

VISUAL-BASED ASSISTANCE FOR ELECTRIC VEHICLE DRIVING*

Michael Finnefrock, Xianhua Jiang, and Yuichi Motai, *Member, IEEE*
Department of Electrical and Computer Engineering
University of Vermont, Burlington, U.S.A.
{mfinnefr, xjiang, ymotai}@cem.uvm.edu

Abstract – This paper presents a unique assistance system for driving a vehicle. We wish to establish an ideal intelligent vehicle that detects obstructions through use of the mounted sensors, which provide sufficient visual-based aid. To this end, we utilize an infrared camera, a laser range finder, and a gyro sensor so that these sensors provide some assisted visual cues for a driver to identify unexpected obstructions. In this proposed vehicle, we develop several interesting modules. For example, in addition to a 2D heat-map provided by the infrared camera, the system provides depth to the map by utilizing the tracking results of the gyro-scope. The depth values in the region of interests are computed in a real-time video-frame rate by motion stereo from a single far-infrared camera, based on typical pinhole lens model. We also develop a graphical representation of a virtual camera so that the driver can see the field of view presented by the mounted sensor. The laser range sensor will compensate the camera by increasing the field of view and providing accuracy for detected objects. Indoor-based experimental results using a portable electric scooter demonstrate the feasibility and merits of this new sensory approach on the driving assistant system.

I. INTRODUCTION

Most advances in vehicle safety have provided protection to both a driver and passengers. Despite the increase in the number of standard safety features, automobiles are still unable to prevent accidents, nor protect those involved in the accident outside of the vehicle. Advances must be made to effectively reduce the number of pedestrian- and animal-related accidents [1]. There exist numerical people who really need some driving help [12]. Our initial target is a small electric scooter, by which a disable people can drive by their own operation. Until advances in autonomous vehicle control become widely accepted and trusted by the average driver, further assist devices must be used. These systems will not have control of the vehicle's trajectory; therefore, they must provide safety by providing early warning to the driver of dangerous obstacles. The goal is to provide the drivers with information that they cannot obtain by themselves [3].

A color-camera based system gives the drivers good visual feedback. Cameras provide an increased field of view and a natural visual feedback. However cameras can only produce a two-dimensional solution. In correcting this problem, binocular-vision systems are considered to be effective candidates [1]. There are two major drawbacks to

these systems. The first drawback is the extra overhead involved in computing correlations between left and right images, as well as the computations involved in extracting image patterns. The second drawback is the added expense of purchasing a second camera. Stereovision does provide a third dimension of data; allowing the driver to judge and respond accordingly.

Far-infrared cameras are able to detect radiated heat of pedestrians, other cars, or animals, yielding a similar visual feedback being supplied to the driver. Using far-infrared cameras allows objects to be detected in low lighting situations where color, and even black-white cameras are unable to produce clear images. Night driving is responsible for more than 50% of fatal accidents, yet less than a 50% of all driving occurs at night. An on-board vision system should be able to assist a driver during expressway and high speed driving as well as in urban conditions where the ability to detect pedestrians is essential [1].

Current pedestrian detection is based on localizing objects using pattern analysis based on a human shape [4]. Using infrared imaging, one can localize 'hot-spots' using relatively simple algorithms and relatively low computations. Once a hotspot has been found its pixel area can be localized and examined for further qualities. This second process will determine the object's classification; human, animal, or automobile. The object can then be examined for trajectory and possible collisions. A few companies, such as Honda (Legend model) and General Motors (Cadillac model), have already placed infrared devices on production vehicles. In order for these options to become popular they must show fascination for the customer as well as effectiveness to prevent the accident [1]. We think that infrared imaging and Heads Up Display (HUD) units will be the next standard and have enough glamour to attract potential drivers/users. All that is missing is functionality which enhances sensory perception of the environment.

We propose to create functionality by using a single infrared camera to provide the driver with pertinent information such as object/collision detection. Using only one infrared camera we will demonstrate a method for creating a three-dimensional data set similar to that of a stereovision apparatus using a gyro-sensor. Also we have developed a

* This work is supported by Vermont Experimental Program to Stimulate Competitive Research Program (EPS-0236976) and by URECA! Awards at the University of Vermont.

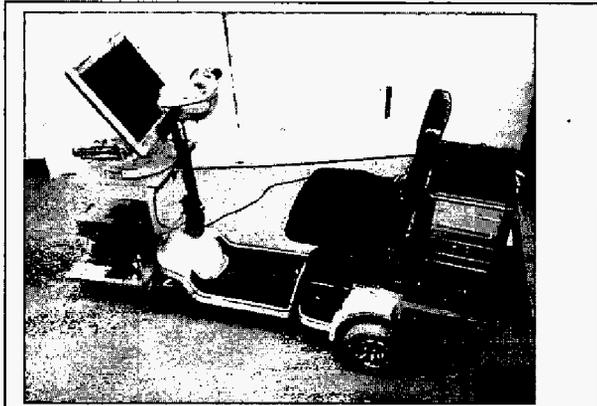


Fig. 1 Electric vehicle for experiments developed.

reliable form of driver-computer interface which, we believe, requires both auditory and visual aids.

In this paper, we will first illustrate our architecture of the sensory system mounted on an experimental electric vehicle in Section II. In Section III, we briefly explain the developed modules for detecting obstructions. We report some interesting experiment results in Section IV to demonstrate the performance of the overall sensory system. We will discuss interface issues in Section V. Finally, the conclusion is given in Section VI.

II. SENSORY MODULES FOR ELECTRIC VEHICLE

The system is comprised of three sensory modules; a far-infrared camera, a Laser Range Finder (LRF), and a gyroscope. Fig. 1 gives the overall outlook of these devices.

The overall purpose for the sensory modules is to detect obstructions. It is very useful to utilize a camera sensor by incorporating machine vision methodologies for driving assistance. Obviously, the lack of depth information between a camera and obstruction is core in our application. In fact, trying to reconstruct three-dimensional data from a two dimensional image has intrigued many people. In [5] and [13], the methods for determining depth based on one camera have been clearly laid out. These methods typically rely on depth from focus and require the object in question satisfy the property, where object distance is less than the focal point. In [6], they explored methods for reconstructing three-dimensional data of a particular cubic object using only a camera and gyroscope. We have expanded this idea into the real-time methodology. We have integrated a gyroscope and

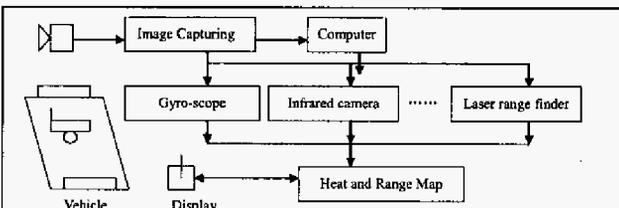


Fig. 2 Block diagram for proposed sensory architecture.

camera to construct three-dimensional data for the entire natural scene, based on two-dimensional images. Under 80 meters the LRF data can be correlated to the infrared camera's images for very reliable and accurate measurements to objects in the vehicle's trajectory. However, over 80 meters the LRF can no longer provide numerical results. In order to overcome this loss of data, moving stereovision techniques have been incorporated to give object-depth feedback to the collision warning system. In the rest of this section, we will describe module specifications of each sensor illustrated in Fig. 2.

A. Infrared Camera

An infrared camera (Raytheon thermal vision 4000B) is used to detect objects and pedestrians. The optical field of view (FOV) is 12 degrees. Focal length is 18 mm. This sensor forms 160*120 pixel images. The goal is to create a robust algorithm which extracts data from an image to alert the driver of dangerous situations. As of now there is an algorithm which detects 'hot-spots' and alerts the user to their location by outlining them with a box. Our simple tracking algorithm extracts 'hot-spots' based on threshold values. In Fig. 3, the white box is a hot-spot locator showing a hand has been found. The hand is hotter than half the objects in the room. The coordinates of the hand and the box's area is computed. This typical program is developed for our human tracking problem of autonomous navigation [11] with a few minor additions to the source codes in [10].

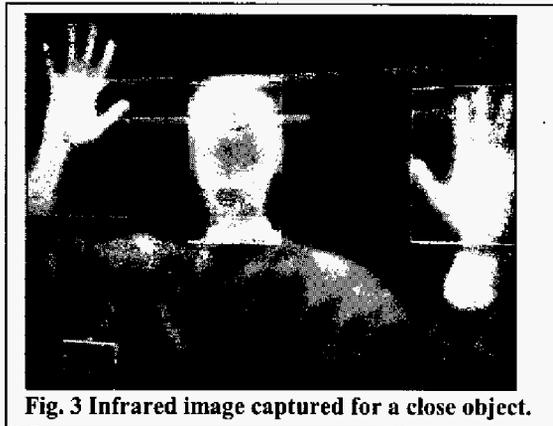


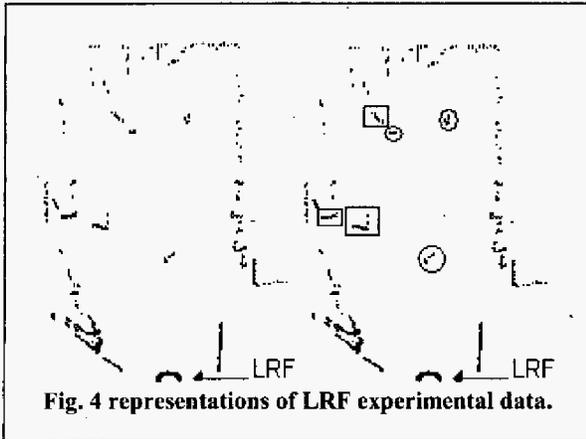
Fig. 3 Infrared image captured for a close object.

B. Range Finder

A laser scanner (SICK LMS200) is used to acquire the range data of a particular area, concentrates on particular area, and usually centered in the camera FOV. The scanner allows the following FOV options using the LMS 200:

- 1) 180 degree with 1 degree or 0.5 degree of resolution
- 2) 100 degree with 1, 0.5 or 0.25 degrees of resolution

There are two measurement settings used to determine distance. The first allows depth accuracy of 1 mm, but is limited to objects within 8.2 meters from LRF. The second setting allows objects up to 82 meters to be measured with 1 cm accuracy. Fig. 4 demonstrates the results of measuring the experimental room with the 1 mm accuracy setting. In the experimental section later in this paper, we used 100 degrees



of vision with the 1/4 degree of resolution. This higher resolution will result in 35 cm between measurements at 81 meter from the laser.

With the LRF working, one can create a 'map' of the surrounding area. One simply has to convert several radius values (aka distance from the laser) into cartesian coordinates. Refer to Fig. 4 for a visual example. Each values is converted from a distance value from source to a (x,y) pair and plotted. Hence, the pictures 'dotted' appearance. In Fig. 4, one can see the room's flat wall on the right side of the image. The back of the room and left side of the image are noisy because of the computer's set up for students. Students have been circled and monitors placed in boxes.

This LRF will support visual cues to enhance the infrared images as shown in Fig. 5. The color bar represents the LRF's data which overlaps the camera's FOV's value. This gives additional information to the far-infrared images. Many papers [9,12] have been written on the subject of fusing LRF data and camera data. The LRF allows up to 180 degrees of data to be analysed, far exceeding the infrared camera's 12 degree FOV. The LRF's wide FOV allows for enhanced city object detection where potential hazards might be out of the camera's FOV. The LRF also allows objects that do not radiate heat, for example a garbage container, to be detected.

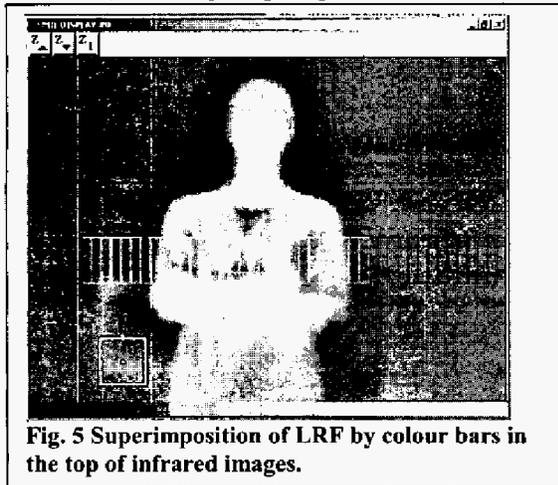


Fig. 5 Superimposition of LRF by colour bars in the top of infrared images.

We are using a simplified method for constructing maps of data based on [8]. If a collision warning, and/or avoidance system is to be effective, it must be aware of its surroundings. In [9] it is discussed how to interpret this data for collisions with pedestrians, curbs, and other vehicles. Using two previous data sets, and subtracting the velocity of the vehicle (found using the gyro) accurate velocity vectors of potential hazards can be calculated. A slight manipulation of the data allows the computer to warn a driver of hazards with a time-to-collision vector constantly being updated.

C. Gyro Scope

A gyro scope (MicroStrain 3DM-G Orientation Sensor) is used to measure location and orientation by 3-angular rate sensors, 1 tri-accelerometer, and a 3-axis magnetic field sensor. The sensor is mounted in the camera so that it can track and measure the location and orientation of the moving camera. Since the camera is moving, the images captured are used for stereo triangulation to estimate the depth of obstructions. This sensor includes an integrated signal processing that allows for orientation vectors to be queried from the sensor directly. It measures magnetic field orientation, angular acceleration around axis x,y,z , and

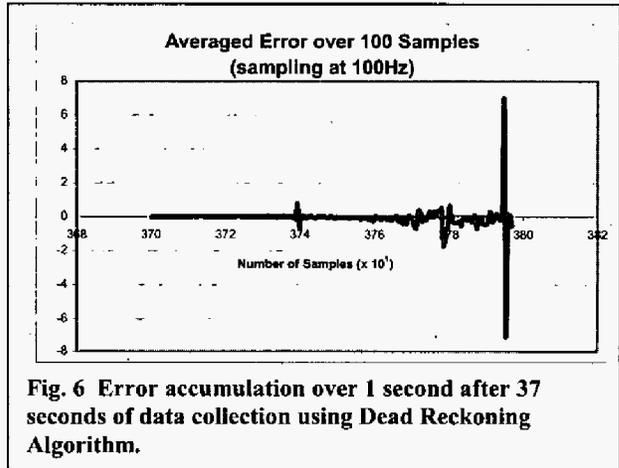
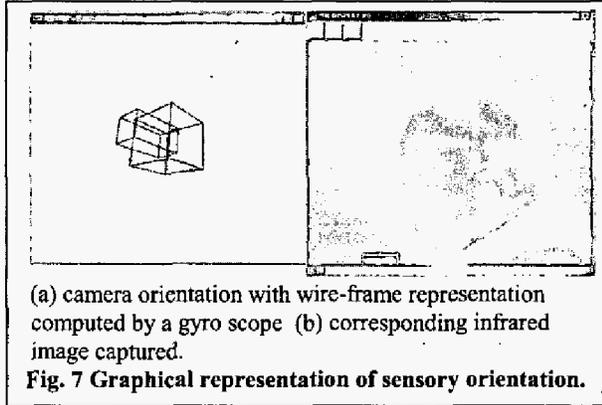


Fig. 6 Error accumulation over 1 second after 37 seconds of data collection using Dead Reckoning Algorithm.

acceleration along x,y,z .

We tried to reduce noises of Gyro Sensor in a dead reckoning algorithm [14], employing several additional filters (as shown in the result above Fig. 6). However, due to vehicle vibrations and constant velocities over extended periods of time, the errors accumulated rapidly. A simpler and seemingly more efficient way of determining camera position is used. Rather than trying to create a virtual map of where the camera has been, our new system simply relates the previous image to the current image with a vector. This memory-less system accumulates fewer errors simply because there is less reliance on previous data values. Vibrations still play a small part in our accuracy, however, applying a moving average filter of our calculated results significantly reduced errors. Fig. 7 (a) represents a graphic output of the sensory output in the 3D graphical space using Open-GL.



III. METHOD FOR COMPUTING DISTANCE

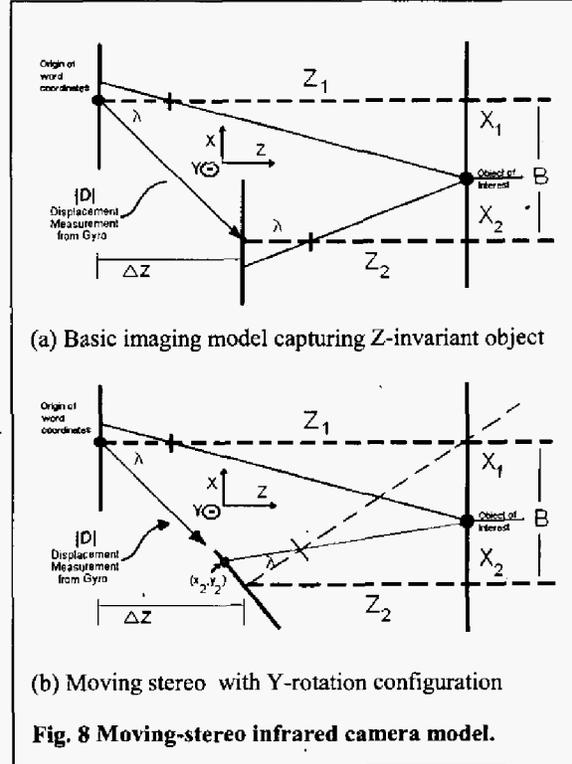
We have created stereo images by tracking the single camera positions using the mounted gyro scope. Among our contributions, in this section, we would like to describe in details in one specific module; depth estimation. This setting is very unique since the camera is moving and provides far-infrared and low-resolution images (160*120 pixel). Thus we are developing a straight forward moving stereo methodology to estimate depth value between a camera and objects by the following steps:

The first step in creating three-dimensional representations is to translate all image planes into the world coordinates. Fig. 8 (a) demonstrates this setup. It is advantageous to always set up dynamic world coordinates based on Image 1 (or the first of the two images being processed). This allows for quicker computations and fewer ambiguities based on angles and incoming data streams. It is pointed out in [7] that based on basic stereovision diagram using a pin-hole camera representation; a set of linear equations can be solved in terms of Z allowing one to find depth. In order to make this method more robust, it must first be adapted to handle a diverse range of values in more than one plane. The second step is to detect yaw, pitch, and roll followed by displacement of the image plane. This is done by reading the gyro's values. Using the kinematics equations, displacement can be determined from the gyros acceleration data. The gyro orientation can be found using Euclidean angles. These two data sets will give us the magnitude of the displacement from image 1 to image 2; this sets up without rotation movement of the single camera as shown in Fig. 8 (a). λ is focal length. From Fig. 8 (a), the following Eqs. (1) and (2) are defined. We wish to derive depth value Z_2 at the current camera position, if given camera rotation β , and translation ΔZ by the gyro-scope. D and B are the length shown in Fig. 8.

$$Z_1 = Z_2 + \Delta Z \quad (1)$$

$$B_x = \sqrt{(D)^2 - (\Delta Z)^2} \quad (2)$$

Next a third degree of freedom is added by allowing the image to vary based on rotations around the Y-axis. For simplicity, in



the explanation, we only compute one translation R_β from the image plane to the world coordinates by Eq. (3).

$$R_\beta = \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

We will explain at the end of this section how more translations can be added for increased accuracy. The Euclidean angles allow image 2 to be related to the world coordinate set by position 1. An image plane shift scales an image based on the camera focal length by Eqs. (4) and (5).

$$L_w = \begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\lambda} & 1 \end{bmatrix} \quad (5)$$

where L_w is linear movement of the world coordinates by (X_0, Y_0, Z_0) . P is the parameters from camera calibration

aligning image plane to a pinhole model. The resulting world point $W(X, Y, Z, 1)$ is transferred into the following coordinates T shown in Eq. (6).

$$T = R_\beta L_w W = \begin{bmatrix} (X_1 - X_0) \cos \beta - (Z_1 - Z_0) \sin \beta \\ Y \\ (X_1 - X_0) \sin \beta + (Z_1 - Z_0) \cos \beta \\ 1 \end{bmatrix} \quad (6)$$

Likewise, the resulting image point x_2 is mapped into the following coordinates shown in Eq. (7).

$$x_2 = \lambda \left[\frac{X_2 \cos \beta - Z_2 \sin \beta}{-X_2 \sin \beta - Z_2 \cos \beta + \lambda} \right] \quad (7)$$

Or in the corresponding world point X_2 is represented in the following coordinates shown in Eq. (8).

$$X_2 = \left[\frac{\lambda \cdot x_2 + \lambda \cdot Z_2 \sin \beta - Z_2 \cdot x_2 \cos \beta}{x_2 \sin \beta + \lambda \cos \beta} \right] \quad (8)$$

Eq. (8) is checked by the specific value in Eq. (9)

$$X_2|_{\beta=0} = \left[\frac{\lambda \cdot x_2 - Z_2 \cdot x_2}{\lambda} \right] = \frac{x_2}{\lambda} (\lambda - Z_2) \quad (9)$$

The image translation should also include a slight linear shift to center the new axis to the image plane. However, due to noise in the system this small adjustment can be ignored. If testing proves a calibration matrix is needed, then this translation can be added in.

Finally the following Eq. (10) is used to estimate depth Z_2 , the latest camera position, which is derived from Eq. (1) (7) and (8).

$$Z_2 = \frac{\beta[x_2 \sin \beta + \lambda \cos \beta] - \lambda \cdot x_2 + x_1[x_2 \sin \beta + \lambda \cos \beta] - \frac{x_1 \cdot \Delta Z}{\lambda} \cdot [x_2 \sin \beta + \lambda \cos \beta]}{\lambda \sin \beta - x_1 \cos \beta + \frac{x_1}{\lambda} \cdot [x_2 \sin \beta + \lambda \cos \beta]} \quad (10)$$

Eq. (10) means that if we have image coordinate x_2 , camera rotation β , and translation ΔZ , we can estimate the depth value Z_2 . If more than one axis has experienced a rotational displacement additional rotational vectors must be multiplied in above steps to make the new rotation equation R corresponding to Eq. (3). These rotational transforms for other axes are found in the textbook [7]. Note that this depth estimation Z_2 is not necessarily meant to be perfect, but to be fast. The amount of data processing that is occurring must not be slowed down by the image processing taking place. We will show the experimental result in the next section.

IV. EXPERIMENTAL RESULTS

We would like to show experimental results on depth map for the proposed vehicle, using an infrared camera, a LRF, and a gyro scope. All the sensory results are computed and measured using the electric scoter, which is an ideal moval testing platform (as shown in Fig. 1) as both in-door and out-

door driving tool. The Windows-based Pentium 4 (3GHz) PC (mini-barebone) is in the backbone of the vehicle. The PC has a PCI image grabber card (Matrox Meteor II) to the infrared camera and serial connections to the Gyro and LRF. The specifications of sensors are described in Section II. Fig. 9 shows the error evaluation of the distance between the object and camera. The error was computed by the difference between the estimated depth value Z_2 in Eq. (10) and LRF (as

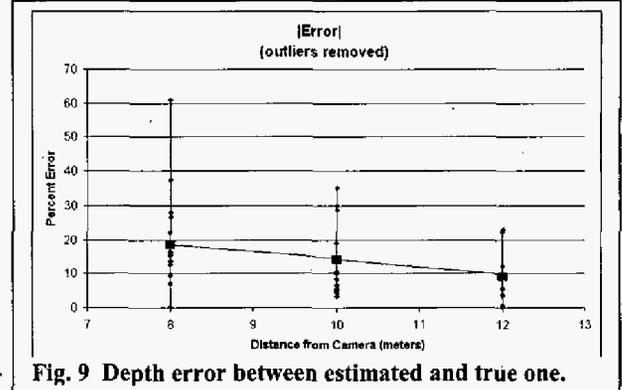


Fig. 9 Depth error between estimated and true one.
a ground truth). The sampled distances shown in Fig. 9 were chosen at 8, 10, and 12 meter, measuring 42 separate points. After excluding outliers, the averaged error values were computed in Table 1. We plan to compare our results of the depth estimation with another method such as [12, 5, 6, 13]. Also the 2D depth map of entire FOV is currently developing with processing of 2D space-constraints.

TABLE I
ACCURACY COMPARISONS WITH THREE OBJECT MEASUREMENTS

Distance (meter)	8.0	10.0	12.0	Total
Average Error (%)	17.5	12.8	8.6	15.3
Standard Deviation	14.6	11.4	8.8	12.7

V. DISCUSSION FOR FURTHER INTERFACES

We will also address the several issues on user-interface for aided driving. As illustrated in the diagram of Fig. 2, all sensory outputs are displayed in the head up monitor by an integrated approach. We started to evaluate the related issue of effectiveness of visual cues for the driver.

Graphical Display for Aided Driving

Correlating the data from the proposed sensory devices in a coherent fashion is of the utmost importance. Data presented to the driver must not divert the driver's attention from the driving tasks at hand [4]. The display should be kept simple and should allow the driver to glance at it from time to time as if it were another mirror. Through experimentation, we feel the heads-up-display captured shot shown is optimal. It fuses data from three different sensors into one subtle display. By finding the calibration vector needed to merge LRF data with the captured images from the infrared camera, we are able to present the driver with a color band shown in Fig. 10. This band quickly allows the driver to associate color to danger, and, in a way, distance. For example, red signifies close to

danger and therefore stop. Red represents close (less than 10 m), Yellow represents varying shades (around 10 and 50m), Green represents neutral (around 50m), and Blue represents safe (far away greater than 74m). This color band was found by synchronizing the LRF with the camera. The proper data points from the LRF are found through a calibrating vector created on vehicle startup.

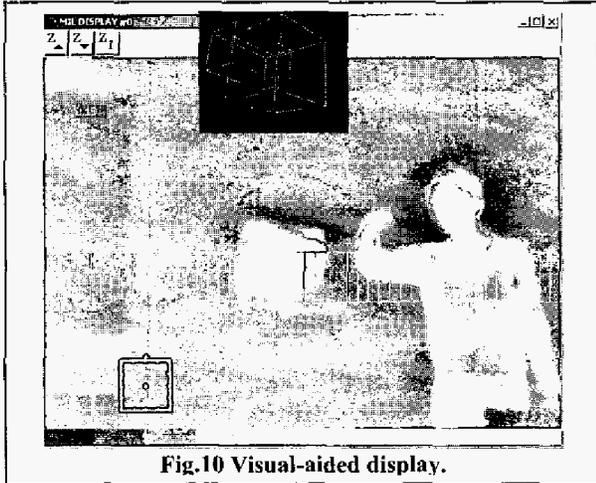


Fig.10 Visual-aided display.

Data fusion for useful driver feedback

As shown in Fig. 10, even if no hot spot exits in the left hand side, the LRF clearly demonstrates there is an obstruction with yellow bars; the sensory fusing compensates each other. As such, merging the data input from the LRF and IR camera allows for two things. One, the LRF inputs 100-180 degrees of range sensing input accurate up to 1 cm. The camera, for example, only views 12 degrees and cannot focus below 8 meters. Therefore the LRF provides the system much needed information for obstacle avoidance. In addition the camera only detects heated objects; it will not pick up many things detected by the LRF. Second, since the IR camera has a range of over 450 meters it far surpasses the range of the LRF. Taking advantage of this large range of sensor inputs allows for a collision detection unit able of proving valuable information (by auditory aids) at almost any speed.

High speed driving

During high speed driving (speeds in excess of 25 meters per second) reaction times are greatly reduced. Therefore, the driver must be made aware of these situations as early as possible. The LRF's data (at these speeds) would give a driver scarcely 3 seconds to react to a dangerous situation. The camera can detect an object's heat up to 450 meters away, giving a driver close to 15 seconds to decide an evasive course of action. Smaller objects, at close distances and large objects at far distances have similar image areas. Therefore for a collision detection and avoidance system, it is crucial to know the object's distance from the vehicle.

VI. CONCLUSION

In this paper, we have developed a new visual-based assistance for driving a vehicle. Particularly, we developed a small electric scooter, by which a disable people (or who really need some driving help) can drive by their own operation. Our proposed sensory system is an integrated approach, by which the number of pedestrian fatalities caused by traffic accidents might be effectively reduced. The laser range finder covered a wide range of the FOV for detecting obstructs. The far-infrared image sensory provided a heat map. The gyroscope tracked the accurate movements (orientation and translation) of the infrared camera to estimate depth maps for some regions of the FOV. We have merged for displaying the overlapped images between a laser range finder and an infrared heat map. The experimental in-door results demonstrated the feasibility and merits of this visual-based driving assistance system for further efficient driving.

REFERENCES

- [1] Furukawa, Y. "Status and Future Direction of Intelligent Drive Assist Technology." Proceedings of IEEE Intelligent Transportation Systems, pp. 113-118, 2000.
- [2] Cherfaoui, Veronique and Burie, Jean-Cristophe. "Dealing with uncertainty in perception system for the characterization of driving situation." Proceedings of IEEE Intelligent Transportation System, pp. 83-88, 2000.
- [3] Yamaguchi, Toru, Hiroshi Nitta, Tomohiro Takagi and Hideki Hashimoto. "Human Centered Sensory Intelligence Architecture for ITS", Proceedings of IEEE Intelligent Transportation System, pp. 446-451, 2000.
- [4] A. Broggi, M. Bertozzi, A. Fascioli, M. Sechi. "Shape-based Pedestrian Detection", Proceedings of IEEE Intelligent Vehicles Symposium, pp. 215-220, 2000.
- [5] A. Pentland, T. Darrell, "A simple, Real-Time Range Camera". Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 258-261, 1989.
- [6] T. Mukai, N. Ohnishi. "The recovery of object shape and camera motion using a sensing system with a video camera and a gyro sensor". Information Fusion, pp. 45-53, 2000.
- [7] K.S. Fu, R.C. Gonzalez, and C.S.G. Lee, "Robotics: Control, Sensing, Vision", and Intelligence, McGraw-Hill, 1987.
- [8] R. Mazl, L. Preucil. "Building a 2d Environment Map from Laser Range-Finder Data", Proceedings of IEEE Intelligent Vehicles Symposium, pp. 290 - 295, 2000.
- [9] A. Kirchner, C. Ameling, "Integrated Obstacle and Road Tracking using a Laser Scanner", Proceedings of IEEE Intelligent Vehicle Symposium, pp. 675 - 681, 2000.
- [10] Source codes in the download site (On-line). <http://www.emba.uvm.edu/~medialab/Fall04/ee214/ElecChair/links.html>
- [11] R. Ketcham and J. King, "Robotic Tracking with Infrared Vision", Undergraduate web-publishing at the University of Vermont (On-line). <http://www.emba.uvm.edu/~medialab/Fall04/ee187/RobotVehicle/>
- [12] Y. L. Murphey, J. Chen, J. Crossman, J. Zhang, P. Richardson, L. Sieh, "DepthFinder: a real-time depth detection system for aided driving, Proceedings of IEEE Intelligent Vehicle Symposium, pp. 122-127, 2000.
- [13] J. Ens and P. Lawrence. "An Investigation of Methods for Determining Depth from Focus, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No.3, pp. 97-108, 1993.
- [14] C. Randell, C. Djalllis and H. L. Muller, Personal Position Measurement Using Dead Reckoning. Proceedings of International Symposium on Wearable Computers, pp. 166--173, 2003.