IEEE INTELLIGENT SYSTEMS, VOL.16, NR.6, PP.77-81, 2001

SENSOR-BASED PEDESTRIAN PROTECTION

D.M. Gavrila

Image Understanding Systems
DaimlerChrysler Research, Ulm 89081, Germany
dariu.gavrila@DaimlerChrysler.com
www.gavrila.net

1. THE PROBLEM

More than 430.000 pedestrians are injured and 39.000 killed yearly in traffic worldwide, see Table 1. For the European Union (EU), the corresponding numbers are 155.000 and 6000, see Table 2. Pedestrian accidents represent the second largest source of traffic-related injuries and fatalities, right after accidents involving car passengers. Especially children prove to be at risk, in situations such as shown in Figure 1.

Figure 1: A typical dangerous situation: a child suddenly crossing the street

The magnitude of the problem has meanwhile caught the attention of the legislative branch. The EU, for example, is studying proposals for legislating maximum tolerated pedestrian "impact coefficients" in the unfortunate event of the latter being hit frontally by a vehicle. Two classes of impact coefficient are considered, one involving the primary impact area, the (lower- and upper-) legs, and the other involving the more dangerous secondary impact area, the (child- and adult-) pedestrian head, at vehicle speeds of 40 km/h. Many aspects of such a specification are still subject of considerable debate. Among the open issues is whether a component-based crash test, where separate impactors are hurled towards the vehicle, can adequately model the kinematics of a human body during a crash. Another issue involves the large variation in pedestrian kinematics itself, when comparing a child and an adult, with widely different centers of mass at impact. Optimizing for the adult case can make things worse for the child case and vice versa.

While final test procedures and numbers have not materialized yet, it is clear that, because of the widely different object properties between pedestrians and vehicles, the room for improvement, in terms of energy absorption during a crash, is certainly limited. Oppos-

	Killed	Injured	Total
Passenger Cars	75.615	3.751.024	3.826.639
Pedestrians	**39.670**	**436.422**	**476.092**
Bicycles	6.872	236.027	242.899
Mopeds	3.151	163.854	167.005
Motor Cycles	10.972	227.946	238.918
Other	28.397	1.303.571	1.331.968
Total	161.677	6.118.844	6.283.521

Table 1: Road Traffic Accidents 1997 - Figures for UN-ECE Countries (Accident Source: UN-ECE)

	Killed	Injured	Total
Passenger Cars	22.502	995.026	1.017.528
Pedestrians	**6.049**	**155.151**	**161.200**
Mopeds	2.421	141.870	144.291
Bicycles	2.385	139.442	141.827
Motor Cycles	3.821	124.023	127.844
Other	4.559	121.816	126.375
Total	41.737	1.677.328	1.719.065

Table 2: Road Traffic Accidents 1997 - Figures for EU (Accident Source: UN-ECE)

ing demands are placed on vehicles; in addition to being "pedestrian-friendly", they should also perform well in crashes with "hard" objects, such as other vehicles and trees, and be appealing from a design point of view. Vehicle manufacturers are addressing these challenges by looking into extendable vehicle body structures (i.e. bumper, motor-hood), to be activated upon first impact with a pedestrian.

A complementary approach is to focus on sensor-based solutions, which enable vehicles to "look ahead" and detect pedestrians in their surroundings. This article probes the state-of-the-art in this domain, reviewing video-based approaches as well as approaches that involve active sensors (e.g. radar, laser range finders).

2. VIDEO-BASED APPROACHES

The use of video sensors comes quite natural for the problem of people detection. Texture information, provided at a fine angular resolution, enables the use of quite discriminative pattern recognition techniques. The human visual perception system is perhaps the best example of what performance might be possible with such sensors, if only the appropriate processing is added. Besides, for practical purposes, video cameras are cheap and because they do not emit any signals, there are no issues regarding interference with the environment.

An extensive amount of computer vision work exists in the area of "Looking-at-People", see [9] for a recent survey. What makes the pedestrian recognition application on-board vehicles particularly challenging is the moving camera, the wide range of possible pedestrian appearances and the cluttered (uncontrolled) backgrounds that are involved. Most work on vision-based pedestrian recognition has taken a learning-based approach, bypassing a pose recovery step altogether and describing human appearance in terms of simple low-level features from a region of interest. One line of work has dealt specifically with scenes involving people walking laterally to the viewing direction, either using the periodicity cue [6, 16] or by learning the characteristic lateral gait pattern [12].

A crucial factor determining the success of learning methods is the availability of a good foreground region. In contrast to applications such as surveillance, where the camera is stationary, standard background subtraction techniques are of little avail here because of the moving camera. Independent motion detection techniques can help [16], although they are difficult to develop, themselves. Yet, given a correct initial foreground region, some of the burden can be shifted to tracking [1, 4, 5, 12, 15, 17]

A complementary problem is to recognize pedestrians in single images; this is particularly relevant for the case of a pedestrian standing still. One general approach involves shifting windows of various sizes over the image, extracting low-level texture features, and using standard pattern classification techniques to determine the presence of a pedestrian. For example, [14] uses wavelet features in combination with a Support Vector Machine (SVM) classifier. More recently, this work has been extended to involve a component-based approach [13]. However, pure brute force window sliding approaches currently do not demonstrate performance-speed characteristics suitable for use on-board vehicles. The system described in [10] called the Chamfer System, addresses this by using a two-step approach for object recognition. In the first step, contour features are used in a hierarchical template matching approach to efficiently "lock" onto candidate solutions using distance transforms. By capturing the objects shape variability by means of a template hierarchy and using a combined coarse-to-fine approach in shape and parameter space, this method achieves very large speed-ups compared to an equivalent brute-force method. Only in the second step it reverts to texture-based pattern classification, on the candidate solutions provided by the first step. Another powerful technique to establish regions of interest (ROIs) is stereo vision. It is used in [8, 20] combination with texture-based pattern classification. Work in [2] also uses stereo vision, but prefers to combine it with a verification technique based on symmetry properties.

Lately, there has been increased interest in video sensors which operate outside the visible spectrum. Longtime used exclusively in the military domain, infra-red sensors are finding their way into civilian applications; a development aided by the advent of cheaper, uncooled cameras. The principle of detecting pedestrians by the heat their bodies emit is very appealing indeed (e.g. [18]). Yet pedestrians are not the only sources of heat that can be observed in a traffic environment, vehicles generate heat too. Even the pavement can appear hotter on a summer day than the pedestrian body. Thus, rather than offering the solution for pedestrian detection *per se*, infra-red sensors provide means to simplify the segmentation problem. Still required are the pattern recognition techniques discussed before.

3. OTHER SENSOR APPROACHES

Video sensors do not directly provide depth information; stereo vision derives depth by establishing feature correspondence and performing triangulation. Active sensors, such as radar and laser range finder, on the other hand, measure distances directly. Radar has already been introduced commercially in the vehicle domain for adaptive cruise control applications (e.g. Distronic System on-board Mercedes-Benz's S-Class). For near-distance applications, such as pedestrian detection, ongoing investigations focus on 24 Ghz radar technology [11]. Object localization is enhanced by placing multiple radar sensors on the relevant parts of the vehicle and employing triangulation-based techniques. Object classification, i.e. distinguishing pedestrians from other objects such as cars and trees, is achieved by examining the power spectral density plot of the reflected signals. Spectral content and reflectivity are the object properties to consider in this context. Objects of smaller spatial extents, such as pedestrians, have narrower peaks in the plot than, say, cars. At the same time, material properties of the object's surface determine the strength of reflected radar signals. Metallic parts of cars and other vehicles reflect much better than human tissue, by at least an order of magnitude. Human tissue, in turn, reflects much better than non-conductive materials, such as the wood of trees.

Eye-safe laser range finders offer other promising means for obstacle detection. Their main appeal lies in their fast and precise measurement of depth and their large field of view. For example, the laser range finder described in [11] has a depth accuracy of +/- 5 cm up and a range of 40 m for objects with at least 5% reflectivity (this includes most, if not all, relevant targets). Furthermore, its horizontal scans cover a 180 degree field of view in increments of 0.5 degree at 20 Hz, making the sensor especially suitable to cover the area just in front of the vehicle.

4. CURRENT SYSTEMS

At least three pedestrian systems are currently integrated on-board vehicle demonstrators [2, 8, 20]. All three are are video-based and use a two-step detection-verification framework for efficient pedestrian recognition; the region of interest is invariably provided by stereo vision.

The Carnegie Mellon University system [20] combines stereo vision with a neural-network pattern classification approach. The texture features used for classification are obtained by applying a high-pass filter to the region of interest and normalizing for size. Their system, running at 3-12 Hz, is in particular aimed at assisting bus drivers in urban traffic. It will later be expanded to cover the sides of the bus, and eventually, to provide full 360 degree coverage.

The University of Pavia system [2], implemented in the ARGO experimental autonomous vehicle, combines stereo vision with template matching techniques for detecting pedestrian head/shoulder shapes. Candidate regions are verified using vertical symmetry considerations. The authors report good detection results in a distance range of 10-40 meter.

At DaimlerChrysler, we have been working on pedestrian recognition as part of our multi-year effort to extend driver assistance beyond the highway scenario into the complex urban environment [8, 7, 10, 12]. Of special interest has been the so-called Intelligent Stop&Go on-board our Urban Traffic Assistant (UTA) demonstrator (for the latter, see Figure 2). It allows UTA to autonomously follow a lead vehicle, while being aware of relevant elements of the traffic infrastructure (e.g. road lanes, traffic signs, and traffic lights) and other traffic participants. Our most recent pedestrian detection system consists of stereo vision-based obstacle detection, and shape-based object classification with the Chamfer System [10] (see Section 2. The reader is referred to web site www.gavrila.net for a few video clips.

Figure 2: DaimlerChrysler's Urban Traffic Assistant (UTA) demonstrator

The above systems will soon be joined by others. The EU has recently initiated a major initiative for pedestrian protection under the fifth-framework project PROTECTOR [3, 11]. It brings together major vehicle manufacturers, sensor suppliers and research institutions, in order to develop intelligent systems on-board vehicles for the reduction of the accident rate involving pedestrians, bicyclists and other unprotected traffic participants. Among the tasks already completed is the analysis of accident statistics and the definition of relevant traffic scenarios. Three different sensor technologies are pursued, radar and laser range finder and video, to be implemented on two passenger cars (FIAT and DaimlerChrylser) and one truck (MAN). Final systems will be evaluated on a test track under standardized and realistic conditions (i.e. using dummies), sometime in 2002. User interface and user acceptance studies will conclude this project.

5. THE ROAD AHEAD

Success or failure of a pedestrian safety system, from a technical point of view, will very much depend on the rate of correct detections versus false alarms that it produces, at a certain processing rate and on a particular processor platform. But what rate will be needed for actual deployment of a sensor-based pedestrian system ? The question is difficult to answer because desired rate will very much depend on the final system concept. If, for example, the system concept only involves a warning function, performance criteria will likely be less stringent than for the case which involves active vehicle control.

Perhaps it is then easier to establish where we currently stand regarding performance. For this purpose, let us pursue the following thought experiment. Consider a (fictional) video-based pedestrian detection system which involves a succession of three components: stereo-based obstacle detection, template-based shape matching and texture-based pattern classification. For argument's sake, assume that the performance of each individual component is independent of the others. We conservatively estimate that, in order not to miss any pedestrians, the stereo component will produce 1 pedestrian ROI per 10 seconds, when tested in urban traffic (in lieu of hard experimental data, we use a value derived from our experience). We assume that the stereo component accomplishes this by employing simple heuristics regarding sizes and locations of the rectangular regions it detects as obstacles. Since we cannot expect the pedestrian ROI to be exactly outlining the pedestrian, we assume that 10 probes are needed to extract the pedestrian correctly. For the shape-based and texture component we estimate a detection rate of 95% at a false positive rate in the order of 10^{-3} and 10^{-1} per candidate region, respectively, based on the figures cited in [10, 14] and [20]. All in all, we arrive, in this *best case* scenario, at a false positive rate of 1 per 10^4 seconds or 1 per 2.8 hours, for a detection rate of 90%. Integrating results over time by tracking will improve this figure somewhat, but this effect will be offset by the lower filter ratios of the shape and texture components which, in practice, cannot be considered independent. Based on this, it is fair to say that the false positive rate will need to be reduced by at least *one order of magnitude* in order to obtain a viable pedestrian system, while maintaining the same detection rate.

Fortunately, there are a number of ways to significantly reduce the false positives rate. Improved multi-cue video-algorithms, i.e. combining distance, shape, texture, and motion cues, hold the promise to successively decimate the false alarm rate, as illustrated in the previous paragraph. Large benefits are also expected from sensor fusion, e.g. combining video and laser range finder approaches. Finally, telematics concepts, involving communication between pedestrian and vehicles combined with GPS-based localization, could close any remaining performance gap. Although it is unrealistic to expect people to buy special-purpose pedestrian protection devices, pedestrian safety systems could piggyback on the pervasiveness of the future communication infrastructure (e.g. UMTS, Bluetooth).

Challenges remain even after the pedestrian detection problem has been solved. After all, what is needed is an assessment of the danger that a particular traffic situation represents. This assessment will take into account positional and speed information of pedestrian and vehicle. But with larger look-ahead of the system, beyond the pre-crash range, prediction quickly becomes unreliable. Pedestrians can very easily change their heading direction, furthermore, accurate risk assessment will increasingly require good scene understanding. For example, the danger associated with a pedestrian heading towards the street will very much depend on the placement of the road boundaries, whether a traffic light exists, and if so, whether it is green or not. This suggests that, in the long run, a reliable and anticipatory pedestrian system will have to be aware of several types of infrastructural elements, by either perception or telematics approaches. At least some of the complexity might be reduced by limiting the scope of a pedestrian protection system to only cover specific traffic scenarios; this will represent a good intermediate solution.

In conclusion, this article has provided an overview of the state-of-the art in the area of sensor-based pedestrian detection. In spite of the very difficult technical challenges that lie ahead, some degree of optimism is warranted given the progress that this domain has seen over the past few years. Considering Tables 1 and 2, the goal certainly appears worthwhile.

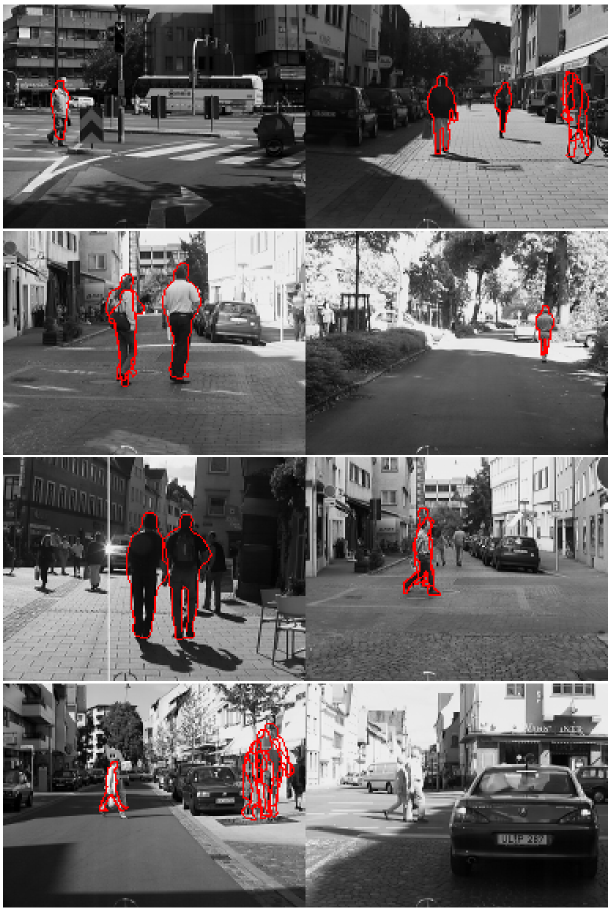

Figure 3: Pedestrian detection results by the Chamfer System [10] shown in red. In addition to correct detections, the figure also illustrates typical shortcomings, such as false detections in heavily textured image areas (e.g. left image, bottom row) or missing detections in areas of low contrast and/or occlusion (e.g. right image, bottom row).

6. REFERENCES

[1] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *Proc. of the European Conference on Computer Vision*, pages 299–308, 1994.

[2] A. Broggi, M. Bertozzi, A. Fascoli, and M. Sechi. Shape-based pedestrian detection. In *Proc. IEEE Intelligent Vehicle Symposium*, pages 215–220, Dearborn, U.S.A., 2000.

[3] P. Carrea and G. Sala. Short range area monitoring for pre-crash and pedestrian protection: the Chameleon and Protector projects. In *Proc. of the 9th Aachener Colloquium on Automobile and Engine Technology*, pages 629–639, Aachen, Germany, 2000.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and applications. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] C. Curio, J. Edelbrunner, T. Kalinke, C. C. Tzomakas, and W. von Seelen. Walking pedestrian recognition. *IEEE Transactions on Intelligent Transportation Systems*, 1(3), 2000.

[6] R. Cutler and L. Davis. Real-time periodic motion detection, analysis and applications. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 326–331, Fort Collins, U.S.A., 1999.

[7] U. Franke, D. M. Gavrila, A. Gern, S. Görzig, R. Janssen, F. Paetzold, and C. Wöhler. From door to door - principles and applications of computer vision for driver assistant systems. In L. Vlacic, F. Harashima, and M. Parent, editors, *Intelligent Vehicle Technologies*, chapter 6. Butterworth Heinemann, Oxford, 2001.

[8] U. Franke, D. M. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler. Autonomous driving goes downtown. *IEEE Intelligent Systems*, 13(6):40–48, 1998.

[9] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

[10] D. M. Gavrila. Pedestrian detection from a moving vehicle. In *Proc. of the European Conference on Computer Vision*, pages vol. II, pp.37–49, Dublin, Ireland, 2000.

[11] D. M. Gavrila, M. Kunert, and U. Lages. A multi-sensor approach for the protection of vulnerable traffic participants - the PROTECTOR project. In *Proc. of the IEEE Instrumentation and Measurement Technology Conference*, volume 3, pages pp. 2044–2048, Budapest, Hungary, 2001.

[12] B. Heisele and C. Wöhler. Motion-based recognition of pedestrians. In *Proc. of the International Conference on Pattern Recognition*, 1998.

[13] A. Mohan, C. Papageorgiu, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361, 2001.

[14] C. Papageorgiu and T. Poggio. A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33, 2000.

[15] V. Philomin, R. Duraiswami, and L. Davis. Quasi-random sampling for condensation. In *Proc. of the European Conference on Computer Vision*, pages 134–149, Dublin, Ireland, 2000.

[16] R. Polana and R. Nelson. Low level recognition of human motion. In *Proc. of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 77–82, Austin, 1994.

[17] G. Rigoll, B. Winterstein, and S. Mller. Robust person tracking in real scenarios with non-stationary background using a statistical computer vision approach.

In *Proc. of Second IEEE Int. Workshop on Visual Surveillance*, pages 41–47, Fort Collins, USA, 1999.

[18] T. Tsuji, H. Hattori, N. Nagaoka, and M. Watanabe. Development of night vision system. In *Proc. IEEE International Conference on Intelligent Vehicles*, pages 133–140, Tokyo, Japan, 2001.

[19] C. Wöhler and J. Anlauf. An adaptable time-delay neural-network algorithm for image sequence analysis. *IEEE Transactions on Neural Networks*, 10(6):1531–1536, 1999.

[20] L. Zhao and C. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(3), 2000.