Stereo- and Neural Network-Based Pedestrian Detection

Liang Zhao, *Student Member, IEEE,* and Charles E. Thorpe, *Senior Member, IEEE*

Abstract—**Pedestrian detection is essential to avoid dangerous traffic situations. In this paper, we present a fast and robust algorithm for detecting pedestrians in a cluttered scene from a pair of moving cameras. This is achieved through stereo-based segmentation and neural network-based recognition. The algorithm includes three steps. First, we segment the image into sub-image object candidates using disparities discontinuity. Second, we merge and split the sub-image object candidates into sub-images that satisfy pedestrian size and shape constrains. Third, we use intensity gradients of the candidate sub-images as input to a trained neural network for pedestrian recognition. The experiments on a large number of urban street scenes demonstrate that the proposed algorithm: 1) can detect pedestrians in various poses, shapes, sizes, clothing, and occlusion status; 2) runs in real-time; and 3) is robust to illumination and background changes.**

Index Terms—**Driver assistance system, neural networks, object detection, pedestrian detection, range image segmentation, stereo vision.**

I. INTRODUCTION

OBJECT detection and pedestrian recognition are essential to avoid dangerous traffic situations. In a driver assistance system for warning the driver of potential collision with nearby objects—especially pedestrians, we have to accomplish the following two tasks in real time. The first is to separate foreground objects from the background; the second is to distinguish pedestrians from other objects in order to protect pedestrians in danger. The first task is a segmentation procedure, the second one a recognition procedure.

In this paper, we employ a video-rate stereo system [20] to provide range information for object detection and pedestrian recognition. Using stereo to guide pedestrian detection carries with it some distinct advantages over conventional techniques. First, it allows explicit occlusion analysis and is robust to illumination changes. Second, the real size of an object derived from the disparity map provides a more accurate classification metric than the image size of the object. Third, using stereo cameras can detect both stationary and moving objects. Fourth, computation time is significantly reduced by performing recognition where objects are detected; it is less likely to detect background area as pedestrian since detection is biased toward areas where objects are detected.

Neural networks have been successfully applied to many real-time intelligent vehicle systems [13]–[15]. In our system, a neural network trained with the back-propagation algorithm [30] is used to discriminate pedestrians from other objects. Unlike similar systems [3], [14], [15], which are limited to detecting walking people, our system uses shape features instead of motions cues to detect both moving and stationary pedestrians. Since neural networks can express highly nonlinear decision surfaces, they are especially appropriate to classify objects presenting high degree of shape variability. In this system, the trained neural network implicitly represents the appearance of pedestrians in various poses, postures, sizes, clothing, and occlusion situations; it performs pedestrian detection in real time. The experiments on a large number of urban street scenes demonstrate the feasibility of the approach in terms of recognition rate and frame processing rate.

Pedestrian detection is of course only one part of a driver warning system. The context for this work is a program titled "Development and testing of performance specifications for a next generation side collision warning system," sponsored by USDOT [1], [2]. The goal of the project is to warn bus drivers of potential collisions, focusing on blind spots along the sides of a city bus. There are several parts to the project: studying accident records and determining causes of bus collisions; developing object detection sensors; predicting bus motion; and developing the driver interface. The preliminary functional goals first call for detecting vehicles, fixed objects, bicycles, and pedestrians. Each of these categories of objects can be detected as a three-dimensional (3-D) object; but each has different motion characteristics and each should be identified separately. Next, the functional goals call for predicting bus motion, for example incipient motion when the bus is about to pull away from a stop, or the sweeping motion of a bus turning through an intersection, or the angled motion of a lane change. Object detection will be especially important in the region the bus is about to traverse. Motion prediction can come directly, from rate gyros and wheel encoders; indirectly, from a route map and positioning system; or by inference, for example noting that the door has just been closed. Finally, the system will generate graded warnings. At the lowest level, the system will generate nonintrusive "situation awareness" information, similar to the driver looking in the rear-view mirrors to check for nearby objects. At the next level, the system will generate "alerts," a gentle notification (audio or video) of a potentially dangerous situation. The system will then progress to one or more levels of "warning," an intrusive alarm designed to notify the driver of impending collision.

Pedestrian detection is especially important for this project. While there are relatively few collisions with pedestrians,

they tend to be more serious: the number of fatalities from bus–pedestrian collisions is about the same as the number of fatalities from bus–vehicle collisions, even though there are many more bus–vehicle crashes. A fixed object will usually stay put; pedestrians are more unpredictable. The driver will receive a higher level alarm for a pedestrian than for a sign post in the same location.

At the same time, the state of the art of commercial side-looking systems is not very useful for pedestrian detection. Most side-looking systems are designed to cover blind spots on the side of a semi-trailer traveling along a limited-access highway. The systems are designed to look for cars or trucks, and therefore the sensors are widely spaced and are tuned to find large metal objects. The users are specifically warned that these systems are not designed to detect pedestrians.

Thus, a successful system such as the one described in this paper should have a large role to play in a collisions warning system. While the initial implementation will be for city busses, pedestrian safety is also a large issues for construction equipment, agricultural equipment, and other large moving vehicles that operate in close proximity to people.

This paper is organized as follows. Section II describes the related work on pedestrian detection. Section III presents the stereo guided object detection algorithm. The neural network based pedestrian recognition is presented in Section IV. Section V gives the experimental results, followed by the summary and conclusions in Section VI.

II. RELATED WORK

Most human tracking and motion analysis systems [3], [9], [10], employ a simple segmentation procedure such as background subtraction or temporal differencing to detect pedestrians. A serious problem with these approaches is the dynamic background caused by illumination changes or background (or camera) motion. Some techniques such as Pfinder [4], W^4 [5], and path clustering [8], have been developed to compensate for small, or gradual changes in the scene or the lighting. However, they cannot deal with large, sudden changes in the background. Although optical flow [11] can be used to detect independently moving targets in the presence of camera motion, it is not feasible for nonrigid object extraction since the movements of the body parts are different. Above all, a common drawback with the above approaches is the assumption that all detected objects are pedestrians; this limits the generalization and application of these schemes.

More sophisticated pedestrian detection techniques include a recognition step to discriminate pedestrians from other objects. These techniques can be classified into motion-based, shape-based, and multi-cue-based methods. Most motion-based approaches [15], [27]–[29], use rhythmic features or motion patterns unique to human beings for pedestrian detection. However, there are several limitations with these schemes. First, the pedestrian's feet or legs should be visible in order to extract the rhythmic features. Second, the recognition procedure requires a sequence of images, which delays the identification until several frames later and increases the processing time. Third, the procedure cannot detect stationary pedestrians and pedestrians

performing unconstrained and complex movement such as wandering around, turning, jumping, etc.

On the other hand, the shape-based approach relies on shape features to recognize pedestrians. Thus, this approach can detect both moving and stationary pedestrians. The primary difficulty in this approach is accommodating the wide range of variations in pedestrian appearance due to pose, nonrigid motion, lighting, clothing, occlusion, etc. References [9] and [10] use hand-crafted human models to detect pedestrians. An advantage of these methods is that they can analyze the motion of each body part; the disadvantage is that the segmentation of body parts is very difficult and even impossible in some situations. Lipton [26] depends on a dispersedness defined as the ratio $perimeter^2/area$ to classify human and vehicle. This classification metric is easy to calculate, but fails to distinguish humans from other objects with similar dispersedness and tends to misclassify pedestrians walking together as a vehicle. Papageorgious and Poggio [18] present a more robust pedestrian detection system based on wavelet analysis and the support vector machine (SVM) technique. However, the system has to search the whole image at multi-scales for pedestrians. This would be an extremely computationally expensive procedure, and it may cause multiple responses from a single pedestrian. More recently, Gavrila and Philomin [19] developed a real-time pedestrian detection algorithm based on Distance Transforms. The method includes an offline generation of the template hierarchy, and an online coarse-to-fine matching between the templates and the image. The algorithm is further sped up by hardware-specific means (i.e. SIMD instructions). Although the template hierarchy can capture the variety of object shapes, it can not handle large shape variations appropriately when pedestrians are very close to the camera.

To increase reliability, some systems [12], [25], integrate multiple cues such as stereo, skin color, face, and shape pattern to detect pedestrians. However, skin color is very sensitive to illumination changes [12]; face detection can only identify pedestrians facing the camera. These systems [12], [25], prove that stereo and shape are more reliable and helpful cues than color and face detection in general situations.

III. STEREO GUIDED OBJECT DETECTION

Our driver warning system is equipped with a stereo-based object detection module that detects foreground objects in real-time. In the following subsections we first describe stereo analysis, then explain the range image segmentation for object detection.

A. Stereo Analysis

Real-time stereo systems [20], [22]–[24], have recently been available and applied to people detection [25], [21], [12]. Among these systems, we chose the small vision system (SVS) developed by SRI [20] to perform stereo analysis, because the SVS can run at video rates on a standard PC without specialized hardware. The SVS has the following four features. First, it computes a dense disparity map using area correlation after a Laplacian of Gaussian transform. Second, the disparity value can be searched at various disparity levels, e.g., 16, 24, or 32
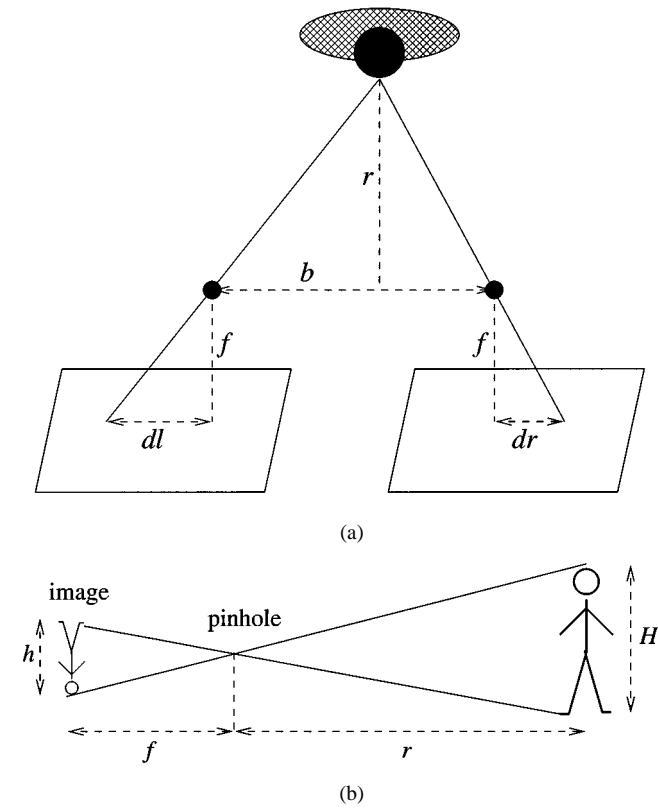
(a)

(b)

Fig. 1. (a) Stereo geometry and (b) image-forming geometry.

Fig. 2. (a) Left image from stereo cameras; (b) disparity map; (c) segmentation result; and (d) detected objects marked by boxes.

pixels. Third, the SVS performs postfiltering with an interest operator, and a left–right consistency check. Fourth, the SVS does $4\times$ range interpolation.

Fig. 1(a) displays stereo geometry. Two images of the same object are taken from different viewpoints. The distance between the viewpoints is called the *baseline (b)*. The focal length of the lenses is f. The horizontal distance from the image center to the object image is dl for the left image, and dr for the right image. Normally, we set up the stereo cameras so that their image planes are embedded within the same plane. Under this condition, the difference between dl and dr is called the *disparity*, and is directly related to the distance r of the object normal to the image plane. The relationship is

$$r = bf/d, \quad \text{where } d = dl - dr. \tag{1}$$

The height H of an object can be derived from the height h of the object image based on the assumption of a pinhole camera and the image-forming geometry, as shown in Fig. 1(b)

$$H = hr/f. \tag{2}$$

Fig. 2(b) shows a typical disparity image produced by the SVS. Higher disparities (close objects) are brighter. There are 64 possible levels of disparity; disparity 0 (black areas) are regions where the range data is rejected by the post-processor interest operator due to insufficient texture.

B. Range Image Segmentation for Object Detection

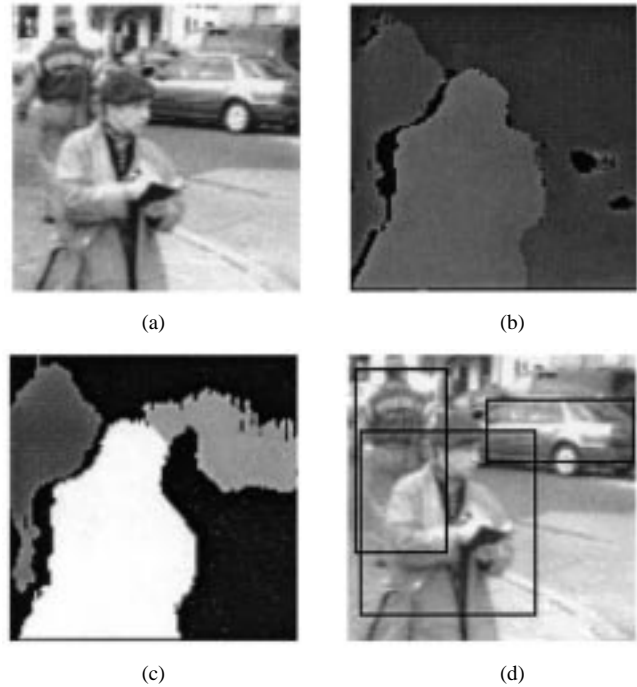Range information is a powerful cue for foreground/background segmentation. Compared with the intensity-based approach, range-based segmentation is less affected by light conditions, shadows and occlusion; compared with the 3-D-based approach, it is less expensive computationally and suitable for real-time implementation. Most stereo-based segmentation algorithms [21], [25], assume a static background and a pair of stationary cameras, thus, they do not work properly with a changing scene.

In contrast, our segmentation and grouping technique does not require a static background and therefore is robust to the changing scene. The algorithm proceeds in several stages of processing as explained below. In a collision warning system, we are only concerned about the objects within a close distance from our vehicle. Thus, we first eliminate background objects from the disparity image by range thresholding. The range is calculated using (1). We then employ a morphological closing operator to remove the noise and to smooth the foreground regions. Then, a connected-component grouping operator is applied to find the foreground regions with smoothly varying range. Finally, small regions are eliminated through size thresholding. The size range of a normal person is obtained from the statistic data given in [31]. Fig. 2(c) illustrates the segmentation result on Fig. 2(a). From this result we can see that the overlapping objects are successfully separated due to their different distances from the cameras.

The segmentation step provides a rough estimate of the object position and size. Subsequently, a 2-D box [shown in Fig. 2(d)] is fitted to each segmented region as an indication of an object. Then, all detected objects are fed into the trained neural network for pedestrian identification, which is presented in Section IV.

C. Postprocessing of the Object Detection Results

The major problem with the above object detection procedure is that each segmented region does not necessarily correspond
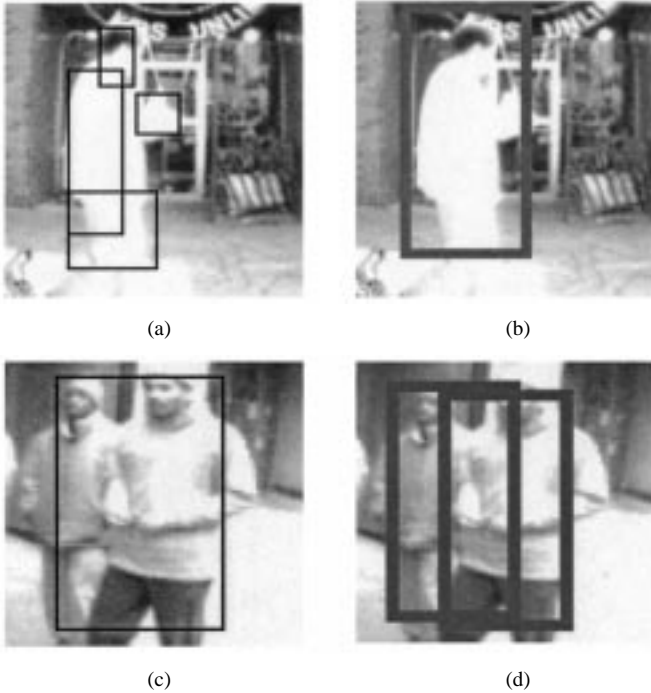
Fig. 3. (a) Pedestrian is segmented into several parts as indicated by the black bounding boxes. (b) These parts are correctly grouped into a single one (indicated by the thick black box) through the hypothesis and verification procedure. (c) Two pedestrians are merged into a single one. (d) The pedestrians are correctly separated and identified.

Fig. 4. Architecture of the neural network.

Fig. 5. ROC curve as a function of the threshold of the neural network output.

to a single object. Due to noise, insufficient texture, and the limited pixel and range resolutions of a disparity map, a single object would be divided into multiple parts [e.g., Fig. 3(a)], while objects close to each other may be merged into a single one [e.g., Fig. 3(c)]. Reference [7] depends upon simple human size and motion cues to determine the correspondence between regions and pedestrians. In this system, we rely on spatial and shape information to achieve the same goal. Since we can estimate the real size of an object from stereo analysis (1) and (2), our approach is much more reliable than that of [7].

This is a hypothesis and verification procedure. For small regions that are close to each other and have similar disparity values, they are temporally grouped into a single region. If the grouped region does not exceeds the size range of a normal person and can be classified as a pedestrian, then the grouping is confirmed, otherwise the small regions remain split. For a big region that exceeds the size range of a normal person, we use a window of a normal human size to search the whole region for pedestrians. If a subregion is identified as a pedestrian, it is separated from the original region; if no pedestrian is detected, the big region remains unchanged. Note that the verification is postponed after the pedestrian detection procedure. Figs. 3(b) and (d) show that the above procedure can correctly group regions belonging to a single object and split a region containing multiple objects.

IV. NEURAL NETWORK BASED PEDESTRIAN RECOGNITION

In this paper a neural network approach is introduced which can be trained for different kind of scenes and can deal with noisy data robustly. In the following subsections we will de-
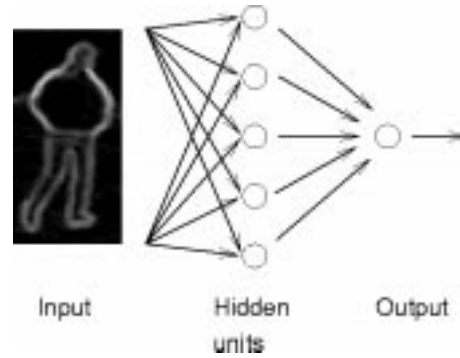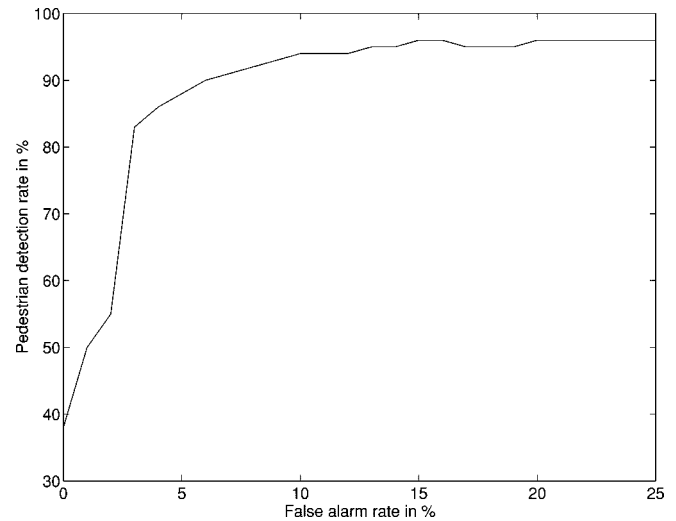
scribe the neural network input processing, training, and classification in details.

A. Preprocessing of the Input data

The design of the input data to the neural network (NN) is important; it directly affects the performance of the network. Our goal is to make the input data maintain the shape information for recognition while be reduced to a manageable amount.

An intensity image is often used as input in many neural network-based systems such as face detection [16] and autonomous vehicle steering [13]. However, there are several disadvantages with the intensity image. First of all, image intensity is not appropriate for encoding the consistent shape information of pedestrians, since pedestrians present a much higher degree of variability in color and texture than human faces and road surfaces. Secondly, image intensity is sensitive to illumination changes. In order to make the NN adapt to the changing illumination, the NN with an intensity image as input must be trained on-line, as shown in [13]. The silhouette extracted from the segmented region is invariant to color and texture changes. However, range segmentation does not always provide useful silhouettes when severe noise or occlusion exists. An alternative choice is edge image, but the edge detection result is greatly influenced by the heuristic thresholding. In this system, we employ the intensity gradient image (as shown in Fig. 4) to encode the shape information of an object, because

Fig. 6. Results of pedestrian detection on typical urban street scenes.

it has the same advantage as an edge image while it avoids the thresholding problem. Furthermore, the gradient image is robust to illumination changes. Therefore, the NN only needs to be trained off-line once.

The range information is not used as an input to the NN, because pedestrians may appear at various distances from the camera. However, we do use the range information to remove the distraction from the background and from other objects. This is done by only using the pixels within the detected object's range and assigning "don't care" or zero to out-of-range pixels.

Each segmented region is normalized to a fixed 30×65 window. The gradient values ranging from 0 to 255 are linearly scaled to range from 0 to 1 so that the network inputs would have values in the same interval as the hidden unit and output unit activations.

B. The Neural Network for Pedestrian Recognition

As shown in Fig. 4, we use a three-layer feed forward network for pedestrian recognition. There are 30×65 units in the input layer, five units in the hidden layer, and one unit in the output layer. Each layer is fully connected to the next, and each unit uses a sigmoid function for activation.

The network is trained by the back-propagation algorithm. The initial training examples are generated from a set of manually labeled example images produced by the object detection module. Totally, we have 5318 training data—1012 of pedestrians among which 10% are occluded, and 4306 of nonpedestrians ranging from traffic sign poles, parking meters, fire hydrants, vehicles, trees, to nonobjects. Then, we use the "bootstrapping" strategy [17] to improve the system performance. For each training step one gradient image is chosen at random from the training set. The network parameters are initialized by small

random numbers between 0.0 and 1.0, and are adapted during the training process. Therefore, the shape features to be extracted are learned from the training examples instead of being imposed *a priori*.

The network is trained to produce an output of 0.9 if a pedestrian presents, and 0.1 otherwise. Thus, we classify the detected object by thresholding the output value of the trained network: if the output is larger than the threshold, then the input object is classified as a pedestrian, otherwise as a nonpedestrian. The threshold determines the tradeoff between the rate of pedestrian detection and the rate of false alarm; we select this threshold by evaluating the receiver operating characteristics (ROC) curve shown in Fig. 5.

V. EXPERIMENTAL RESULTS

The system has been implemented on a Pentium II 450 Mhz system under Microsoft Windows NT with an Imagination PXC200 digitizer. It has been tested extensively on large amounts of live video in urban areas obtained from a pair of cameras mounted on the top of a minivan. Over 8400 instances of pedestrians and other objects have been presented. By adjusting the threshold of the neural network output, we achieve a pedestrian detection rate of 85.2% and a false alarm rate of 3.1%. The system can detect and classify objects over a 320×240 pixel image pair at a frame rate ranging from 3 frames/s to 12 frames/s, depending on the number of objects presented in the field of view of the cameras.

In Fig. 6, we show the results of our pedestrian detection system on some typical urban street scenes under different weather conditions. Fig. 7 illustrates the results of processing the video of a walking pedestrian. Figs. 6 and 7 show that our system can detect pedestrians in different size, pose, gait,

Fig. 7.   Pedestrian detection results of the video of a walking pedestrian.

clothing, and occlusion status. However, there are some cases where the system fails. Most of the false alarms are from the objects with outlines similar to those of human beings. Other failures occur when a pedestrian is almost similar in color to the background, or two pedestrians are too close to each other to be separable by stereo and shape cues. We believe that including motion cues to the current system will improve the system performance. Instead of gating the output from the NN, we keep the raw output as an estimate of the confidence of a person being presented. By tracking an object over a sequence of images, we will be able to get a confidence interval and get a more reliable indication if an object is a person or not.

It is worthwhile to compare our (CMU) method with MIT's wavelet-based method [18]. To allow a fair comparison, we let both methods work on the same set of test images. Table I gives the results of comparing the performance of the CMU system and MIT systems. MIT I is a fast version which only uses intensity information, while MIT II employs color information for pedestrian detection. The test set includes 617 people candidates generated from 100 images through the range segmentation procedure. Among these candidates, there are 254 pedestrians and 363 nonpedestrians. The results demonstrate that our pedestrian detection system is faster and more accurate than MIT's system, and the color information helps to improve the system performance.

TABLE I
COMPARING THE PERFORMANCE OF THE CMU SYSTEM AND MIT SYSTEMS

	Detection rate	False alarms	Frame rate
CMU	85.4%	18	3-12f/s (PC 450MHz)
MIT I	50%	100	10f/s (SUN)
MIT II	70%-81.6%	1:15,000 windows	1f/h (SGI)

based on stereo vision. Each object is then classified as pedestrian or nonpedestrian by a trained neural network. The two key elements which make this system robust and real-time are stereo-guided object detection, and neural network-based pedestrian detection. The system effectively combines 2 1/2 D information with the intensity information to detect pedestrians from a moving vehicle. The neural network is trained on a large number of pedestrian and nonpedestrian data extracted from complex scenes, therefore, it is applicable to various real world situations. The experiments on a large number of urban street scenes demonstrate that the proposed algorithm:

1) can detect pedestrians in various poses, shapes, sizes, clothing, and occlusion status;
2) runs in real-time;
3) is robust to illumination and background changes.

VI. SUMMARY AND CONCLUSIONS

This system is part of the Bus Driver Assistance project [1], [2] that aims at developing the next generation side collision warning system. To achieve this goal, foreground objects are first detected through foreground/background segmentation

ACKNOWLEDGMENT

REFERENCES

[1] C. Mertz, S. McNeil, and C. Thorpe, "Side collision warning systems for transit buses," in *IEEE Intelligent Vehicle Symp.*, Oct. 2000.

[2] S. McNeil, C. Thorpe, and C. Mertz, "A new focus for side collision warning systems for transit buses," in *Intelligent Transportation Society of America's 10th Annual Meeting and Exposition*, May 2000.

[3] Y. Guo, G. Xu, and S. Jsuji, "Understanding human motion patterns," *Proc. 12th Int. Conf. Pattern Recognition*, vol. 2, pp. 325–330, Jan. 1994.

[4] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 780–785, July 1997.

[5] I. Haritaoglu, D. Harwood, and L. Davis, "W^4—Real time detection and tracking of people and their parts," University of Maryland, College Park, Technical Report, Aug. 1997.

[6] H. Fujiyoshi and A. J. Lipton, "Real-time human motion analysis by image skeletonization," in *Workshop on Applications of Computer Vision*, 1998.

[7] O. Masoud and N. P. Papanikolopoulos, "Robust pedestrian tracking using a model-based approach," in *IEEE Conf. Intelligent Transportation Systems*, 1997, pp. 338–343.

[8] J. Segen and S. Pingali, "A camera-based system for tracking people in real time," in *Proc. 13th Int. Conf. Pattern Recognition*, 1996, pp. 63–67.

[9] D. Hogg, "Model-based vision: A program to see a walking person," *Image and Vision Computing*, vol. 1, no. 1, pp. 5–20, 1983.

[10] K. Rohr, "Toward model-based recognition of human movements in image sequences," *CVGIP, Image Understanding*, vol. 59, no. 1, pp. 94–115, Jan. 1994.

[11] P. J. Burt and J. R. Bergen *et al.*, "Object tracking with a moving camera: An application of dynamic motion analysis," in *Proc. IEEE Workshop Visual Motion*, 1989, pp. 2–12.

[12] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 601–608.

[13] D. A. Pomerleau, "Knowledge-based training of artificial neural networks for autonomous robot driving," in *Robot Learning*. Norwell, MA: Kluwer, 1993, pp. 19–43.

[14] M. Y. Siyal, M. Fathy, and F. Dorry, "Neural-vision based approach for real-time road traffic applications," *Electron. Lett.*, vol. 33, no. 11, pp. 969–970, May 1997.

[15] C. Wohler, J. K. Aulanf, T. Portner, and U. Franke, "A time delay neural network algorithm for real-time pedestrian recognition," in *Int. Conf. Intelligent Vehicle*, Germany, 1998.

[16] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 38–44.

[17] K.-K. Sung and T. Poggio, *Example-Based Learning for View-Based Human Face Detection*. Cambridge: Massachusetts Institute of Technology, Dec. 1994.

[18] C. Papageorgiou, T. Evgeniou, and T. Poggio, "A trainable pedestrian detection system," in *1998 IEEE Int. Conf. Intelligent Vehicles*, 1998, pp. 241–246.

[19] D. M. Gavrila and V. Philomin, "Real-time object detection for 'Smart' Vehicles," in *Int. Conf. Computer Vision*, Corfu, Greece, 1999.

[20] K. Konolige, "Small vision systems: Hardware and implementation," in *Proc. ISRR*, Hayama, 1997.

[21] C. Eveland, K. Konolige, and R. C. Bolles, "Background modeling for segmentation of video-rate stereo sequences," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1998, pp. 266–271.

[22] T. Kanade, A. Yoshida, K. Oda, H. Hano, and M. Tanaka, "A stereo machine for video-rate dense depth mapping and its new applications," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1996, pp. 196–202.

[23] J. Woodfill and B. Von Herzen, "Real-time stereo vision on the PARTS reconfigurable computer," in *IEEE Workshop FPGA's for Custom Computing Machines*, 1997, pp. 242–250.

[24] L. Matthies, A. Kelly, and T. Litwin, "Obstacle detection for unmanned ground vehicles: A program report," in *Proc. ISRR*, 1995.

[25] I. Haritauglu, D. Harwood, and L. S. Davis, "W^4S: A real-time system for detecting and tracking people in 2 1/2D," in *Eur. Conf. Computer Vision*, 1998, pp. 877–892.

[26] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Workshop Applications of Computer Vision*, Princeton, NJ, Oct. 1998.

[27] H. Mori, N. M. Charkari, and T. Matsushita, "On-line vehicle and pedestrian detection based on sign pattern," *IEEE Trans. Ind. Electron.*, vol. 41, no. 4, pp. 384–391, Aug. 1994.

[28] S. A. Niyogi and E. H. Adelson, "Analyzing and recognizing walking figures in xyt," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 469–474.

[29] ——, "Analyzing gait with spatiotemporal surfaces," in *IEEE Workshop Motion of Non-Rigid and Articulated Objects*, Austin, TX, 1994, pp. 64–69.

[30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distribution Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986, vol. I, Foundations, pp. 318–362.

[31] A. R. Tilley, *The Measure of Man and Woman: Human Factors in Design*. New York: H. D. Associates, 1993.

Liang Zhao received the B.E. degree in computer science from Tsinghua University, China, in 1991, and the M.E. degree in pattern recognition and intelligent control from Chinese Academy of Sciences **(CITY?)** in 1996. She is currently pursuing the Ph.D. degree in robotics at the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. Her research interests include computer vision and its applications in autonomous vehicles, intelligent user interface, and video analysis.

Charles E. Thorpe (M'87–SM'93) received the degree in natural science from North Park College, Chicago, IL, in 1979, and the Ph.D. degree in computer science from Carnegie Mellon University (CMU), Pittsburgh, PA, in 1984.

He is a Principal Research Scientist at CMU. During the 2000–2001 academic year, he is Acting Director of the Robotic Institute. His interests are in computer vision, planning, and architectures for outdoor robot vehicles. Since 1984, his Navlab group has built a series of ten robot cars, HMMWVs, minivans, and full-sized passenger buses, for military and civilian research. The Navlab group has pioneered new methods in stereo vision, laser rangefinding, 3-D terrain modeling, neural nets for perception, route planning, driver performance modeling, traffic simulation, teleoperation, vehicle control on rough terrain, and system architectures. He has also been involved with automated helicopters, walking robots and underwater robots. He also founded and directs the Master's Degree program in Robotics at CMU.

Dr. Thorpe is a Fellow of the American Association for Artificial Intelligence.