IN-VEHICLE PEDESTRIAN DETECTION USING STEREO VISION TECHNOLOGY

Wei Zhang, Ph.D., P.E.

Highway Research Engineer, Office of Safety Research & Development, HRDS-10
Federal Highway Administration
6300 Georgetown Pike, McLean, VA 22101, USA, e-mail: wei.zhang@dot.gov

Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA

ABSTRACT

In the last 3 decades, between 4378 (in 2008) and 8090 (in 1979) pedestrians were killed each year in motor vehicle related crashes, representing 11% to 17% of the total roadway fatalities. Although the numbers of pedestrian deaths have been in decline steadily since 1980, their distributions have become more and more concentrated in urban areas. There is an urgent need to develop reliable pedestrian detection technologies that can warn drivers in time to take corrective actions to avoid collision with pedestrians. This paper presents the research findings of the development of an in-vehicle pedestrian detection system using stereo vision technology. Stereo vision images contain both color and depth (distance) information of each pixel, giving researchers the option to implement more efficient filtering algorithms to quickly reduce the regions of interest (ROI). The developed system consists of one pair of stereo vision cameras, one vision accelerator, and one Dual Quad-Core computer. A layered approach was implemented to systematically remove irrelevant pixel regions, reject non-pedestrian objects, and then using pattern matching techniques to identify and track pedestrian like objects. The developed system can recognize pedestrian like objects, and other objects such as ground, vehicles, buildings, trees, and tall structures. It was tested under day light and twilight conditions. The completed system can process videos at 7 to 10 Hz rates, detect pedestrian like objects up to 30 meters away while driving at speed of up to 35 mph, and achieved a 90% overall positive detection rate. The project was funded under the FHWA Exploratory Advanced research Program.

Keywords: stereo vision, 3D vision, pedestrian safety, pedestrian detection, vertical support histogram.

INTRODUCTION

Traffic related pedestrian deaths from 1975 to 2009 are shown in Figure 1 (Insurance Institute for Highway Safety, 2009 Fatality Facts: Pedestrians). It reached a peak of 8,000 in 1980 and has been on the decline in general. Figure 2 shows the percent of pedestrian deaths in urban and rural areas during the same period. One trend revealed was that as the urban areas expand and urban populations increase over the years, the proportion of pedestrian deaths in urban areas increased. In 2009, 71% of pedestrian deaths occurred in urban areas versus 28% pedestrian deaths in rural area.

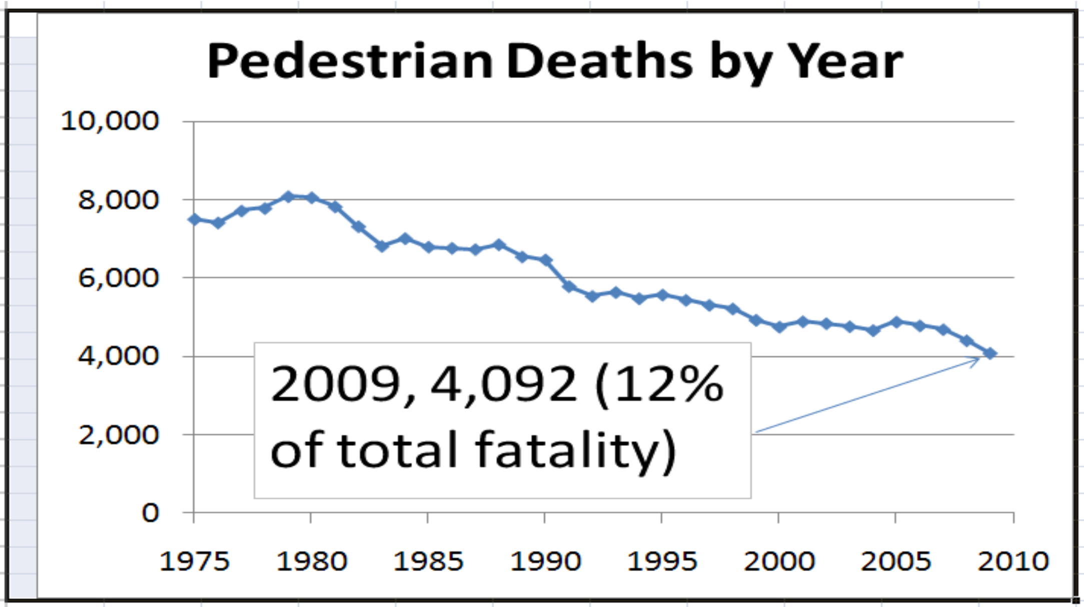


Figure 1 Pedestrian deaths in traffic accidents from 1975 to 2009

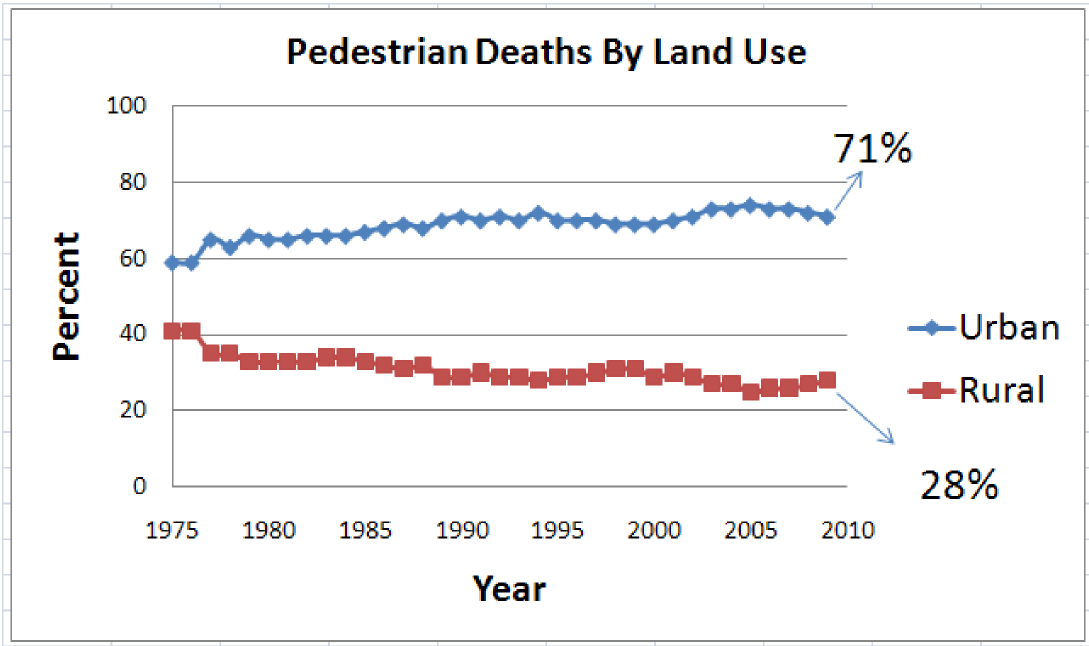


Figure 2 Percent of pedestrian deaths by land use from 1975 to 2009

According to statistics provided by the Insurance Institute for Highway Safety (IIHS), in 2009, 4,092 Pedestrian were killed in traffic crashes, most fatalities occurred:

In urban areas (71%)
At non-intersection locations (74%)
Under normal weather conditions (90%)
Under low to poor visibility from 6:00 p.m. to 6:00 a.m. (66%)
On non-freeway types of roads (84%)
On lower speed roads with speed limits less than 40 mph (54% urban, 46% rural)
Half of the pedestrians killed after dark had blood alcohol concentration (BAC) >=0.08

Detailed statistics of traffic related pedestrian deaths can be found at the IIHS Web site - Fatality Facts: Pedestrians: http://www.iihs.org/research/topics/pedestrians.html.

The above pedestrian fatality data shows the urgent need to implement policies and technologies to improve pedestrian safety. The conditions under which most pedestrian deaths occurred largely specified the performance requirements a technology product must possess for real world applications. Obviously, traffic accident related pedestrian fatalities and injuries are serious social issues, and the needs of developing vehicle based or infrastructure based real-time pedestrian detection and warning systems have existed since the 1970s or earlier. Only recently, the explosive breakthroughs in computing, machine vision and relevant sensing technologies, coupled with their ever decreasing costs make it possible and affordable to develop and mass market this type of products.

In September 2007, FHWA elected to fund a proposal titled "Layered Objective Recognition System for Pedestrian Collision Sensing" under the Exploratory Advanced Research program. The contract was awarded to Sarnoff Cooperation, which became part of SRI International in January 2011.

PERFORMANCE REQUIREMENTS OF REAL TIME IN-VEHICLE PEDESTRIAN DETECTION SYSTEM

According to the Virginia driver's manual, the relationship between driving speed and stopping distance is shown in Table 1.

Table 1 Average Stopping Distance on Dry, Level Road
(Source: Code of Virginia Section 46.2-880)

Speed MPH	Stopping Distance	
	Feet	Meters
25	85	26

35	135	41
45	**195**	**59**
55	265	81
65	344	105

This table, derived from extensive real life experience, defines the requirements of real time detection distances under various driving speeds in order to avoid collision with the detected subject. Looking at the 2009 pedestrian fatality statistical summary and the driving speed versus stopping distance table, it comes naturally that for any such system to work in real life environment, the following requirements must be met:

1. Can detect pedestrians while driving up to 45 mph
2. Can detect pedestrians up to 60 m away
3. Can detect pedestrians under day and evening visibility conditions
4. Can detect pedestrians under urban and rural settings
5. Can detect pedestrians at intersection and non-intersection locations

Upon satisfying requirements 1 to 5, the system must be trained and optimized to detect pedestrians under certain common scenarios, such as

a. Pedestrian darts out in mid block to cross the street
b. Pedestrian appears suddenly from behind a parked vehicle
c. Vehicle 1 stops for a crossing pedestrian and blocks the view of Vehicle 2 that is passing or overtaking vehicle 1
d. Vehicle is making a left or right turn while a pedestrian is crossing the same intersection
e. ….

The above requirements are not demanding for the human eyes, but are very challenging for automated pedestrian detection and warning systems, at least currently. We are stilling facing limitations on sensing technologies, computing power, and the efficiency of pedestrian detection algorithms. To our knowledge, no known system in the market has achieved the practical performance requirements outlined above. This research is one of the many efforts made to advance the state of the art in this field.

RELATED WORK

Dalal and Triggs (2005) proposed the Histogram of Oriented Gradients (HoG) algorithms for pedestrian detection. They characterized pedestrian regions in an image using HoG descriptors, which are a variant of the well-known SIFT descriptor. They reported significantly better results compared to previous approaches based on wavelets and PCA-SIFT, of around 90% correct pedestrian detection at 10^{-4} false positives per number of windows (FPPW) evaluated. A false alarm rate of 10^{-4} FPPW corresponds to about 0.4 false positives per frame (FPPF).

Tuzel, et.al (2007) proposed the covariance descriptor to characterize global image regions and used learning on a Riemannian manifold for the pedestrian detection. They reported improved results compared to the HoG descriptor, of about 93.2% correct detection, at the same false alarm rate of 10^{-4} FPPW.

Leibe et al. (2007) describe a stereo based system for 3D dynamic scene analysis from a moving platform, which integrates sparse 3D structure estimation with multi-cue image based descriptors (shape context} computed at Harris-Laplace and DoG features to detect pedestrians. The authors showed that the use of sparse 3D structure significantly improved the performance of pedestrian detection. Still, the best performance cited was 40% probability of detection at 1.65 false positives per frame.

Gavrila and Munder (2007) propose PROTECTOR, a real-time stereo system for pedestrian detection and tracking. PROTECTOR employs sparse stereo and temporal consistency to increase the reliability and to mitigate misses. The authors reported 71% pedestrian detection performance at 0.1 false alarms per frame without using a temporal constraint with pedestrians located less than 25 meters from the cameras. However, the datasets used were from relatively sparse, uncluttered environments.

RESEARCH AND DEVELOPMENT APPROACH

In this research, one pair of Elmo NTSC stereo vision (3D) sensors, one Acadia I 3D Vision Accelerator, and one dual quad-core computer were used to develop the in-vehicle pedestrian detection system. Figure 3 shows the major components. This approach offers the following distinctive advantages:

- 3D vision sensors simulate human eyes to provide both color and depth information of each pixel in the field of view. Depth information allows researchers to develop efficient algorithms to quickly extract out pixels within a defined distance, and identify objects of interest from noisy visual background.
- 3D vision accelerator provides a board-level solution for stereo disparity and depth, relieving the computer CPU from this demanding task.
- The dual quad-core computer offers a cheap and flexible platform for development, testing and demonstration of detection, classification, and tracking.

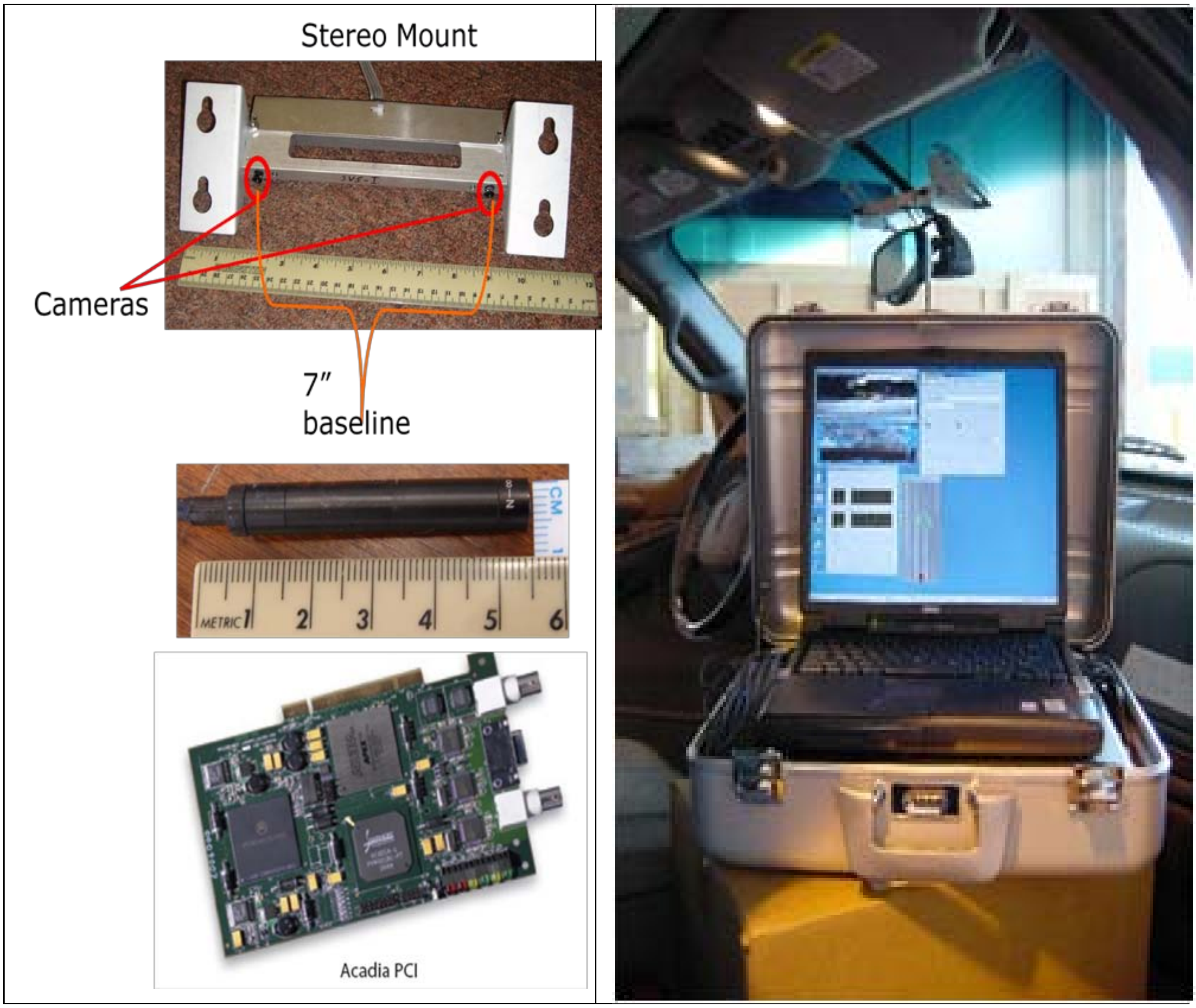


Figure 3 Elmo NTSC Stereo Vision (3D) sensors, Acadia I Vision accelerator, Dual Quad-Core Computer

THEORY AND LIMITATIONS OF STEREO VISION

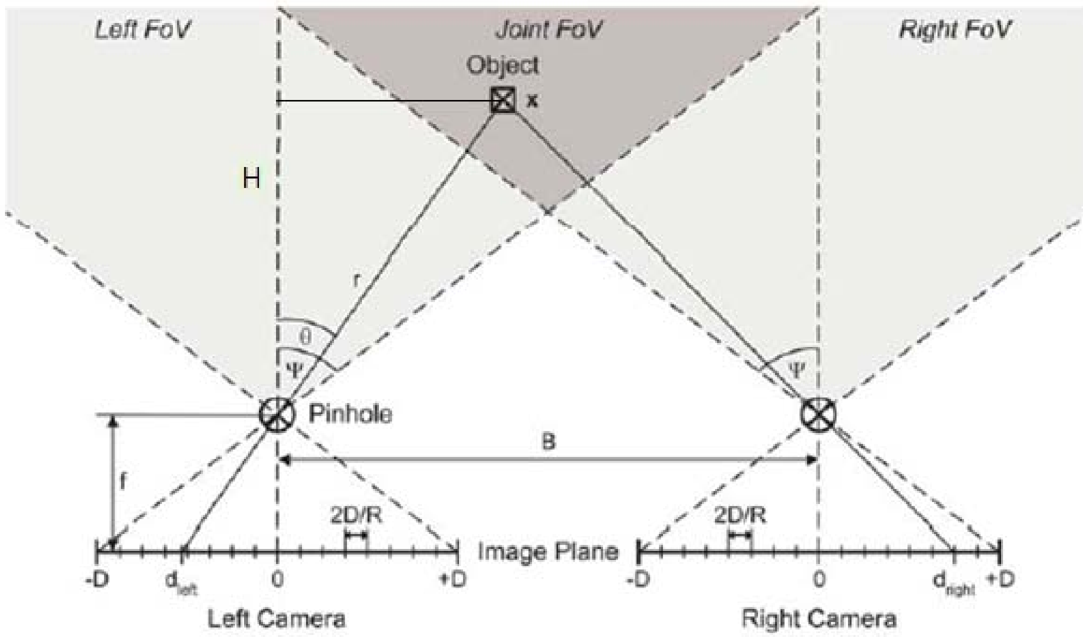


Figure 4 Illustration of 3D Vision sensor range finding

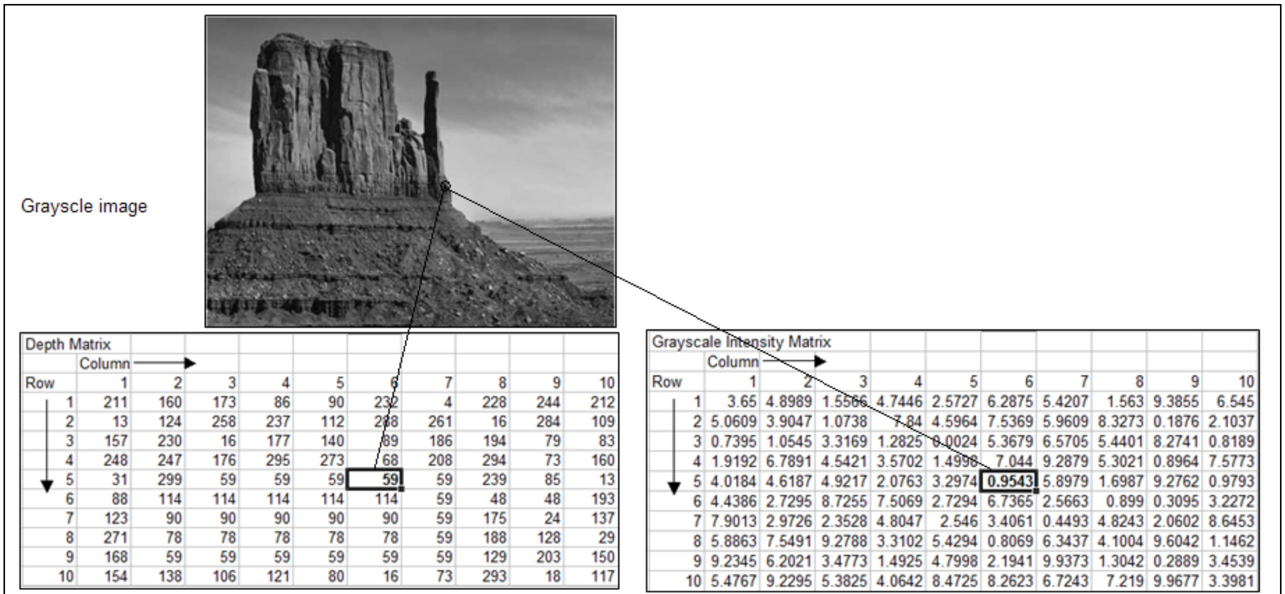
In Figure 4, let B be the base distance between two cameras of identical specifications, f be the focal length of the cameras, and d_{left} and d_{right} be the left and right disparities for object x, applying the principle of similar triangles in trigonometry

—

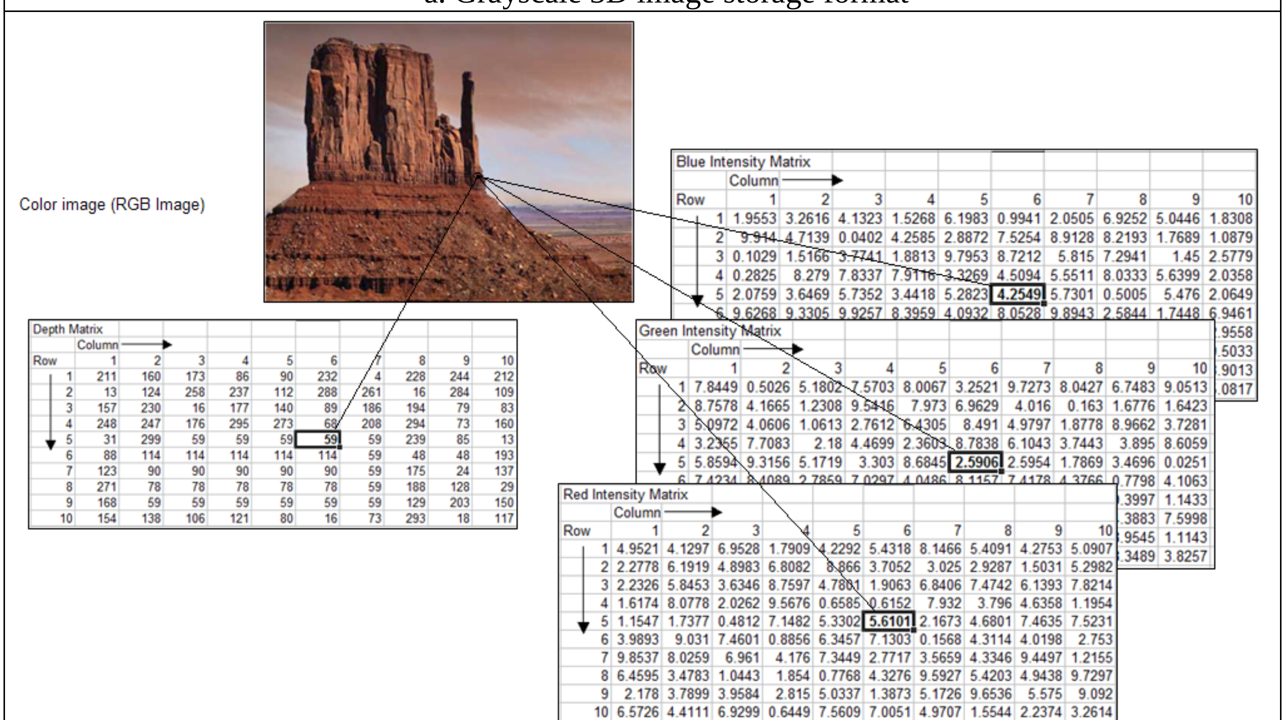
Where $d=d_{left}+d_{right}$

Using the depth information as a reference, stereo vision can eliminate the effects of shadows, identify both moving and stationary objects, and quickly distinguish different objects that are not in the same range of distance from the cameras. When storing color and depth information pixel-by-pixel, 3D images are stored as 3-dimensional matrices as shown in Figure 5.

From Figure 5, one can see that for the same image resolution, the number of data elements stored in a colored 3D image is twice that of a grayscale 3D image. Figure 6 indicates that if the resolution of a 3D image is increased from N1 x N2 to 2N1 x 2N2, the corresponding number of data elements in a 3D image file would be increased by 6 times for a grayscale 3D image, and 12 times for a colored 3D image. Note that computational effort is closely tied to the number of data elements that must be processed.



a. Grayscale 3D image storage format



b. Color (RGB) 3D image storage format

Figure 5 3D Image storage format

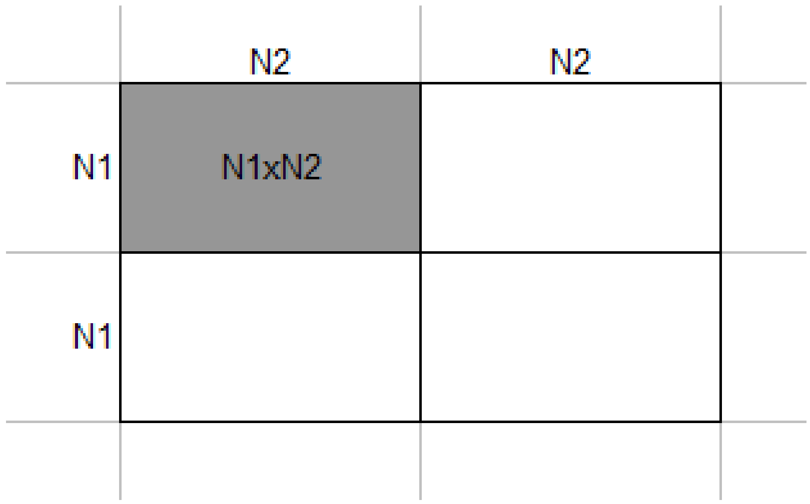


Figure 6 Number of data elements vs image resolution.

Currently, machine vision technology is still highly restrained by the CPU's computing power to achieve real time performance (the ability to process at least 25 image frames per second). When developing stereo vision in-vehicle pedestrian detection system, the researchers must balance performance needs and the following constrains:

- Vision sensor resolution (governed by the required detection distance and driving speed)
- Image type (monochrome or color, as they relate to computational effort)
- Mapping and processing speed of the stereo vision accelerator
- Processing speed of the computer
- Efficiency of the object detection algorithm

In this research, the cameras used are standard NTSC cameras of 720 x 480 resolutions with a 46^0 horizontal field of view. The image type chosen is grayscale; the video accelerator used is Acadia I™ vision accelerator; the computer is powered by Intel dual quad-core processor; and the algorithm implemented is a layered approach to be described later.

LAYERED PEDESTRIAN DETECTION ALGORITHM

The algorithm developed includes the following key components:

Structure Classifier

The purpose of the structure classifier is to quickly identify objects that frequently appear in the scene, but are clearly not pedestrians. These objects include buildings, trees, and tall vertical structure such as poles, etc. By rejecting pixel regions associated with these objects, the region of interest (ROI) needs to be analyzed further is greatly reduced.

Pedestrian Detector

The purpose of the pedestrian detector is to locate all pedestrian-like objects. This component also detects the ground plane to evaluate if the pedestrian-like objects are footed on the ground. Analyses done up to this stage typically still have high false positive rate (in the order of ten pedestrians per second).

Pedestrian Classifier and Verifier

The purpose of the pedestrian classifier and verifier is to perform second stage analysis within the reduced ROI to explicitly recognize pedestrians. It then performs a third stage analysis to further reduce false positives by explicitly rejecting detections of vehicles, bushes/shrubs. Finally, a fourth stage analysis is done to verify pedestrian detections by comparing the remaining pedestrian like objects to a set of human contour shapes stored in the learning library.

Pedestrian Tracker

The purpose of the pedestrian tracker is to compensate for missed classifications (typically caused by noise, and lighting changes, etc.), and support trajectory estimate.

Figure 7 shows the overall system flow diagram. Figure 8 illustrates how easily objects can be separated out from background objects in the 3D image using the vertical support histograms. Refer to the pedestrian in upper left corner of Figure 8, his background included buildings, and parked cars, etc., which seems noisy; however, if we examine the depth information, we will found that the pedestrian was separated from the building and the parked cars by a sizable distance.

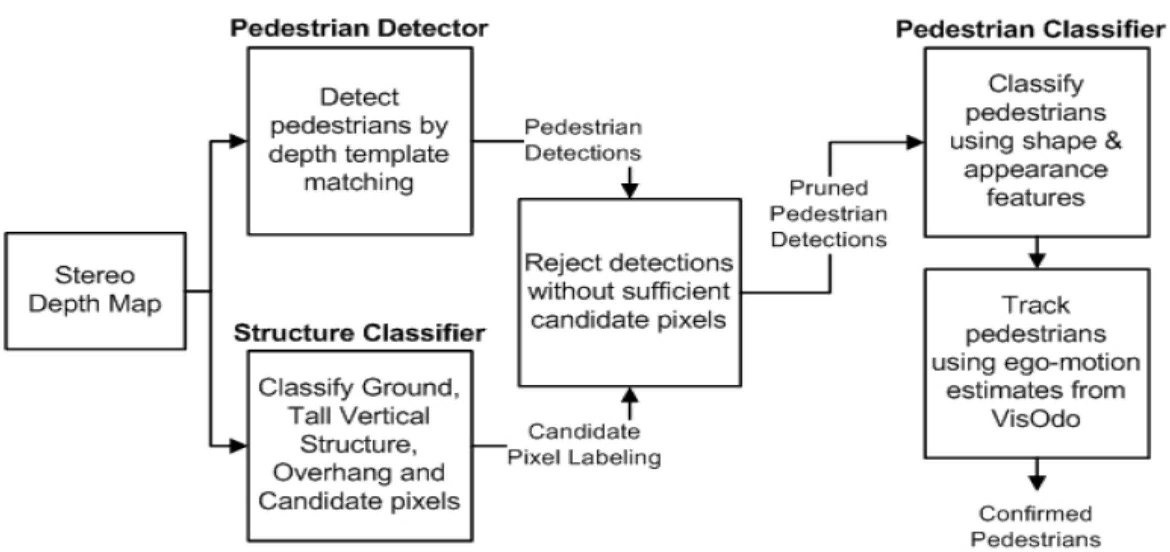


Figure 7 Layered pedestrian detection system flow diagram

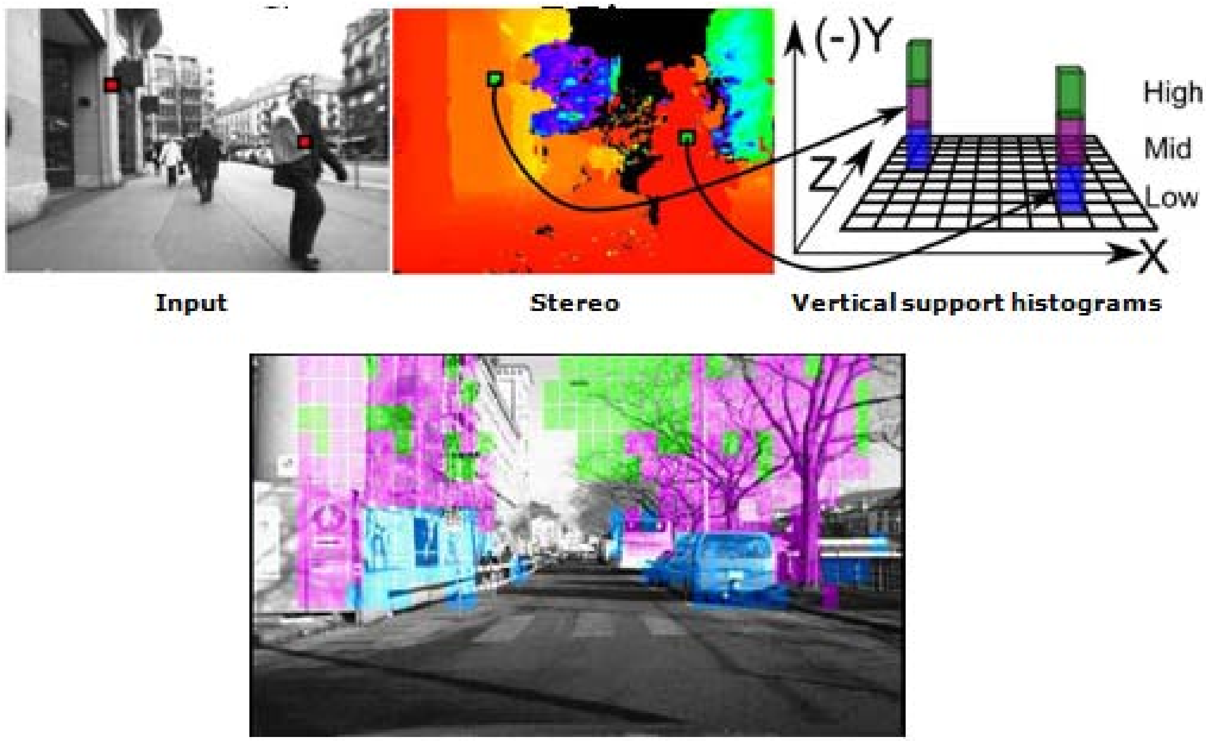


Figure 8 Isolation of objects using vertical support histograms

Details of the mathematical equations, numerical procedures, noise filtering techniques, and statistical pattern matching techniques for pedestrian detection and tracking can be found in the FHWA research report "Layered Object Recognition System for Pedestrian Collision Sensing", currently under editorial review. This paper describes the main assumptions, high level procedures, testing scenes, and pedestrian detection results. When this contract was completed in December 2009, the developed system could recognize the following types of objects:

- Ground
- Cars
- Bushes
- Trees
- Tall Slender Structures (Poles)
- Buildings
- Pedestrians

The research team used the following assumptions to positively identify pedestrians:

- Pedestrians are footed on the ground
- Pedestrians are between 1 to 2 meters above the ground (height)
- Pedestrians are less than 1 meter wide
- Pedestrians' body contours match one of the contours in the image library (Figure 9)

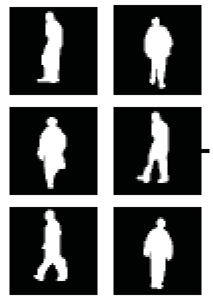


Figure 9 Pedestrian like object body contours

Assuming d1 (say 5 meters) and d2 (say 60 meters) represent the distance range within which in-vehicle pedestrian detection is desired, the detection process can be summarized as follow:

1. Extract pixels within the depth range of d1 and d2.
2. Recognize non-pedestrian objects such as buildings and trees, and eliminate the pixels associated with these objects.
3. Eliminate pixels associated with objects not footed on the ground, or objects with dimensions outside the height and width ranges defined for pedestrians.
4. Compare pedestrian like objects within the reduced ROI to library stored images, and then use Bayesian method to determine the likelihood of positive pedestrian detection.

In real urban scenes, when pedestrians appear in the traffic mix, usually part of their bodies are blocked from the view by other objects. To improve accuracy, pedestrian contour matching was done at the body parts level. This is done by building a library of the various poses of the human body parts (head, arm, leg, and back, etc., see Figure 10), when the defining shapes of a pedestrian like object are detected, they are compared to the library stored shapes to determine the probability of a match. This approach produced efficient and reliable performances on high resolution pedestrian ROIs. When the distance gets further, the pedestrian ROI gets smaller, contour extraction becomes fragile under low resolution images.

For the vision sensors and video accelerator used in this research, on the screen, an average sized human would appear as a 25-pixel tall object if standing 25 meters away, and as a 17-pixel tall object if standing 35 meters away. If detection distance needs to be increased, higher resolution of the 3D image is required. As already explained, higher resolution would mean dramatically increased number of data elements that must be processed in time. In the implementation, the research team developed three pedestrian classifiers corresponding to distance ranges of [0 – 20], [20 – 30], and [30 – 40] meters. The appropriate pedestrian classifier would be triggered to analyze an object based on its detected distance.

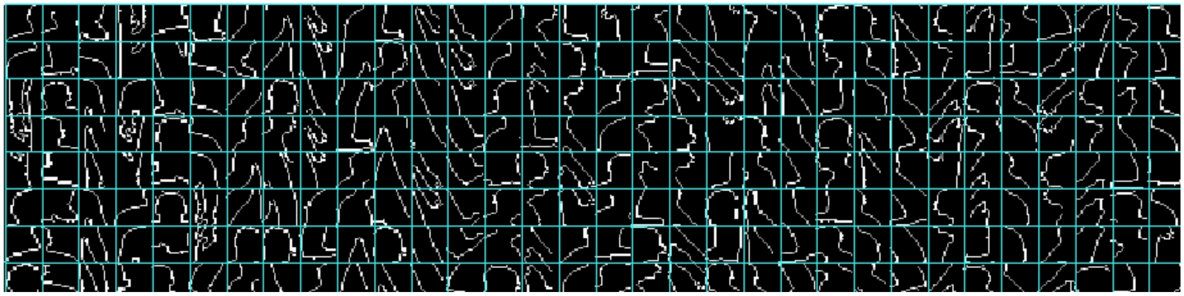


Figure 10 Body parts shape library

TEST RESULTS

The layered pedestrian detection algorithm was exercised using several publicly available 3D video datasets. After the research team felt confident with the result, the portable system was installed on a sport utility vehicle (a Toyota Highlander) and tested on roads inside the Sarnoff Corporation's campus and downtown Princeton under normal weather conditions. Test runs were conducted during the morning, afternoon, and evening hours, up to 8:00 p.m. Figures 11 to 14 show the typical scenes under which the system was tested.

Figure 11 Pedestrians crossing at an intersection during the day

Figure 12 Pedestrians crossing at an intersection while the vehicle was making a (right) turn

Figure 13 Pedestrians crossing at an intersection during evening

Figure 14 Pedestrians crossing at midblock

The developed algorithms were tested using videos collected from Sarnoff Campus, a challenging Europe dataset, and a public dataset (http://www.vision.ee.ethz.ch/~aess/dataset) that is recognized in the machine vision world as very challenging.

Tables 2 shows some typical performance results that the developed system achieved using the datasets mentioned. In some cases, both in-path detection and full field of view detection were performed. Usually in-path only detections gave better overall results.

Table 2 Representative performance results of the developed system

Sequence Name: 080613111722_BM-SHJ_cross-in-front (Parking Lot)			
Parameters: In-path (-1m to +1m from the center of the vehicle); Distance: 0 – 40 m			
Mode	Detection rate (%)	False positives/frame	# of persons
Detector-only	100	0.04	70
Detector + classifier	87.14	0	70
Detector + classifier + tracker	95.71	0	70
Sequence Name: 080613111722_BM-SHJ_cross-in-front (Parking Lot)			
Parameters: Full field of view; Distance: 0 – 40 m			
Mode	Detection rate (%)	False positives/frame	# of persons
Detector-only	100	7.09	383
Detector + classifier	87.73	0.36	383
Detector + classifier + tracker	96.87	1.1	383
Sequence Name: EuropeTour_Innsbruck.0_20070128_42_SVS_Data			

Parameters: Full field of view; Distance: 0 – 40 m			
Mode	Detection rate (%)	False positives/frame	# of persons
Detector-only	90.54	4.436	134
Detector + classifier	72.97	0.58	134
Detector + classifier + tracker	85.14	1.36	134
Sequence Name: seq00_rerun (Ess Sequence)			
Parameters: In-path; Distance: 0 – 40 m			
Mode	Detection rate (%)	False positives/frame	# of persons
Detector-only	94.56	0.82	584
Detector + classifier	66.61	0.16	584
Detector + classifier + tracker	92.81	0.45	584
Sequence Name: seq00_rerun (Ess Sequence)			
Parameters: Full field of view; Distance: 0 – 40 m			
Mode	Detection rate (%)	False positives/frame	# of persons
Detector-only	91.91	10.78	1816
Detector + classifier	66.13	1.56	1816
Detector + classifier + tracker	89.21	3.55	1816

Note: Suppose a video clip contains 100 images, and 80 of them contained a pedestrian, the detection rate should be interpreted as follow:

If the system correctly identified the pedestrian in 75 images, then detection rate would be 75/80 = 93.75%, and the missed detection rate would be 5/80 = 6.25%.

If in one image, the system identified 3 pedestrian objects but the 3 objects identified were actually not pedestrians, then the false positive for this frame would 3. The false positive/frame values shown in Table 2 are the average false positives of all the frames included in that data set.

Figure 15 shows the developed system's performance and its comparison with reported results (shown in dashed lines) found in the literature. When interpreting Figure 15, just keep in mind that under given requirement of false positive/frame (this would be a performance requirement for the system), the system that can deliver higher detection rate is the better system. It can be seen that in all cases, the results produced by the developed system were superior. The research team reported that at specific operating points of the classifiers, in crowded urban environments, the false alarm performance can be reduced to 1 every few minutes or lower, but this reduces the true detection rate. For highway driving, the false detection rate can approach to zero due to the performance of the structure classifier and the vehicle false-positive rejection classifiers.

The system was tested while driving at 15 mph and 30 mph. The frame rate of the developed system was between 7.5 and 10 Hz. The live system performed better at slower speeds (standstill to 15 mph) than at higher speeds. The offline system showed that if the frame rate were improved to 15 Hz, the performance would improve such that speeds of 30 mph could be easily handled.

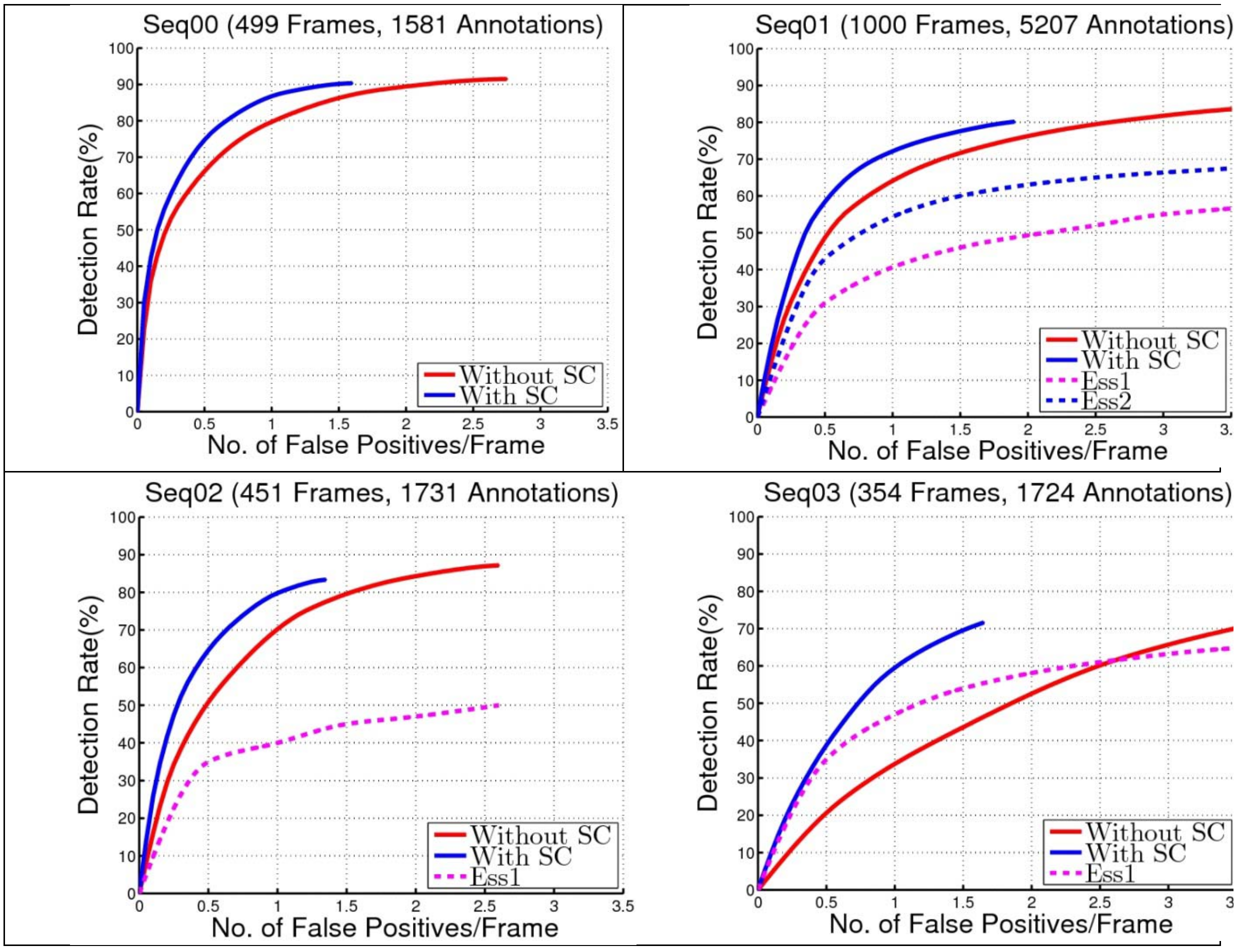


Figure 1 Developed system's performance on 4 datasets (Seq00-03), and comparisons with another representative approach from the literature (Andreas Ess from ETH Zurich)

Note that SC in Figure 15 denotes structural classification.

The developed system was able to detect, classify and track pedestrian like objects, and achieved the following performance results:

- Can estimate depth close to ± 10% of the true distance
- Can process grayscale 3D images at the rate of 7 to 10 fps
- Can detect pedestrian like objects while driving at up to 30 mph
- Can detect pedestrian like objects up to 35 meters away under good visibility condition, and up to 25 meters away under reduced visibility up around 8:00 p.m.
- Can track objects between 2 m and 35 m within the field of view
- Achieved 90% overall positive detection rate

The above results indicate that there are still gaps between the achieved performance and the desired performance described before.

MARKET STATUS AND FEASIBILITY OF COMMERCIALIZATION

During the fourth quarter of 2009, when this FHWA sponsored research was near completion, Mobileye, Inc, an Israel based company, started offering an in-vehicle pedestrian detection and warning system as an aftermarket product at around $1,000/unit, to selected passenger car models from BMW, Volvo, and GM. The product is based on Mobileye EyeQ2 systems, and uses a monocular camera. Mobileye claims the product can detect pedestrians up to 30 meters away, and can process videos at 15-20 FPS. The Mobileye C2-270, as shown in Figure 16, is a single camera-based safety solution integrating pedestrian, bicycle, motorcycle, and vehicle detection into one product.

Figure 16 Mobileye C2-270

Also during the fourth quarter of 2009, Trafficon, a Belgium based company, started offering Safe Walk, as shown in Figure 17. This is an infrastructure based 3D vision pedestrian detection and warning system, using a pair of CMOS 1/3” B&W vision sensors, and can detect both moving and waiting (stationary) pedestrians. Its recommended installation height is 3.5 m, and its recommended monitoring area is 3m by 4m. Safe Walk can be installed at intersection or mid block, and can be integrated into traffic signal controllers to help improve safety for pedestrians and reduce delay for motor vehicles when no pedestrian is present. The video processing speed and pricing information are unknown.

Figure 17 Trafficon Safe Walk

Constructing one stereo vision in-vehicle pedestrian detection system identical to the one developed for FHWA (Figure 3) would cost $30,000 or more (parts and labor). The quoted prices (May 2010 quote) for two Elmo NTSC Stereo Vision Sensors and one Acadia I Vision accelerator were $4,524 and $9.375 respectively. To make it into a commercial product, someone must integrate the vision accelerator and the CPU into one circuit board. The contractor indicated that if tens of thousands of units are ordered, the unit price could be lowered to one or few thousands dollar range.

DISCUSSIONS:

There are still gaps between the performance results achieved by the developed system and the performance targets desired. Currently we are restrained by both the hardware processing power and the efficiency of software algorithm. The system can be optimized in different ways:

1. Reduce the detection zone from full field of view to a smaller area.
2. Detect the moving direction and change the detection zone to in-path, left side, or right side based on whether the vehicle is moving straight, turning left, or turning right.
3. Dynamically set the detection range based on the vehicle's driving speed detected.

Stereo vision technology has been used in the development of both vehicle based and infrastructure based pedestrian detection systems. As of 2011, commercial 3D vision pedestrian detection product in vehicle based category is not yet available, and at least one infrastructure based 3D vision pedestrian detection system has been available since late 2009.

Vehicle based systems are designed as consumer products, and needs to be priced at $1,000 or lower to generate a sustainable market. Infrastructure based systems are designed for use by public agencies, and their prices may be set on case by case basis. These two types of products address different pedestrian safety issues and target different customers. Vehicle based systems can work anywhere the equipped vehicle is present (including intersections, mid blocks, and parking lots, etc), infrastructure base systems can be installed at intersections or mid block locations, it is suited for high pedestrian traffic areas. Vehicle based systems must meet stringent performance requirements (moving backgrounds, long detection range, fast driving speed) before they can be accepted by consumers. Infrastructure based systems deal with stationary background and covers a much shorter range and smaller monitor area, and therefore is easier to achieve desired performance requirements.

To enhance detection performance under low visibility, light enhancing techniques can be implemented to increase the image's contrast, the reported improvement was limited. Infrared sensors can be fused with vision sensors. This approach is effective in rural environment where the body temperature is higher than the surroundings. In urban setting, too many heat emitting objects may exist in the vicinity of the pedestrian(s), limiting the effects of infrared sensors.

Since the project was completed in late 2009, the costs of stereo vision technologies have been decreasing steadily, evidenced by the mass introduction of 3D televisions, 3D game consoles, and 3D camcorders, etc., all at very affordable price ranges in their categories. Although further advancements in both hardware and software are needed to achieve real time performance of the stereo vision pedestrian detection system, the relevant technologies (vision sensors, video accelerators, CPUs, infrared sensors, etc.) have been improved to the level that can readily support breakthroughs in pedestrian detection and warning applications. Applied researches in this area are expected to increase, and competition is expected to intensify

CONCLUSIONS

Based on test results achieved and the market status of pedestrian detection products, the following conclusions can be drawn:

1. The developed stereo vision in-vehicle pedestrian detection system achieved state-of-the-art performance for detection rate and false alarm rate when compared to other published results. However, there are still gaps to achieve the desired performance requirements.

2. 3D vision sensors alone cannot satisfy the performance requirements under low visibility conditions. Light enhancing (software) techniques can be implemented to help improve performance to a limited degree. Infrared sensors can be fused with 3D sensors to enhance detection performance under low visibility.

3. The false alarm rate achieved by the system is not low enough for deployment as a stand-alone system. Either the use of an additional sensor (e.g., radar or LIDAR) or reducing the horizontal field of view of the stereo camera is needed to achieve production-level performance.

4. The system needs further optimization to improve its performance to 15 Hz or higher. A higher frame rate is needed so that the system can be used on vehicles traveling at speeds higher than 30 mph.

ACKNOWLEDGEMENTS

The author would like to thank Jayan Eledath and his research team for conducting the research and development work.

REFERENCES

Chang P., Camus T., and Mandelbaum R. (2004). “Stereo-based vision system for automotive imminent collision detection”, In Intelligent Vehicles Symposium, IEEE, 274–279.

Chang P., Hirvonen D., Camus T., and Southall B. (2005). “Stereo-based object detection, classification, and quantitative evaluation with automotive applications”, In IEEE International Workshop on Machine Vision for Intelligent Vehicles, San Diego, USA.

Dalal N. and Triggs B. (2005). “Histograms of oriented gradients for human detection”, In CVPR, 886–893.

Eledath J., Matei B., Bansal M., Jung S. H. and Sawhney H. “Layered Object Recognition System for Pedestrian Collision Sensing”, Final Report for FHWA, under review.

Ess A., Leibe B., and Van Gool L. (2007). “Depth and appearance for mobile scene analysis”, In ICCV.

Ess A., Leibe B., Schindler K., and Gool L. V. (2008). “A mobile vision system for robust multi-person tracking”, In CVPR, pages 734–741.

Gavrila D.M. and Munder S. (2007). “Multi-cue pedestrian detection and tracking from a moving vehicle”, IJCV, 73:41–59.

Insurance Institute for Highway Safety, Fatality Facts: Pedestrians. http://www.iihs.org/research/topics/pedestrians.html

Leibe B., Cornelis N., and Cornelis L. V. G. K. (2007). “Dynamic 3d scene analysis from a moving vehicle”, In CVPR.

Mohan A., Papageorgiou C., and Poggio T. (2001). “Example-based object detection in images by components”, PAMI, 23:349–361.

Shashua A., Gdalyahu Y. and Hayun G. (2004). “Pedestrian Detection for driving assistance systems: single frame classification and system level performance” (2004). “IEEE Intelligent Vehicles Symposium”, Parma, Italy. http://www.mobileye.com/sites/mobileye.com/files/PedestrianDetectionforDrivingAssistance%20Systems.pdf

Stein G., Rushinek E., Hayon G and Shashua A. (2005). “A computer vision system on a chip: a case study from the automotive domain”, First IEEE workshop on embedded computer vision, San Diego, USA. http://www.mobileye.com/sites/mobileye.com/files/EyeQPaperCVPRworkshop-2005.pdf

Viola P., Jones M. J., and Snow D. (2005). “Detecting pedestrians using patterns of motion and appearance”, IJCV, 63:153–161.