ADVANCED SITUATIONAL AWARENESS AND OBSTACLE DETECTION USING A MONOCULAR CAMERA

Abhijit Bhoite, Nikhil Beke, Sashank Nanduri, Tim Duffy, Myra Torres.

Impact Technologies, LLC, 300 Canal View Boulevard, Rochester, NY 14623, United States

ABSTRACT

This paper presents a modular approach for a high resolution monocular camera based system to detect, track, and display potential obstacles and navigational threats to soldiers and operators for manned and unmanned ground vehicles. This approach enhances situational awareness by integrating obstacle detection and motion tracking algorithms with virtual pan-zoom-tilt (VPZT) techniques, enabling soldiers to interactively view an arbitrary region of interest (ROI) at the highest captured resolution. Depth determination from a single imager is challenging and an approach for depth information, along with size and motion information is developed to assign a threat level to each obstacle.

***Index Terms* - Obstacle detection, Motion tracking, Depth perception, VPZT**.

1. INTRODUCTION

Passive sensors, such as high-resolution monocular cameras, have become a key technology that soldiers rely on in the field due to their simplicity and relatively low cost and weight [1]. However, the application of these high-resolution imaging technologies upon military ground vehicles is hindered by the limited resolution of their display panels and the bandwidth of their content delivery systems. To overcome this problem, the high resolution video is down sampled to match the resolution of the display screen. Thus, the soldiers are unable to view a local region of interest (ROI) at the "highest captured resolution" [2]. One solution to this problem is a pan-zoom-tilt (PZT) camera system, which allows for camera movement and zoom. For field applications, this solution is unsuitable as it introduces unnecessary complexities and numerous points of failure. A fixed camera solution solves the same problem with the use of virtual pan–zoom-tilt algorithms (VPZT). VPZT algorithms allow soldiers to pan, tilt, and zoom through a high-resolution video feed in real-time without mechanical components, enabling them to view any arbitrary ROI at the highest captured resolution.

In some situations, soldiers do not have any direct visual access to the environment in which the vehicle is being operated. Under such circumstances of total reliance on indirect sensory input, early detection of obstacles and distant navigational threats are of critical importance. Many algorithms have been proposed in the past to detect obstacles on a moving platform [1], [3], [4]. However, there no known technology that can integrate VPZT algorithms with obstacle detection and threat assessment to increase the situational awareness of the soldier [2].

The paper presents a modular system approach to improve situational awareness in the field. As seen in Figure 1, the process starts with image capture using a camera system that can resolve obstacles at a distance of 200-500m. Environmental effects present many constraints to object classification. Fog and rain reduce the visibility and range of the surroundings due to absorption and scattering of natural or artificial illumination by fog particles. The captured image is thus degraded, losing its contrast and color fidelity. An image quality enhancement approach comprising of histogram equalization and contrast enhancement addresses and overcomes the image degradation due to environmental factors. The post-processed image is then evaluated for road and obstacle detection in real time.

Once obstacles are identified, the region encompassing the obstacles is input to the motion tracking algorithms. The captured image sequence is used to distinguish stationary and moving obstacles. The relative distance of the obstacle from the vehicle is determined using a depth perception algorithm discussed in section 2.4. All stationary and moving obstacles are continuously analyzed for their threat level with respect to navigation of the vehicle depending on their size, speed, direction of motion, and distance from the vehicle. The key challenge is the ability to detect depth using monocular video. Most conventional approaches for depth perception use stereo vision or range based sensors. This approach for depth perception relies on optical geometry of the camera and motion detection to present relative distance.

2. TECHNICAL APPROACH

2.1 Obstacle detection

Obstacles in the context of navigation are defined as stationary or moving objects that may impede the motion of the vehicle. All pixels in the image are classified as belonging to either an obstacle or background. Any pixel that visually differs from the background is considered an obstacle. An edge based segmentation technique for obstacle detection emphasizes enclosed boundaries corresponding to the obstacle while neglecting spurious edges due to shadows and occlusions. Figure 2 details the edge based segmentation

978-1-4244-9300-5/10/$26.00 ©2010 IEEE

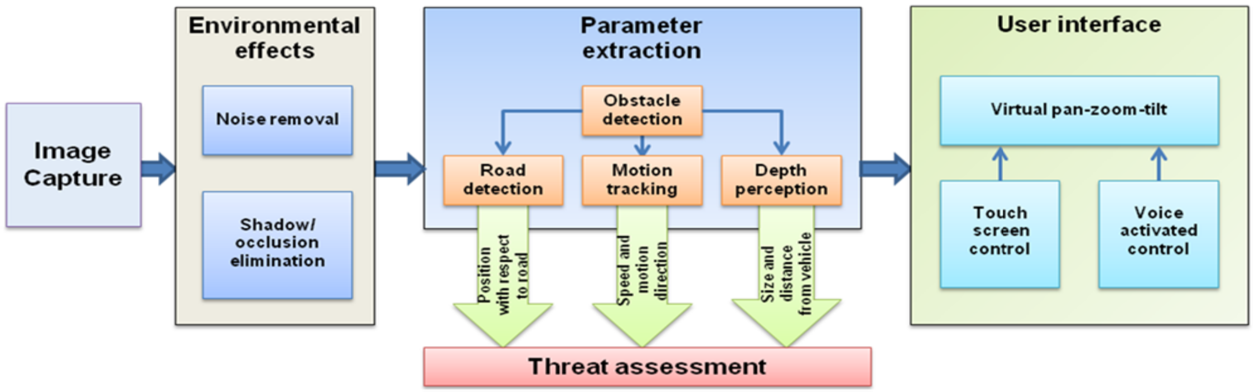

Figure 1: System block diagram for high resolution monocular vision system awareness

technique.

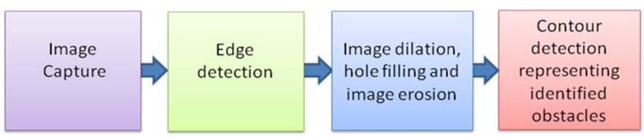

Figure 2: Edge-based obstacle detection technique

An edge detection filter, such as the canny filter, is applied to the image which, after thresholding, gives a binary image representing all edges in the image. An image dilation filter is applied to the binary image (following hole filling) to fill only the enclosed boundaries in the image. This step removes unnecessary edges, giving only the dominant edges corresponding to the obstacles. An erosion filter is applied to further remove the unwanted blobs in the image, followed by contour detection, which gives a representation of obstacles in the form of white pixels. The method is suitable for real-time applications as it uses very basic image processing kernels. Figure 3 shows results of the described technique.

(a) (b)

Figure 3: Representation of identified obstacles in red (images taken from IEEE Performance Evaluation of Tracking and Surveillance (PETS) 2001 image sequence)

2.2 Road detection

For threat assessment, it is necessary to distinguish obstacles on road from obstacles on the side. Road detection is accomplished using the histogram comparison method [3]. The image in Red, Green, and Blue (RGB) color space is converted to Hue, Saturation, and Intensity (HSI) space. Hue, which gives the color component in the image, is used to process the image. Considering only hue makes processing faster and more efficient while making the implementation independent of the light variations. A trapezoidal area in the bottom middle section of the image is taken as a reference area for the road. Here, the ground plane is considered to have constant color distribution. Since the bottom middle trapezoidal area is nearest, directly in front of the vehicle, it is safe to assume that it does not contain any obstacle. A histogram of the hue values of the trapezoidal area is plotted. Each hue value in the image is compared with the histogram. If the histogram value corresponding to the hue value is greater than a particular threshold, that pixel is considered as belonging to the road. The result of the classification approach can be seen in Figure 4.

Figure 4: Representation of identified road in white

2.3 Motion tracking

The motion tracking block determines the speed and direction of obstacles within the sequences captured by the imaging system. Obstacle boundaries and a contour output from the obstacle detection algorithms are input to the motion tracking block. The output of the tracking block is a set of parameters to be used for navigational threat assessment. Motion tracking is performed in three steps: optical flow calculation, moving/stationary obstacle classification, and calculation of navigation parameters.

Optical flow algorithms approximate the apparent motion of obstacles within an image sequence [4][5][6]. These algorithms produce a two dimensional projection of motion on an image of a three dimensional motion field. For a given image sequence, flow vectors are generally calculated on a pixel by pixel basis. Conventional approaches use these algorithms applied directly to the original image sequence. In this work, the optical flow algorithms are applied only on the post-threshold object and the delineated image sequences produced by the object detection block. The density and accuracy of flow vectors calculated is improved for the objects in the image sequence. Implementation of flow vector calculation using traditional processors and digital signal processors is hindered by the computational complexity of these algorithms. In recent years, FPGA technology has successfully implemented these algorithms in real-time [7][10][11]. In addition, based on the performance requirements, the number of frames (per second) used for flow calculation can be adjusted to facilitate real-time implementation. Figure 5 shows an illustration of optical flow vectors calculated using the Lucas-Kanade method [4].

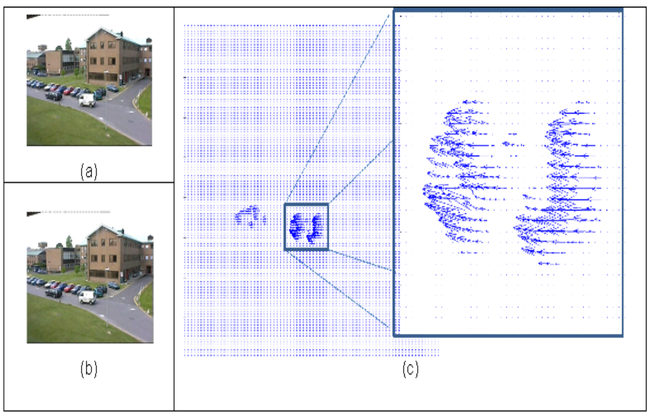

Figure 5: (a), (b) Image sequence of moving objects (c) Zoomed representation of identified moving obstacles on quiver plot

The optical flow algorithms produce a flow map of all the moving obstacles within the image sequence. Therefore, if the imaging system is mounted on a moving platform, stationary obstacles will also appear to be moving due to the effect of relative velocity. Data regarding camera motion and orientation can be accessed via the integrated vehicle bus system. This information is then used to calculate a baseline flow map that can be subtracted from the optical flow block output.

The output from the moving/stationary object detection block is a set of absolute flow vectors associated with each delineated object that has been detected. The mean magnitude of the vectors is used to calculate speed of the obstacles and the net orientation of the flow vectors can be used to calculate the direction of motion of obstacles [10].

2.4 Depth perception

For indirect driving, depth perception is critical for vehicle navigation. Stereo vision is a widely used technique for depth perception in which images are captured from two identical cameras with fixed position [8]. Depth is determined by establishing a relationship between the two images. The challenge in a monocular camera approach is to use a single imager for depth perception. The advantage of a monocular system is the decreased usage of channel bandwidth and cost as compared to stereo system. A technique developed by Murphy is used to achieve depth perception using monocular vision [9]. The system makes use of the motion of the vehicle. The obstacle is located in two different frames placed a few milliseconds apart in time. The contours of the obstacles detected using the obstacle detection approach explained earlier can be matched to localize the same obstacle in the two different images. The distance of the obstacle from the camera can then be found using the geometrical properties of the camera.

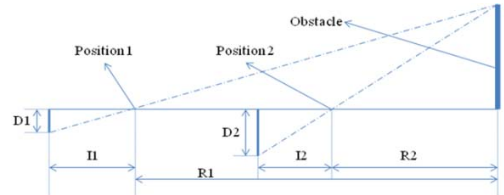

Figure 6: Camera geometry for obstacle depth perception

With reference to Figure 6, the distance $R2$ of the obstacle from the vehicle is calculated by Equation 1

$$R2 = (I/(D2 - D1))^{*}L$$

Equation 1: Determination of distance of the obstacle from the vehicle

where:

$I = I2 - I1$ is the disparity between the distances between the lens and the image plane at moment $t2$ and $t1$, respectively;

$D2 - D1$ is the disparity between the heights of the obstacle in images taken at moment $t2$ and $t1$, respectively; and

$L = R1 - R2$ is the distance that the vehicle has moved during the time period that the two image frames were captured.

With the distance between the obstacle and camera known, the size S of the obstacle is determined by Equation 2.

$$S = D2 * R2/I2$$

Equation 2: Determination of size of the obstacle

Thus, the size and the distance of the obstacle can be calculated using only monocular vision.

2.5 Threat assessment

Each obstacle is prioritized with a threat level that indicates the threat from that obstacle to the navigation of the vehicle. The threat level is calculated according to weighted sum of obstacle size, speed, direction of motion, and distance from the camera which are calculated from the methods explained in Section 2. The other major consideration in determining the threat of the obstacle is its relative position from the vehicle. As an example, obstacles along the path of the vehicle and obstacles on the road will have the highest threat level, whereas stationary obstacles on the roadsides will have lowest threat level. However, a moving obstacle detected off the road yet approaching the path of the vehicle will have a higher threat level. Thus, moving obstacles will always have higher threat level than those that are stationary.

2.6 Virtual pan-zoom-tilt

A VPZT system with a high-resolution imager and intelligent user interface will allow for both high speed PZT as well as the appropriate level of fidelity at high zoom levels. Traditional PZT systems utilize moving parts and moving lens assemblies to move the camera to a POI as well as provide the ability to zoom in on the POI. This system allows for a wide range of vertical and horizontal motion, as well as high levels of zoom while maintaining resolution. The ability to zoom mechanically allows for the system to have a lower resolution imager; however, at low zoom levels such as 1:1, the image fidelity suffers. A solution to maximizing fidelity at low zoom levels while maintaining appropriate resolution at high zoom levels is a VPZT system.

With commercially available cameras at high-resolution formats greater then 1920x1080 pixels, there are a plethora of hardware platforms that can be used to host a VPZT system. With a target resolution of 1920x1080, commonly referred to 1080P video, effective zoom levels of over 6x can be achieved with no image processing or decimation. Even higher zoom levels can be achieved through digital zoom techniques and interpolation.

The use of standardized compression techniques reduces both system cost, as well as bandwidth requirements for processing. MJPEG has been chosen as the standard for transport of image data from the imaging system to the real-time processing device. This method of transport holds benefits over other techniques such as MPEG due to its separation of dependence of frame-to-frame data. The system will call for breaking up the image stream into different portions for processing, which can be done on the fly and quickly when in JPEG2000 format. A distributed processing architecture is used to break the image into overlapping portions for near real-time performance.

2.7 User interface

The user interface of the system must be simple (minimum commands) yet powerful, allowing for all operations of the device to be executed by a single operator while driving or operating another piece of equipment. The interface input of choice is a touch screen, with the option of using actual buttons as well. With the use of a resistive touch screen technology, the touch surface will work even when the user is wearing gloves. The interface will be limited to a single point of contact; however, gained functionally from being able to use the interface with covered or dirty hands outweighs the benefits of a multi-touch environment.

Figure 7: VPZT enabled user interface

The system (Figure 7) provides feedback to the user to allow for multiple view levels of the entire video stream or an identified ROI. ROI will be identified by the object detection and movement detection algorithms explained in Section 2. Objects within each ROI will be identified with a simple box as well as information such as approximate size, distance, and direction. These metrics are presented to the operator to allow for operator-assisted threat identification.

3. CONCLUSION

This paper presents a real time, cost effective, modular solution to increase the situational awareness of soldiers in the field by using the monocular camera to its fullest potential. Threat prioritization in regards to vehicle navigation is achieved using features extracted from obstacle detection, motion tracking, and depth perception approaches. Significant in this approach is the ability to provide high-resolution contextual information that is useable to the soldier. With commercial off-the-shelf computing platforms increasing in processing power and decreasing in cost, the proposed system is possible to implement in low-cost hardware.

ACKNOWLEDGEMENT

The authors thank Patrick Kalgren, Mike Roemer, Kevin McCormick and Jessie O'Connell at Impact Technologies for their guidance and support.

REFERENCES

[1] Rostislav Goroshin, “Obstacle Detection Using Monocular Camera,” Thesis presented to the Academic Faculty of Georgia Institute of Technology, August 2008.

[2] DOD SBIR A10-072, “Obstacle Detection and Situational Awareness via High Resolution Monocular Video.”

[3] Iwan Ulrich and Illah Nourbakhsh, “Appearance-Based Obstacle Detection with Monocular Color Vision,” *AAAI Conference on Artificial Intelligence*, Austin, TX July/August 2000.

[4] M. Grunewald, “A Resource-Efficient Approach to Obstacle Avoidance via Optical Flow,” *Systems and Circuit Technology*, Heinz Nixdorf Institute, University of Paderborn, Germany.

[5] D.J. Fleet and K. Langley, “Recursive Filters for Optical Flow,” *IEEE Trans. Pattern Anal. Mach. Intel.*, Vol. 17, No. 1, pp. 61–67, Jan. 1995.

[6] H.C. Liu, T.S. Hong, M. Herman, T. Camus, and R. Chellappa, “Accuracy vs. Efficiency Trade-Offs in Optical Flow Algorithms,” *Computer Vision .Image Understanding*, Vol. 72, No. Issue 3, pp. 271–286, Dec. 1998.

[7] J. Díaz, E. Ros, F. Pelayo, E.M. Ortigosa, S. Mota, “FPGA-Based Real-Time Optical-Flow System,” *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 16, No. 2, pp. 274-279, Feb 2006.

[8] Juyang Weng, Thomas S. Huang, and Narendra Ahuja, *Motion and Structure from Image Sequences*, Springer-Verlag, 1993.

[9] Y.L. Murphey, J. Chen, J. Crossman, and J. Zhang, “A Real-time Depth Detection System using Monocular Vision,” SSGRR Conference, 2000.

[10] Z. Wei, D.-J. Lee, and B. E. Nelson, “FPGA-based real-time optical flow algorithm design and implementation,” Journal of Multimedia, vol. 2, no. 5, pp. 38–45, 2007.

[11] J. L. Martín, A. Zuloaga, C. Cuadrado, J. Lázaro, U. Bidarte, “Hardware implementation of optical flow constraint equation using FPGAs”, Computer Vision and Image Understanding, Vol. 98, pp. 462-490, 2005.