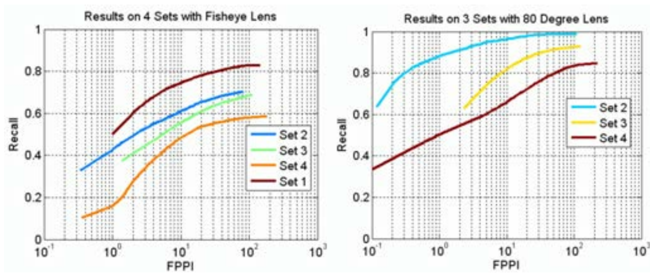

Figure 3: Results of the classification system on a per-image basis, on four subsets of the data. The fisheye (left) and 80 degree (right) demonstrate competitive results at long range.

Figure 3 shows classification results for four data sets using the fisheye camera (left) and latter three subsets for the 80 degree camera (right, note that the first subset was only collected using the fisheye camera). Overall, competitive detection rates at less than 1 false positive per frame can be obtained, especially when using the 80 degree.

		Total Travelled Distance (m)	Loop Closure Error (m)	Drift Rate (%)
Outdoor	Loop 1 Visodo	124.9396	1.1138	0.89
	Loop 1 Visodo+IMU	124.0460	1.0812	0.87
	Loop 2 Visodo	122.4757	0.8724	0.71
	Loop 2 Visodo+IMU	122.3237	0.7168	0.58
Indoor	Loop 1 Visodo	51.2833	0.4648	0.91
	Loop 1 Visodo+IMU	51.3082	0.3699	0.72
	Loop 2 Visodo	105.9501	0.5210	0.49
	Loop 2 Visodo+IMU	105.9180	0.5015	0.47

Table 2: Drift rate and loop closure error for visual odometry with and without IMU data.

Visual Odometry

We formulated and implemented the fisheye camera model into our visual odometry module. The module tracks 2D features across video frames and estimates the 6 degree of freedom relative pose change from one frame to another. We did several long loop closure tests for this new fisheye visual odometry module. Table 2 shows the drift rate and loop closure errors using either fisheye visual odometry or Extended Kalman Filter fusing both IMU and fisheye visual odometry for several test runs, both indoor and outdoor. The drift rates for all experiments are all less than 1%.

Position Filter

Both the GPS positions and the relative range measurements from the RF-ranging radios are noisy. We designed an Extended Kalman Filter that combines the GPS and relative range measurements to provide more stable position estimates. We developed a new relative-polar formulation in EKF for our application (moving RF-ranging nodes, no odometry information).

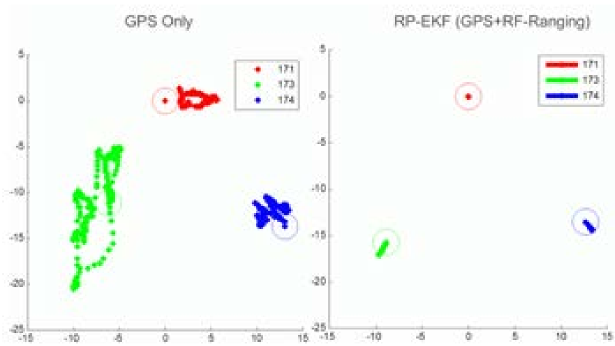

Figure 4: GPS (left) output for three stationary nodes. Position filter (right) output for the same nodes.

Figure 4 shows an example with three stationary GPS/RF-ranging nodes. The left side displays the raw GPS output; note that the drift can be on the order of 10 meters. The right side shows the position filter output, which is much more stable. Figure 5 shows the estimated trajectories for a loop closure test. In this case we used the RF/GPS unit mounted on the robot, and compared the raw GPS (yellow) with position filer output (green) and visual odometry (blue). The total distance traveled was 75m.

Figure 5: Loop closure test: Yellow – GPS, Green – GPR + RF; Blue – Visual Odometry.

CombatID – Combat Identification System.

	Loop Closure Error (m)
GPS	7.63
GPS+RF	4.37
Visodo	0.26

Table 3: Loop closure error for localization modalities.

Table 3 compares the loop closure error estimated by raw GPS, position filter, and visual odometry. The filter result is better than GPS only results, but not as good as the visual

Friend/Foe Labeling

The Friend/Foe Labeling module takes as inputs the RF/GPS position and people detections from stereo and produces a list of associations.

The first step is to determine the camera orientation with respect to the GPS global frame of reference, to enable transforming the RF/GPS filter positions to the camera frame of reference. Each pair of human detection ROI and RF/GPS filter position provides a hypothesis for the rotation that brings the two frames into alignment. We select the hypothesis with the largest support (the rotation that produces the largest number of associations). In case when the robot is moving, the change of camera orientation is estimated by the Visual Odometry module and the initial camera orientation is updated accordingly.

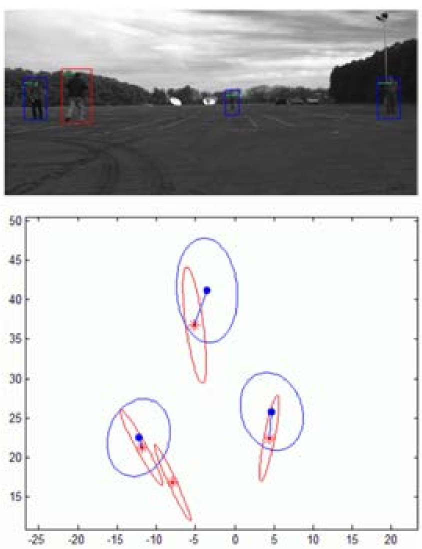

Figure 6: Friend/Foe labeling.

We then apply maximum bipartite matching to assign the Friend/Foe labels. The objective is to find the maximum number of matches between the two parties (positions of video based detections and positions from the RF/GPS filter) that minimizes the sum of Mahalanobis distances. We use the Kuhn-Munkres algorithm, also known as the Hungarian algorithm.

Figure 6 shows an association example. The top image recorded from the camera and contains four people. The colored boxes indicate detections: the blue boxes are the Friends, and the red box is a foe. The figure on the bottom shows people detection positions (red stars) on a map view and the RF/GPS filter outputs after transforming into the camera frame of reference (blue dots). The ellipses indicate their uncertainties. The blue lines show which pairs have been associated.

Display

The three different display types of the system (seen on the right of the system diagram in Figure 2) are shown in Figure 3. The image display (top) shows the image from the forward-looking camera with the classified detections overlaid on the image. Blue boxes correspond to detections classified as Friend, red boxes indicate Foes. In the example shown in the figure there are four people in the image, at about 40m in front of the camera: two Friends on the left and two Foes on the right.

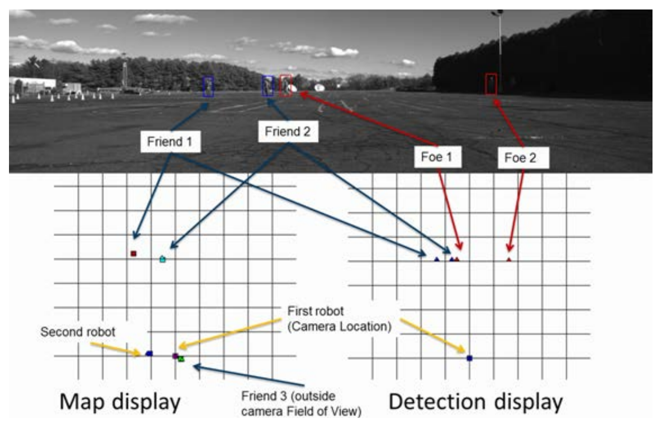

Figure 7: Display types for CombatID system.

The Map display (lower left in Figure 7) shows the location of the RF/GPS units carried by Friends on a grid (cell size is 10m in this example). Finally, the Detection display (lower right) shows the classified detections (blue triangles for Friends, red triangles for Foes) on a grid similar to the one in the Map display. The blue square on the bottom indicates the location of the robot which is the same as the location of the camera used for the image display.

SYSTEM PERFORMANCE EVALUATION

To demonstrate the systems functionality, the CombatID robots took a two phase approach. (1) The systems went through a set of engineering evaluation and tests (EET) to determine baseline performance of the system. (2) The systems where put into tactical scenarios in a relevant environment to determine the effectiveness of the system.

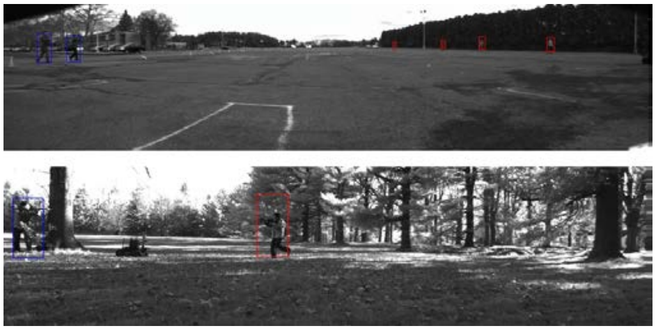

Figure 8: Sample screen captures for CombatID system.

Baseline Testing

Baseline testing for the system was performed with combination of Friends, Foes and vehicles at varying distances. The Friends (up to three) and Foes (up to six) were systematical tested in varying combinations moving in front of the robots at ranges from 10 to 100 meters. The Friends/Foes varied in speed and motion from a slow crawl to a fast sprint. Similar testing was then preformed with automobiles. One to three vehicles varying from parked to moving at 25 mph at ranges from 10 to 100 meters.

The EETs then became more complicated. Introducing various sets of Friends, Foes and vehicles in random patterns to try and find the failure point of the system. Figure 8 shows sample output from the baseline (top) and scenario (bottom) testing. In both cases, the personnel in the scene are detected and correctly classified (blue box for Friends, red box for Foes).

Scenarios

The experiment was conducted with three different scenarios which were parallel to the basics tactical operations used on some current Forward Operating Bases (FOB). The first Scenario was setup with friendly forces being dug into their fighting positions with two fixed Combat Identification robots monitoring the fields of fire. The concept was that the enemy could attack at any moment and the robots would have to identify if the personnel approaching the FOB were friendly forces or enemy forces before any friendly forces could engage the target.

The second scenario was identical to the first scenario with the exception that one of the Combat Identification robots could move across the field of fires in order to establish a better line of sight to Identify the targets as friendly or enemy threats. In the third scenario, the friendly forces conducted patrols from the FOB to a local village, upon returning from the mission the two fixed Combat Identification robots would have to identify the objects as friendly before access would be allowed into the FOB.

CONCLUSION

Since the beginning of time, technology has played a critical role in the way battles are fought and won. Leaders are always looking for ways to increase their available resources by eliminating tasks that are conducted by humans and having robots complete though tasks. The Combat ID system is one of those technologies. The system allows for a boarder field of view/line of sight and object movement detection then one single person can accomplish.

The CombatID program successfully showed that a unmanned robotic equipped with the CombatID payload could scan the same line of sight as a Solider. As Soldiers and commanders become more accustomed robots on the battlefield, the acceptance and utility of CombatID like capabilities will become combat multipliers for the operational commander.

REFERENCES

[1] Scaramuzza, D., Martinelli, A. and Siegwart, R., (2006). ”A Toolbox for Easy Calibrating Omnidirectional Cameras”, Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS 2006), Beijing China, October 7-15, 2006.

[2] Sizintsev, M. and Kuthirummal, S. and Samarasekera, S. and Kumar, R. and Sawhney, H.S. and Chaudry, A., ”GPU accelerated realtime stereo for augmented reality”, in Proceedings of the 5th International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT), 2010.

[3] Kira, Z. and Hadsell, R. and Salgian G. and Samarasekera S., ”Long-range Pedestrian Detection using Stereo and a Cascade of Convolutional Network Classifiers”, in Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS 2012), Algrave, Portugal 2012 – To appear.

[4] P. Dollar, C. Wojek, B. Schiele and P. Perona, ”Pedestrian Detection: An Evaluation of the State of the Art”, PAMI, 2011.

[5] Oskiper, T.; Zhiwei Zhu; Samarasekera, S.; Kumar, R., "Visual Odometry System Using Multiple Stereo Cameras and Inertial Measurement Unit," Computer Vision and Pattern Recognition, 2007. June 2007.

[6] Kuhn, H. W., "The Hungarian Method for the assignment problem", Naval Research Logistics Quarterly, **2**:83–97, 1955

[7] MA Ranzato, FJ Huang, YL Boureau, Y LeCun Unsupervised learning of invariant feature hierarchies with applications to object recognition, CVPR 2007.

CombatID – Combat Identification System.