Improved Vision-based Pedestrian Detection System for Collision Avoidance and Mitigation

M. A. Sotelo, D. Fernández, I. Parra, E. Naranjo

***Abstract*— This paper describes an improved stereo vision system for anticipated detection of car-to-pedestrian accidents. An improvement of previous versions of the pedestrian dection system is achieved by compensation of the cameras pitch angle, since it results in higher accuracy in the location of the ground plane and more accurate depth measurements. The pedestrian detection system has been applied to collision avoidance and mitigation. Collision avoidance is carried out by means of deceleration strategies, whenever the accident is evitable. Likewise, collision mitigation is accomplished by activating an active hood system. For that purpose, the system has been mounted and tested on two different prototype cars and tested on private circuits using dummies.**

I. INTRODUCTION

Each year, thousands of pedestrians and cyclists are struck by motor vehicles. Most of these accidents take place in urban areas where serious or fatal injuries can be sustained at relatively low speed. Only in the European Union about 8.000 pedestrians and cyclists are killed and about 300.000 injured. In North America approximately 5.000 pedestrians are killed and 85.000 injured. In Japan approximately 3.300 pedestrians and cyclists are killed and 27.000 injured [1]. To reduce these figures, the European Commission is forcing the automotive industry to introduce safety measures to drastically cut the number of fatalities by 50% by 2010 compared to the figures of 2001. While in the first phase (up to 2005) passive measures were introduced to achieve the requirements, the way to reach the final requirements in the second phase (from 2005 to 2010) seems to demand the introduction of active or preventive safety measures.

The range of active safety measures is quite wide [2] including ideas like active hood systems, outside airbags, active bumpers or automatic deceleration [3]. Since these actuators have to be activated just before the crash occurs, sensors such as radar and cameras have compusorily to be used in order to provide a measure of the time-to-collision well in advance. The study of the cumulative frequency of crashes between vehicles and pedestrians [1] shows that a crash speed of up to 40 km/h can cover more than 75% of total pedestrian injuries. Thus, if a speed of up to 40 km/h is considered, the levels of injury suffered by pedestrians involved in frontal impacts with motor vehicles will be significantly reduced. Furthermore, some accidents are likely to be avoided, for velocities well bellow 40 km/h, if pedestrians are detected by the sensors onboard the car with enough anticipation. In such cases, the deployment of deceleration strategies makes sense not only for collision mitigation but also for collision avoidance.

One of the most popular measures for collision mitigation is the use of the so-called active hood system [4] [5]. These types of systems raise the hood of the vehicle in case of an unavoidable crash. This way a more elastic deformation of the hood can be achieved to get a reduced force over the pedestrian, especially over the head. Most of the pedestrians involved in a car-to-pedestrian accident have the first contact with the car's frontal region. This usually means that the legs make contact with the front bumper and after 50 to 150 ms the body, and especially the head, hit the bonnet or the windscreen of the car as stated in [6]. For adult leg injuries, the major source is the front bumper of vehicles. When an adult pedestrian is struck by a vehicle, the first impact is generally between the pedestrian knee region and the vehicle's front bumper. Because this initial contact is below the pedestrian's center of gravity, the upper body begins to rotate toward the vehicle. The pedestrian's body accelarates linearly relative to the ground because the pedestrian is being carried along by the vehicle. The second contact is between the upper part of the grille or front edge of the bonnet and the pedestrian's pelvic area. The final phase of the collision involves the head and thorax striking the vehicle with a linear velocity approaching that of the initial striking velocity of the vehicle. Research has shown that the linear head impact velocity is about 90 percent of the initial contact velocity [1]. Child and adult heads adults legs are the body regions to be most affected by contact with the front end of vehicles. On vehicles, the bonnet top and the windscreen are the vehicle regions mostly identified with a high potential for contact in a car-to-pedestrian accident. These areas can cover more than 65 per cent of the fatal and serious injuries.

Sensor systems onboard the car are mandatorily required for predicting the car-to-pedestrian distance and the time-to-collision, both for collision avoidance and for collision mitigation. Cameras are the most commonly used sensors for that purpose. As a matter of fact, a number of remarkable pedestrian detection systems have been developed by scientists around the world using cameras in the visible espectrum [7] [8] [9] [10] and infrared cameras [11] [12] [13]. The stereo-vision system for pedestrian detection developed by the ITS Research Group at the University of Alcalá [14] is currently being deployed on real cars with a double purpose: collision avoidance and collision mitigation. In either case, the accuracy of depth measurements is a critical issue for

M. A. Sotelo, D. Fernández, and I. Parra are with the Department of Electronics. Escuela Politécnica Superior. University of Alcalá. Alcalá de Henares, Madrid, Spain sotelo,llorca, parra@depeca.uah.es

E. Naranjo is with the Informatics Department at the Industrial Automation Institute. CSIC, Madrid, Spain jnaranjo@iai.csic.es

it has a direct influence on the estimation of the time-to-collision. An improved pedestrian detection system can be achieved by compensating for variations of the cameras pitch angle, since these variations can modify the ground plane position and distort depth measurements. Accordingly, a pitch angle estimator has been developed and deployed, as explained in this paper. Collision avoidance is carried out by means of deceleration strategies, whenever the accident is evitable. Likewise, collision mitigation is accomplished by activating an active hood system. The collision avoidance module has been tested on a Citron C3 Pluriel car equipped with a stereo vision system. Tests were carried out on a private circuit. The collision mitigation module was mounted on a Seat Cordoba car equipped with an active hood system that is triggered by the stereo vision system. Real experiments were performed for both collision avoidance and collision mitigation using dummies. Practical results are provided in this paper.

The rest of the paper is organized as follows: section II provides a description of improvements in time-to-collision estimation accuracy with regard to the already existing vision-based pedestrian detection system developed in previous work by the authors. Section III presents current improvements by means of pitch angle compensation. Implementation and experimental results are described in section IV. Finally, section V summarizes the conclusions and future work.

II. Pedestrians Detection

Pedestrians detection is carried out using an improved version of the system described in [14], where a combination of feature extraction methods was implemented for vision-based pedestrian detection. The basic components of pedestrians are first located in the image and then combined with a SVM-based classifier. Candidate pedestrians are located using a subtractive clustering attention mechanism based on stereo vision. An efficient candidate selection mechanism is a crucial factor in the global performance of the pedestrian detection system. The candidate selection method must assure that no miss-detection occurs. Candidates, that are usually described by a bounding box in the image plane, must be detected as precisely as possible. All the improvements presented in this paper deal with the candidate selection stage, and are intended to increase the accuracy of the time-to-collision estimation. In [14] subtractive clustering techniques are used for candidates selection after obtaining a cloud of points corresponding to obstacles in the 3D scene. Nonetheless, these points are subject to outliers due to correlation noise. In the current work, some improvements have been implemented in the correlation step in order to increase robustness and reduce noise. These post-processing measures are described next:

- Only strong responses of the correlation function along the epipolar line are considered as correspondents.
- If the global maximum of the function is not strong enough relative to others local maximums, then the current left image point is rejected (*unique maximum*).
- Right image correlated points are also correlated over the left image (*mutual check strategy*). If the new left matched points are not exactly the same than the original ones, these correspondences are considered as noise (*multi-correlation*).
- In case different left image points would be correlated over the same right image point, two strategies could be taken: maximum correlation criterion or minimum disparity criterion. The second one is used so as the noise due to structured backgrounds, which usually produces close 3D points, is avoided (*minimum diparity*).

After applying the previous steps, the number of correlated points gets decreased by an average of 24% after using multi-correlation. By using both multi-correlation and minimum disparity methods an average of 37% of points are selected as noise.

An adaptive subtractive clustering is applied then to obtain the 3D coordinates of pedestrians candidates. Another improvement with regard to [14] is the implementation of a multi-candidate generation strategy in order to improve the accuracy of 3D position of detected candidates since this is also a critical issue for accurate time-to-collision estimation. The purpose is to produce several candidates around each selected cluster in an attempt to compensate for the effect of the candidate bounding box accuracy in the recognition step. Accordingly, several candidates are generated for each candidate cluster, by slightly shifting the original candidate bounding box in the u and v axes in the image plane. The candidate selection method yields generic obstacles with a 3D shape that is similar to that of pedestrians. The 2D candidates are then produced by projecting the 3D points over the left image and computing their bounding box. Nonetheless, the 2D bounding box corresponding to a 3D candidate might not perfectly match the candidate appearance in the image plane, yielding false pedestrian depth measurements.

Two strategies are proposed to solve the "bounding accuracy effect". The first one consists in training the classifier with additional badly-fitted pedestrians in an attempt to absorb either the extra information due to large bounding boxes containing part of the background, or the loss of information due to small bounding boxes in which part of the pedestrian is not visible. The second strategy consists in performing a multi-candidate (MC) generation for every extracted candidate, trying to hit the target and add redundancy. Three window sizes are defined: the window size generated by the candidate selection method, a 20% oversized window, and a -20% downsized one. These three windows are shifted 5 pixels in each direction: top, down, left, and right. Thus, a total of 15 multicandidates are generated for each original candidate.

A majority criterion is followed in order to validate a pedestrian. Thus, the MC strategy yields a pedestrian if more than 5 candidates are classified as pedestrians. This number has been defined after extensive experiments. In average, the candidate selection mechanism generates 6 windows per frame, yielding a total of 90 candidates per frame after the multi-candidate process. In case the number of candidates

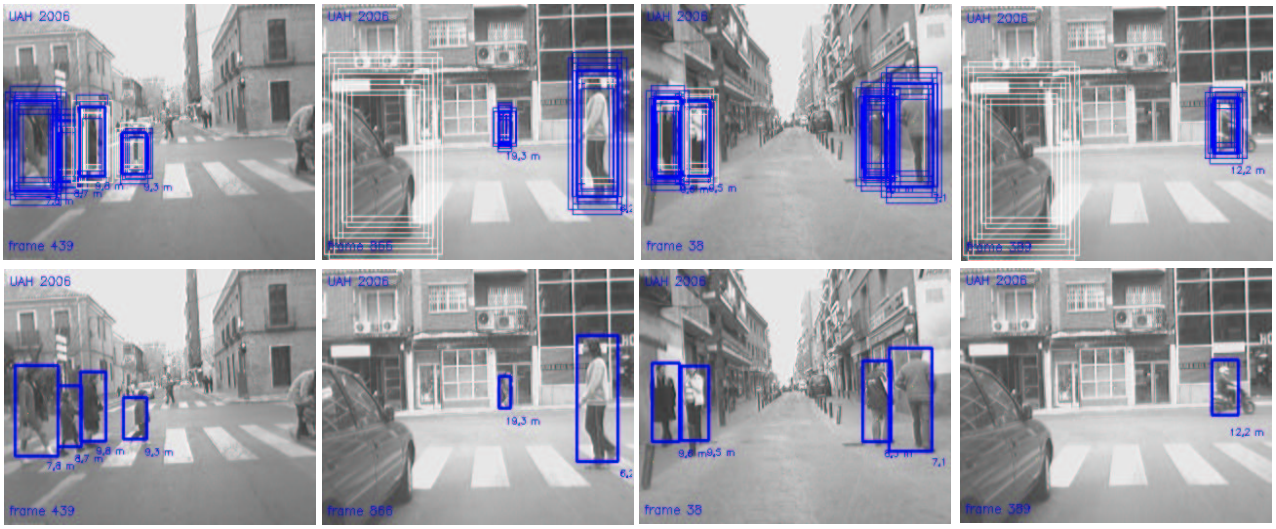

Fig. 1. Upper row: multi-candidate generation. Bottom row: results after classifying the 15 candidates.

Method	Detected	Missed	False alarms
Single	138	10	9
Multicandidate	143	5	8

TABLE I

GLOBAL PERFORMANCE EVALUATED IN A SET OF SEQUENCES WITH A TOTAL DURATION OF 20 MINUTES

generated by the attention mechanism increases abruptly the MC approach might become impractical. A major benefit derived from the MC approach is the fact that classification performance of pedestrians at long distance increases. Figure 1 depicts typical images from our test sequences. The number below the bounding box represents range. The right most image shows a motorcyclist that is detected as a pedestrian (false positive). In the left most image two kids are properly detected and their range is correctly measured.

The results obtained in practical experiments are listed in Table I. For each row in the table the following information is provided: type of method, number of detected pedestrians (only pedestrians below 25 m are considered), number of missed pedestrians, and number of false alarms issued by the system.

The analysis of the results reveals that detection rate is improved using multi-candidate generation for the single frame classification while the false positive rate remains practically unchanged. The single and MC methods exhibit a ratio of 9 and 8 false alarms, respectively, in 20 minutes of operation. This yields ratios of 27 and 24 false alarms per hour, respectively. The MC method does not exhibit an appreciable improvement in FPR because false alarms are usually due to candidates with features similar to pedestrians, such as motorbikes, fences, windows reflections, etc. These candidates are not better classified by using multi-candidate generation since at least 5 out of the 15 candidates continue being classified as pedestrians. The DR is 93.24% using the single method, and 96.62% using the multi-candidate one. Five non-detected pedestrians by the single method have been detected using the multi-candidate approach. Several cases are corrected by this approach. For example, kids, which are usually selected as candidates with very few points, are better detected by using the MC method. When several people are together in the same area the candidate selection method usually yields bounding boxes which fall between two people due to the 3D approach. Thanks to the MC method these pedestrians are well classified. Let us clarify the fact that the other missed pedestrians were partially occluded or completely out of the vehicle path. There is other important effect due to the use of MC classification: pedestrians can be classified at larger distances. This implies that the system can anticipate pedestrian detection with more time in advance.

III. PITCH ESTIMATION

Detection range specification in vision based pedestrian detection applications is usually no longer than $30m$ due to several constraints like camera resolution, pedestrian size, etc. Thus, Flat road geometry is considered, i.e., road curvature can be neglected in the near range. Thanks to the stereo approach the vertical road profile can be directly extracted. The robust correlation process reduces the number of 3D points under the road (which is directly proportional to the amount of correlation errors). Taking into account a base plane without pitch change, the height of the camera relative to the base plane, and the camera vertical field of view, the origin of the world coordinate system is placed at the intersection point between the base plane and the lower boundary of the vertical field of view. Figure 2 depicts the lateral projection of 3D points on the YOZ plane.

The number of 3D projected points over the same 2D point in the lateral view are coded in a gray scale image. Thus the weight of matching errors is reduced. As in [15] we consider the vertical displacement due to roll negligible in

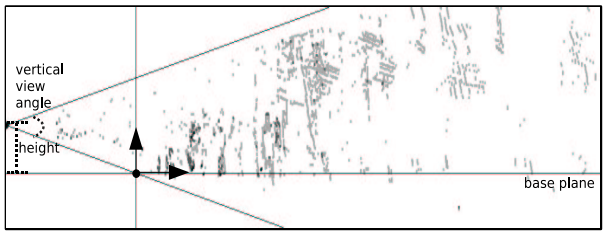

Fig. 2. 3D projected points on the YOZ plane up to 30m.

comparison to the displacement due to pitch. From the point of view of the world coordinate system origin, and varying the slope to cover all possible pitch values, uniformly spaced rays are cast. Gray level values (number of points) along each ray i are counted in a histogram $H(i)$. The histogram is normalized and the mean value $\bar{h}$ is computed. A stable *jump* over $2/3\bar{h}$ in the histogram is looked for from under the road upwards. Being $i = 0$ the lowest ray and $i = N$ the highest one, pitch angle is selected as follows:

$$\begin{aligned}&\text{for } i = 0 \text{ to } N\\ &\quad \text{if } (H(i) > \tfrac{2}{3}\bar{h} \text{ and } H(i+1) > \tfrac{2}{3}\bar{h} \text{ and } H(i+2) > \tfrac{2}{3}\bar{h}\\ &\qquad \text{and } H(i), H(i+1), H(i+2) > H_{min})\\ &\qquad \text{then } \alpha = \alpha_i\text{; break;}\\ &\quad \text{else } \alpha = 0\text{;}\end{aligned} \tag{1}$$

The parameter H_{min} is used to avoid pitch estimation errors when there are not enough road points detected. Figure 3 depicts three examples for positive, negative and zero pitch angle values. The darker the ray the higher the number of accumulated points. The estimated pitch angle is drawn in bold.

In order to have a steady estimation of the pitch angle, a linear Kalman filter is applied. The state vector is composed by the pitch angle and its velocity, $x_k = \{\alpha_k, \dot{\alpha}_k\}$ and the measurement vector by the pitch angle, $z_k = \{\alpha_k\}$. The following equations show the proposed pitch angle estimation:

$$\vec{x_k} = \begin{pmatrix} \alpha_k \\ \dot{\alpha}_k \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{k-1} \\ \dot{\alpha}_{k-1} \end{pmatrix} + \vec{r_k} \text{ state eq.} \tag{2}$$

$$z_k = \alpha_k + \vec{o_k} \text{ measurement eq.} \tag{3}$$

where $\vec{r_k}$ and $\vec{o_k}$ are the state vector noise and the measurement vector noise, respectively. Accordingly, a smoother pitch angle estimation is obtained. So, the transformation matrix that has to be applied in order to perform 3D points correction is:

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{pmatrix} \tag{4}$$

Once the longitudinal profile of the road has been extracted, and 3D points corrected, road surface points, which are not obstacle points, can be easily removed by using their Y coordinate value. By doing so, these points do not perturb the clustering step. The average computation time required to estimate the pitch angle is less than 10ms. This implies that the pitch estimation module takes less than 5% of total computation time required by the pedestrian detection system.

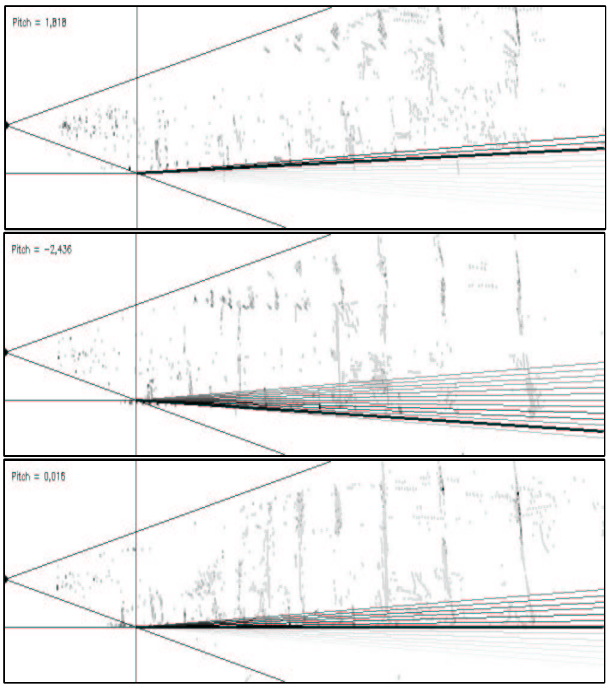

Fig. 3. Pitch angle estimation. From top to bottom: positive pitch angle, negative pitch angle and pitch angle about 0 degrees.

IV. Implementation and Results

The system was implemented on two different platforms for testing collision avoidance and collision mitigation, respectively. On the one hand, a Pentium IV at 2.4 GHz running the Knoppix GNU/Linux Operating System and Libsvm libraries [16] was mounted on a SEAT Cordoba prototype vehicle equipped with an active hood system. This vehicle was tested on a private circuit using light dummies so that the vehicle can run them down as many times as required without being damaged. On the other hand, a second prototype based on a G4 Power PC was installed on a Citroen C3 Pluriel equipped with automatic steering wheel, brake and accelerator pedals. This vehicle was tested on a different test circuit, emulating an urban quarter, where velocity deceleration techniques and avoidance manoeuvres can be properly executed. In both cases, the stereo vision system uses 320×240 pixel images. The complete algorithm runs at an average rate of 15-20 frames/s, depending on the number of pedestrians being tracked and their position. The average rate has a strong dependency on the number of pixels being matched because of the correlation computational cost, which consumes, in average, 80% of the whole processing time.

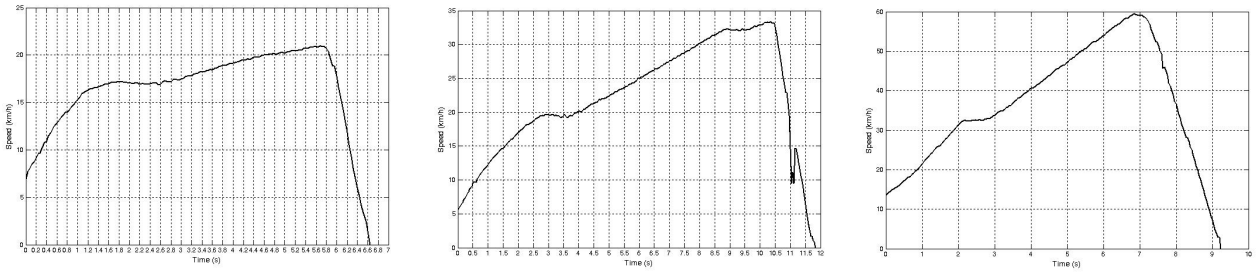

Fig. 4. Emmergency stops at different velocities.

A. Collision avoidance

As previosly stated, collision avoidance experiments consists on emmergency stop manoeuvres by brusquely decelerating the vehicle. Several experiments were conducted to achieve an estimation of the distance that is need for preventing an accident as a function of the current vehicle speed. Although this computation can be easily done using theoretical equations, we decided to execute real emmergency stop manoeuvres and derive a realistic estimation of the stopping time at different velocities. Figure 4 depicts the evolution of the vehicle velocity in several experiments where an emmergency stop manoeuvre was executed. Based on the previous experiments, an approximate equation for computing the stopping time was estimated using linear regression techniques.As depicted in figure 5 the estimated curve is a reasonable approximation of the real curve. The final expression of the estimated stopping time t_{stop}, including actuator latencies, is provided in equation 5.

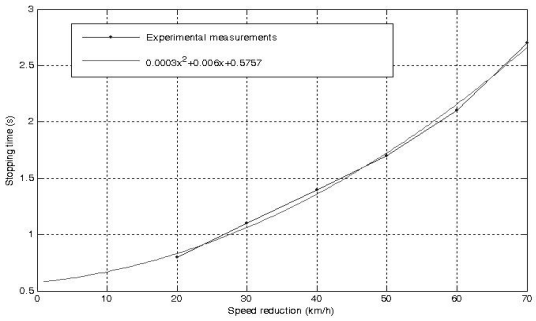

Fig. 5. Regression curve.

$$t_{stop} = 0.0003 \cdot x^2 + 0.006 \cdot x + 0.5757 \quad (5)$$

where x stands for vehicle velocity at the time of starting the emmergency stop manoeuvre. Similarly, the estimated distance required by the vehicle to come to a full stop D_{stop} is provided equation 6.

$$D_{stop} = 0.5 \cdot (0.0003 \cdot x^3 + 0.006 \cdot x^2 + 0.5757 \cdot x) \quad (6)$$

B. Collision mitigation

Collision mitigation experiments were carried out using a light-weight dummy. The car ran over the dummy in several experiments at different velocities bellow 50 km/h. In all cases, the active hood had to be activated at a pre-programmed time before the collision took place, ranging between 200-350ms. The experiments were recorded using a high speed camera providing 1000 frames per second. Thus, the ground truth was obtained from recorded videos given that each frame corresponds to 1ms. With the use of special software applications digital videos recorded with high speed cameras can be played frame by frame in order to precisely determine the exact time between the activation of the active hood and the instant the car-to-dummy collision occurred. This measurement can be done with an accuracy of 1ms. Figure 6 shows the arrangement used in the experiments.

Fig. 6. Arrangement used in collision mitigation experiments.

Normally, the specifications for hood activation are in the range 200-350ms. In previous versions of this work, the authors achieved an accuracy of 120ms in the time-to-collision estimation. This figure depends on the vehicle velocity and the number of correlated points for candidates selection. Let us take into account the fact that the exact desired activation time can occur while an image is being processed. Considering that the execution time of the algorithm is in the average of 50ms, and that during that time the system remains *blind*, issuing the activation signal

TABLE II

TIME-TO-COLLISION ESTIMATION IN THREE DIFFERENT EXPERIMENTS.

Impact velocity	Timer activation	Pre-programmed time	Ground truth
25 km/h	373 ms	350 ms	350 ms
34 km/h	308 ms	250 ms	235 ms
40 km/h	290 ms	250 ms	200 ms

at the end of the frame process would result in a poor resolution activation accuracy that depends on the processing time of the vision-based algorithm. To avoid this problem, a timer is programmed 2 or 3 frames before the activation time is reached. After that, the process is resumed. The timer interrupts the main process when the programmed time expires. In that moment, the vehicle hood is activated and the vision process is overrun. After the improvements described in this paper, the accuracy in the issuing of the activation signal has been increased to some 50ms (worst case). Table II shows the results for three collision mitigation experiments at different velocities ranging between 20-50 km/h. The first column represents the vehicle velocity just before the collision. The second column shows the time between the activation of the timer and the real collision. The third column is the desired (pre-programmed) activation time, and the fourth column represents the ground truth provided by the high-speed camera recording the experiment.

V. CONCLUSIONS AND FUTURE WORK

In this work, we have implemented some improvements on our previous vision-based pedestrian detection system intented to increase the accuracy of the time-to-collision estimation in car-to-pedestrian accidents. The final goal is to use the system in collision avoidance and mitigation. The improvements described in this paper are summarized bellow:

1) Enhanced correlation method by using mutual check consistency, multi-correlation, and minimum disparity criteria.
2) Adaptive subtractive clustering technique in the candidates selection stage.
3) Multi-candidate (MC) generation strategy. This has proven to be useful not only to increase the accuracy of the time-to-collision estimation, but also to improve the performance of the SVM-based pedestrian classification system.
4) Pitch estimation. The estimated value is used to compensate for road slope.

After implementation of the previous techniques, a remarkable increase in performance has been achieved in the accuracy of the time-to-collision estimation, as demonstrated in practical experiments. Nonetheless, further improvement is still necessary, especially in collision mitigation applications based on active hood, where activation of the vehicle hood must be absolutely precise to maximize the mitigation effect. This can be achieved by reducing the computation time required to execute the algorithm. Two alternatives are being considered at present: implementation of the code on a FPGA-based platform and optimization using ALTIVEC. Finally, comparisons with ground-truth data and further quantitative assessment of improvements are being carried out at present. These figures will be provided in future publications.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish Ministry of Education and Science by means of Research Grant DPI2005-07980-C03-02.

REFERENCES

[1] U. T. D. W. F. for Harmonization of Vehicle Regulations., "Pedestrian safety global technical regulation preamble."
[2] M. M. Meinecke and M. A. Obojski, "Potentials and limitations of pre-crash systems for pedestrian protection," in *Second International Workshop on Intelligent Transportation*. Hamburg, Germany, March, 2005.
[3] M. M. Meinecke, M. A. Obojski, D. Gavrila, E. Marc, R. Morris, M. Tons, and L. Letellier, "Strategies in terms of vulnerable road user protection," in *EU Project SAVE-U - Deliverable D6.*, 2003.
[4] www.siemensvdo.com.
[5] www.autoliv.com.
[6] K. C. Fuerstenberg, "Pedestrian protection using laserscanners," in *Proc. IEEE Conference on Intelligent Transportation Systems*. Vienna, Austria, September, 2005.
[7] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: single-frame classification and system level performance," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 1-6, Parma, Italy, June, 2004.
[8] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proc. IEEE Intelligent Vehicles Symposium*. Dearborn, Michigan, USA, October, 2000.
[9] D. M. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: The protector system," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 13-18, Parma, Italy, June, 2004.
[10] G. Grubb, A. Zelinsky, L. Nilsson, and M. Rilbe, "3d vision sensing for improved pedestrian safety," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 19-24, Parma, Italy, June, 2004.
[11] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," in *IEEE Transactions on ITS*. Vol. 6, No. 1, March, 2005.
[12] B. Fardi, U. Schuenert, and G. Wanielik, "Shape and motion -based pedestrian detection in infrared images," in *Proc. IEEE Intelligent Vehicles Symposium*. Las Vegas, USA, June, 2005.
[13] M. Bertozzi, A. Broggi, A. Lasagni, and M. D. Rose, "Infrared stereo vision-based pedestrian detection," in *Proc. IEEE Intelligent Vehicles Symposium*. Las Vegas, USA, June, 2005.
[14] M. A. Sotelo, I. Parra, D. Fernández, and J. E. Naranjo, "Pedestrian detection using svm and multi-feature combination," in *Proc. IEEE Conference on Intelligent Transportation Systems*. Toronto, Canada, September, 2006.
[15] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, P. Ciprian, R. Schmidt, and T. Graf, "High accuracy stereovision approach for obstacle detection on non-planar roads," in *IEEE Intelligent Engineering Systems (INES)*. Cluj Napoca, Romania, pp. 211-216, 2004.
[16] Chih-Chung and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.nut.edu.tw/ cjlin/libsvm.