Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

3D Candidate Selection Method for Pedestrian Detection on Non-Planar Roads

D. Fernández, I. Parra, M. A. Sotelo, P. Revenga, S.Álvarez and M. Gavilán

Abstract— **This paper describes a stereo-vision-based candidate selection method for pedestrian detection from a moving vehicle. Non-dense 3D maps are computed by using epipolar geometry and a robust correlation process. Non-flat road assumption is used for correcting pitch angle variations. Thus, non obstacle points can be easily removed since they lay on the road. Generic obstacles are selected by using *Subtractive Clustering* algorithm in a 3D space with an adaptive radius. This clustering technique can be configurable for different types of obstacles. An optimal configuration for pedestrian detection is presented in this work.**

I. INTRODUCTION

The most successful human detection systems from a moving vehicle are being accomplished through the computer vision as the main sensor. The fact that it was the same sensor humans use for driving is not a triviality. It provides the main clues for pedestrian detection although other sensors, such as laser-scanners, have also been tested [1]. Designing pedestrian detection systems for Intelligent Transportation Systems is quite different from detecting and tracking people in the context of surveillance applications because the background is no longer static and pedestrians significantly vary in scale.

The candidate selection method can be implemented by performing an object segmentation either in the 3D scene or in the 2D image plane. The first solution requires the use of stereo vision, while the second one tackles the problem of object detection using a single camera. Monocular approaches constitute a cheap solution, are less demanding from the computational point of view and they have an easier calibration maintenance process. On the contrary, the main problem with candidate selection methods in monocular systems is that, in average, they are bound to yield a large amount of candidates per frame, in order to ensure a low number of pedestrians that are not selected. In [2] a monocular attention mechanism generates up to 75 windows per frame which are fed to the classifier as potential pedestrians. Another problem in monocular systems is the fact that depth clues are lost unless some constrainsts are applied, such as the flat terrain assumption, which is not always applicable, specially in urban areas where vehicles are exposed to changes in their pitch angle due to braking, accelarating, bumped pedestrian and pelican crossings, etc. Not-flat road assumption becomes compulsory for a robust object detection algorithm. It was introduced by non-flat road approximations by series of planar surface sections over the v-disparity map [11], [12]. In [18] road vertical profile is modelled with a clothoid curve fitting directly on the detected 3D road surface points. A monocular pedestrian detection system, using vertical symmetries, is proposed in [3]. After a refinement process based on maximizing the symmetry and vertical edges density, along with a stereo refinement, they fit the bounding boxes by releasing candidates which might be pedestrians [4],[5]. In a later work [6] the stereo vision is used for both the bounding box validation and the pedestrian position estimation.

Among the frameworks that use stereo vision for candidate selection we emphasize the next three. In [8] a stereo vision system to generate 3D representation of the scene with disparity maps, is propounded. The candidates are classified as pedestrian or non-pedestrian using a trained neural network. Since this is the first stereo approach in the literature, the segmentation algorithms are very basic. In [9] an obstacle detection procedure is done by using a multiplexed depth map, and selecting regions of interest whose number of depth features exceeds a percentage of the window area. Then they extract edge images and match them to a set of learned examples using chamfer distance [10]. In order to extract information from 3D scene in [11] a segmentation based on v-disparity maps is performed. The information for performing generic obstacles detection is defined with vertical lines. This implies managing very little information to detect obstacles, which may work well for big objects detection, such as vehicles [12], but might not be enough for small, thin objects detection, such as pedestrian, especially in city traffic due to the heavy disparity clutter.

Infrared cameras have been also tested with both monocular [13][14] and stereo [15][16] vision. These cameras provide better visibility at night but they also provide images with lower resolution, where the appearance of pedestrians is not clear compared to that of day-time images. In addition the use of infrared cameras is quite an expensive option that makes mass production an untractable problem nowadays.

In this work a more comprehensive analysis of the candidate selection method developed in [7] is presented. A stereo vision based method for generic obstacle detection in the framework of ITS is described. A robust correlation process is applied in order to minimize stereo-matching errors. Pitch angle is modelled by a set of planar surfaces over the lateral view. Road points like shadows, lane markings, etc., are separated from the obstacle points. A subtractive clustering attention mechanism is used in order to select candidates as generic obstacles. This technique is applied by using

D. Fernández, I. Parra, M. A. Sotelo, P.Revenga, S.Álvarez and M. Gavilán are with the Department of Electronics of the University of Alcalá. Alcalá de Henares, Madrid, Spain llorca@depeca.uah.es

an adaptive 3D radius with an optimal configuration for pedestrian detection. The main goal is to have an obstacle detection process with low false negative rate (number of pedestrians not detected) and low false positive rate (number of non-pedestrians selected). Thus the number of candidates per frame is reduced and next stages (classification, tracking, etc.) can be done easier.

The paper is organised as follows: section II provides a detailed description of the system. The results achieved up to date are presented and discussed in section III. Finally, section IV summarizes the conclusions and future work.

II. System Description

This section includes the description of the 3D reconstruction, the pitch angle computation for the non-flat road assumption and finally the 3D clustering technique used in order to detect generic obstacles.

A. Non-dense 3D maps

Stereo vision refers to the ability to infer information on the 3D structure and distance of a scene from two or more images taken from different viewpoints. A stereo system must solve two problems: which parts of the left and right images are projections of the same scene element? (*the correspondence problem*) and given a number of corresponding parts, what can we say about the 3D location and structure of the observed objects? (*the reconstruction problem*)[19].

Our stereo vision system consists of two calibrated and synchronized cameras with a Firewire connection and a baseline of approximately 0.3m. These ones are mounted near the rear-view mirror of the vehicle. The distance between the cameras and the front of the car is large enough to use a larger baseline distance since the dead zone does not reach the front of the car. In addition as the camera baseline increases the sensing range increases and the detph resolution becomes denser. Nevertheless the baseline can not be much more larger than 0.3m since the response time becomes intractable.

The calibration process yields the fundamental matrix, F_{lr}, that defines the system epipolar geometry. This way the perfect physically alignment between cameras that implies the assumption of parallel epipolar lines, is not necessary because the stereo correlation process defines mathematically the geometric relationship for the cameras. While image rectification provides a simple area for correspondents and straightforward 3D reconstruction, the general geometry mode, without rectification, provides a better solution since no image resampling has to be done. Moreover, epipolar geometry is precomputed when the application starts, and stored in a lookup table, so that, epipolar lines computation is avoided at runtime. Radial and tangential distortions are eliminated and the intensities of the left and right images are normalized to correct for differences between them. Relevant points in both left and right images are extracted using the well known Canny algorithm, with adaptive thresholds depending on the histogram distribution of the image. By focusing to the image edges the computational cost is reduced and the correlation process improved, since the interest points are placed in non-uniform image areas. In addition, Canny image provides a good representation of the discriminating features of pedestrians: features such heads, arms and legs are distinguishable, when visible, and are not heavily affected by different colors or clothes. Among the wide spectrum of matching techniques that can be used to solve the correspondence problem we implemented the *Zero mean Normalized Cross Correlation* because of its robustness, which can be can be computed as follows:

$$ZNCC(p,p') = \frac{\sum_{i=-n}^{n}\sum_{j=-n}^{n} A \cdot B}{\sqrt{\sum_{i=-n}^{n}\sum_{j=-n}^{n} A^2 \sum_{i=-n}^{n}\sum_{j=-n}^{n} B^2}} \quad (1)$$

where A and B are defined by:

$$A = (I(x+i, y+j) - \overline{I(x,y)}) \quad (2)$$

$$B = (I(x'+i, y'+j) - \overline{I(x',y')}) \quad (3)$$

where $I(x,y)$ is the intensity level of pixel with coordinates (x,y), and $\overline{I(x,y)}$ is the average intensity of a $(2n+1) \times (2n+1)$ window centered around that point. As the window size decreases, the discriminatory power of the area-based criterion gets decreased and some local maxima appear in the searching regions (epipolar lines). On the contrary, an increase in the window size causes the performance to degrade due to occlusion regions and smoothing of disparity values across boundaries. According to previous statements and to a computational cost criterion a practical 7×7 correlation window size is selected.

A post-processing is applied in the correlation step in order to increase robustness and reduce noise:

- 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. If the new left matched points are not exactly the same than the original ones, these correspondences are considered as noise (*mutual check*).
- 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 resulting correaltion maps look much more noise-free. In general, the number of correlated points gets decreased by an average of 24% after using mutual check strategy. By using both, mutual check and minimum disparity methods an average of 37% of points are selected as noise.

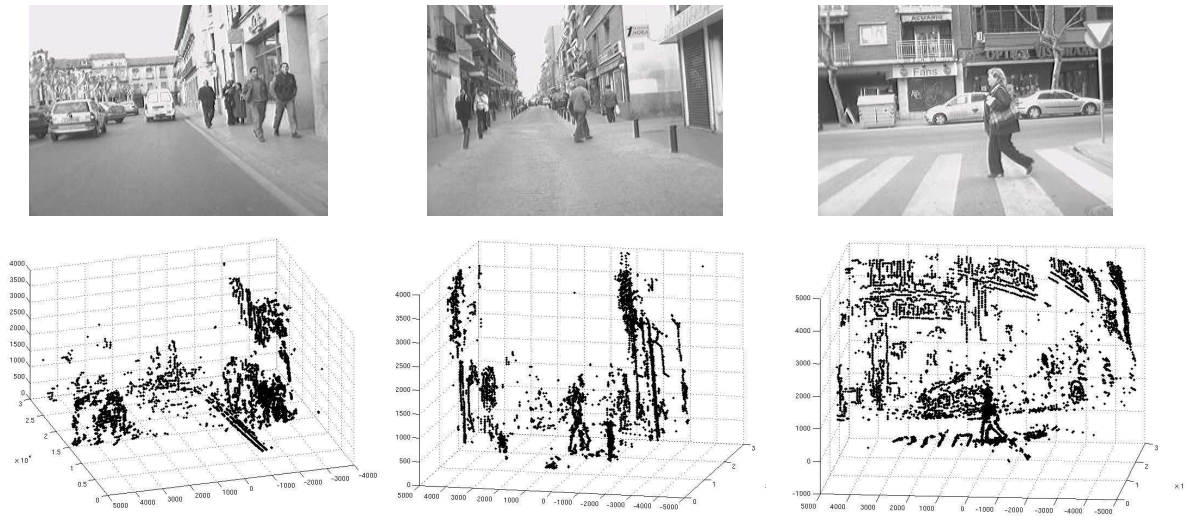

Fig. 1. Upper row: original images, Lower row: non-dense 3D maps

As we know both intrinsic and extrinsic parameters of the stereo system, a scene point can be reconstructed by using a purely algebraic approach: give the projection matrices M_l and M_r and the matching points p_l and p_r, we can rewrite the constraints $s_l p_l = M_l P_l$ and $s_r p_r = M_r P_l$ as

$$\left\{ \begin{array}{ll} p_l \times M_l P_l & = 0 \\ p_r \times M_r P_r & = 0 \end{array} \right. \Leftrightarrow \left(\begin{array}{c} [p_{l\times}]\, M_l \\ [p_{r\times}]\, M_r \end{array} \right) P_l = 0 \qquad (4)$$

This is an overconstrained system of four independent linear equations ($AP_l = b$) in the homogeneus coordinates of $P_l = X_l, Y_l, Z_l$, that is easily solved using the linear least-squares techniques ($P_l = (A^t A)^{-1} A^t b$). After solving both correspondence and reconstruction problems, non-dense 3D maps are created like the ones depicted in figure 1

B. Pitch angle estimation

Detection range 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 view angle, the origin of the world coordinate system is placed at the intersection point between the base plane and the lower boundary of the vertical view angle. Figure 2 depicts the lateral projection of the 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 [18] we consider the vertical displacement due to roll negligible in comparison to the displacement due to pitch. From the point of view of the world coordinate system, 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:

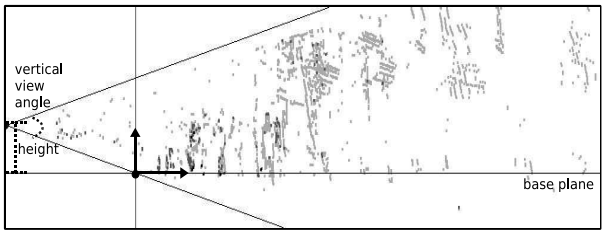

Fig. 2. 3D projected points on the YOZ plane up to 30m

$$\begin{array}{l} \text{for } i = 0 \text{ to } N \\ \quad \text{if } (H(i) > \frac{2}{3}\bar{h} \text{ and } H(i+1) > \frac{2}{3}\bar{h} \text{ and } H(i+2) > \frac{2}{3}\bar{h} \\ \quad\quad \text{and } H(i), H(i+1), H(i+2) > H_{min}) \\ \quad\quad \text{then } \alpha = \alpha_i; \text{ break;} \\ \quad \text{else } \alpha = 0; \end{array} \qquad (5)$$

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:

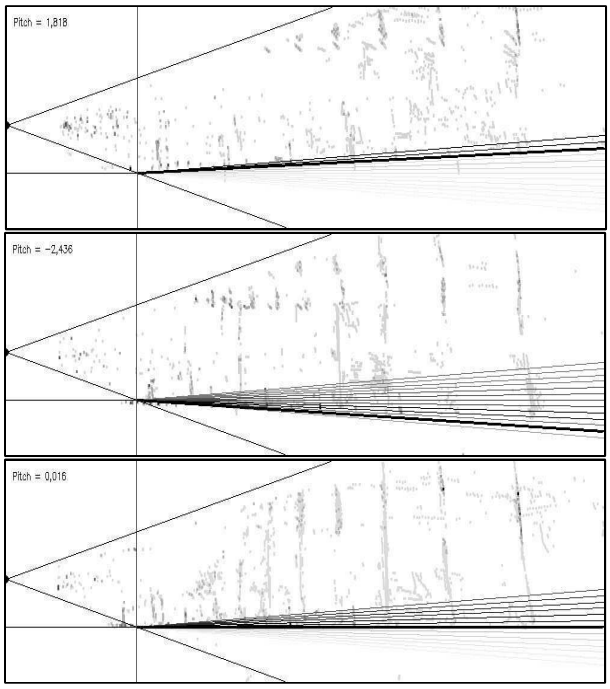

Fig. 3. Pitch angle estimation. From top to bottom: positive pitch angle, negative pitch angle and pitch angle about 0 degrees.

$$\vec{x_k} = \begin{pmatrix} \alpha_k \\ \dot{\alpha}_k \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{k-1} \\ \dot{\alpha}_{k-1} \end{pmatrix} + \vec{r_k} \text{ state eq. (6)}$$

$$z_k = \alpha_k + \vec{o_k} \text{ measurement eq.} \quad (7)$$

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, a transformation matrix has to be applied in order to perform 3D points correction:

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{pmatrix} \quad (8)$$

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.

C. Adaptive 3D Subtractive Clustering

Data clustering techniques are related to the partitioning of a data set into several groups in such way that the similarity within a group is larger than among groups. Normally the number of clusters is known beforehand. This is the case of *K-means* based algorithms. The needed clustering technique should be subject to some constraints:

- The number of clusters is considered unknown, since no a priori estimate of the number of pedestrians in scene can be reasonably made.
- The effects of outliers have to be reduced or completely removed in order to absorb correlation errors.
- It is necessary to define specific space characteristics in order to group different pedestrians in the scene.

For these reasons, a *Subtractive Clustering* method [17] is proposed, which is a well known approach in the field of *Fuzzy Model Identification Systems*. The clustering is carried out in the 3D space, based on a density measure of data points. The idea is to find high density regions in 3D space. Objects in the 3D space are roughly modelled by means of Gaussian functions. It implies that, on principle, each Gaussian distribution represents a single object in 3D space. Nonetheless, objects that get too close from each other can be modelled by the system as a single one and, thus, represented by a single Gaussian distribution. The complete representation is the addition of all Gaussian distributions found in the 3D reconstructed scene. Accordingly, the parameters of the Gaussian functions are adapted in line with the depth by the clustering algorithm, so as to best represent the 3D coordinates of the detected pixels.

The 3D coordinates of all detected pixels are then considered as candidate clusters centers. Thus, each point p_i with coordinates (x_i, y_i, z_i) is potentially a cluster centre whose 3D spatial distribution D_i is given by the following equation:

$$D_i = \sum_{j=1}^{N} exp\left(- \frac{(x_i - x_j)^2}{\left(\frac{r_{ax}}{2}\right)^2} - \frac{(y_i - y_j)^2}{\left(\frac{r_{ay}}{2}\right)^2} - \frac{(z_i - z_j)^2}{\left(\frac{r_{az}}{2}\right)^2} \right) \quad (9)$$

where N represents the number of 3D points contained in a neighborhood defined by radii $r_a = (r_{ax}, r_{ay}, r_{az})$. Cluster shape can then be tuned by properly selecting the parameters r_{ax}, r_{ay}, r_{az}, which are related to 3D actual dimensions. As can be observed, candidates p_i surrounded by a large number of points within the defined neighborhood will exhibit a high value of D_i. Points located at a distance well above the radious defined by r_a will have almost no influence over the value of D_i. Equation 9 is computed for all 3D points measured after stereo reconstruction. Let $p_{cl} = (x_{cl}, y_{cl}, z_{cl})$ represent the point exhibiting the maximum density denoted by D_{cl}. This point is selected as the cluster centre at the current iteration of the algorithm. Densities of all points D_i are corrected based on p_{cl} and D_{cl}. For this purpose, the subtraction represented in equation 10 is computed for all points.

$$D_i = D_i - D_{cl} exp\left(- \frac{(x_i - x_{cl})^2}{\left(\frac{r_{bx}}{2}\right)^2} - \frac{(y_i - y_{cl})^2}{\left(\frac{r_{by}}{2}\right)^2} - \frac{(z_i - z_{cl})^2}{\left(\frac{r_{bz}}{2}\right)^2} \right) \quad (10)$$

where parameters $r_b = (r_{bx}, r_{by}, r_{bz})$ define the neigborhood where the correction of points densities will have the largest influence. The density of data point which is close to the first cluster centre will be reduced, so that these data points can not become next cluster centre. Normally, parameters (r_{bx}, r_{by}, r_{bz}) are larger than (r_{ax}, r_{ay}, r_{az}) in order to prevent closely spaced cluster centres. Commonly

let $r_{bx} = 1.5r_{ax}$, $r_{by} = 1.5r_{ay}$ and $r_{bz} = 1.5r_{az}$. After comprehensive experiments the parameters r_{ax} and r_{ay} have been set to $r_{ax} = 0.8m$ and $r_{ay} = 1.5m$, while an adaptive value has been proposed to r_{az} taking into account the depth resolution of the stereo sensor. Even though stereo geometric relationship is known, for this case, we suppose the depth computation as follows:

$$z = \frac{f_x B}{d_x} \tag{11}$$

where B is the stereo baseline length, f_x is the focal length expressed in units of horizontal pixels and $d_x = x_l - x_r$ is the horizontal disparity in pixels. Depth resolution is computed by using next equation:

$$\Delta z_i = f_x B \quad \frac{1}{d_{xi}} - \frac{1}{d_{xi} - 1} \quad = f_x B \quad \frac{1}{d_{xi}^2 - d_{xi}} \tag{12}$$

According to equations 11 and 12 the adaptive value of r_{az} for a 3D point $p_i = (x_i, y_i, z_i)$ is given by:

$$r_{az} = 2\Delta z_i = 2\frac{z_i^2}{f_x B + z_i} \tag{13}$$

After the subtraction process, density corresponding to the cluster center p_{cl} gets strongly decreased. Similarly, densities corresponding to points in the neighborhood of p_{cl} get also decreased by an amount that is a function of the distance to p_{cl}. All these points are associated to the first cluster computed by the algorithm and will have almost no effect in the next step of the subtractive clustering. In fact these points are *subtracted* and restored as a 3D candidate. After the correction of densities, a new cluster center $p_{cl,new}$ is selected, corresponding to the new density maximum $D_{cl,new}$ and the process is repeated whenever the condition expressed in equation 14 is met.

$$\text{if } U_{rel} > \frac{D_{cl}}{D_{cl,new}} \text{ and } D_{cl,new} > U_{min} \Rightarrow \text{new cluster} \tag{14}$$

where U_{rel} and U_{min} are experimentally tuned parameters that permit to define a termination condition based on the relation between the previous cluster density and the new one, and a mimimum value of the density function. The process is repeated, *subtracting* the points of each new cluster, until the termination condition given by equation 14 is not met.

After applying subtractive clustering to a set of input data, each cluster finally represents a 3D candidate. 2D candidates are selected by projecting the 3D points over one of the images and computing the box that bounds these points.

III. Results

The system has been implemented on a Pentium IV at 2.4 Ghz running GNU/Linux Operating System. Using 320×240 pixel images, the complete algorithm runs at an average rate of 20 frames/s, depending on the number of obstacles being detected and their positions. The candidate selection method has proved to be robust in various illumination conditions, thanks to the adaptive canny edge detector, and distances up to 30m. Pitch angle correction yields more accurate depth measure, but the main advantage consists in the fact that road lane markings, shadows, etc., can be easily detected as non obstacle points without removing points from actual obstacles, like pedestrians. If road points are not removed, these ones will affect negatively the clustering process. On the other hand if road points are not correctly removed, obstacle points could be incorrectly eliminated and not taken into account in the clustering process. Figure 4 depicts two examples of the candidate selection method with and without pitch angle correction. From the point of view of the classification step, in a global pedestrian detection system like in [7], 2D candidates selected in the upper row of figure 4 will consider relevant information of the obstacle as non obstacle (road) information. In figure 5 the Kalman filter estimation for pitch angle in a sequence is shown.

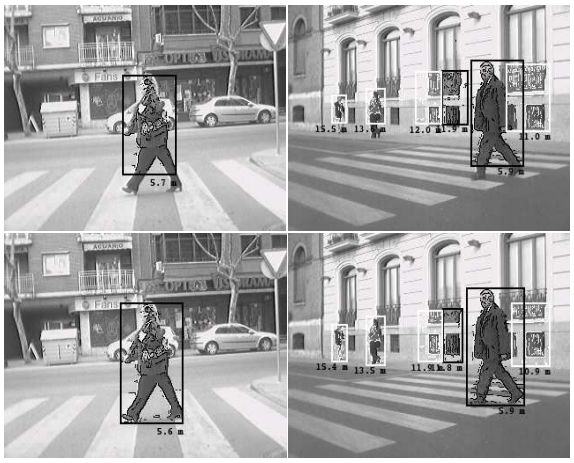

Fig. 4. Obstacle detection examples. Upper row: without pitch angle estimation; lower row: with pitch angle correction

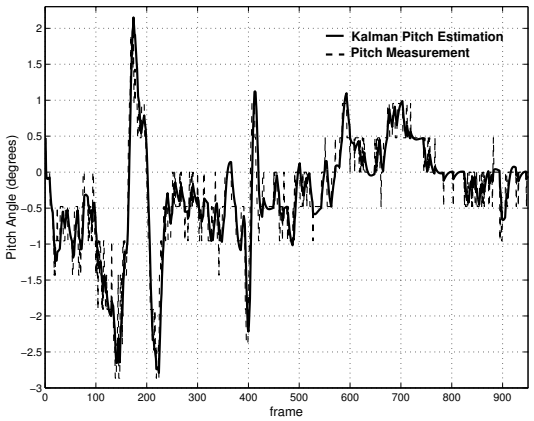

Fig. 5. Kalman filter estimation and measurement for pitch angle in a sequence

The subtractive clustering process yields excellent results. Only 3D volumes similar to pedestrians are selected as possible candidates like trees, bins, lamposts, motorbikes, etc. Cars are usually splitted in two or three parts. This clustering technique can be used for different kind of objects by

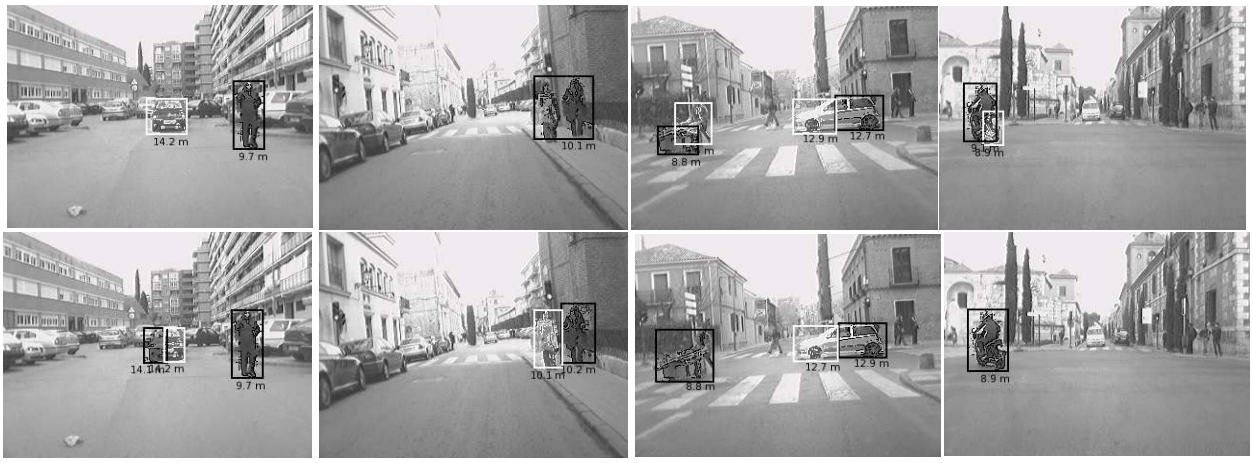

Fig. 6. Upper row: several examples of oversized r_{ax} radius and non-adaptive r_{az} one. Obstacles can be incorrectly separated in two parts due to stereo depth resolution constraints and two obstacles can be merged in one because an oversized value of r_{ax}. Lower row: the same examples correctly detected by using an adaptive value of r_{az} and a correct value of r_{ax}.

configuring its 3D radius. As the parameter r_{ax} increases the probability of merging two pedestrians in one increases. On the other hand a very small value of r_{ax} causes undesirable partitions by dividing one pedestrian in two. The adaptive value of r_{az} corrects obstacles which are splitted in two parts, due to stereo detph resolution constraints. On average our candidate selection mechanism generates 8 windows per frame which is a reasonable amount of obstacles for being classified and tracked in next stages. Figure 6 depitcs several examples with both incorrect (upper row) and correct (lower row) configurations for the radii.

IV. Conclusions and Future Works

In this paper we have described a stereo vision based obstacle detection system with the aim of creating a candidate selection method for a global pedestrian detection system. Non dense 3D maps were created by using edge points. A robust correlation method reduced the amount of stereo-matching errors. Pitch angle was estimated so that a correct obstacle-road separation was possible. In addition depth accuracy was increased and the number of points with which pedestrians were detected was improved. An adaptive subtractive clustering technique has proved to be robust in order to detect generic obstacles with volumes similar to pedestrians. As a future work a hardware implementation of the system would be advisable with the aim of using higher image resolution, without losing performance, and thus more accurate depth measures and higher range values could be obtained.

V. Acknowledgments

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

References

[1] Fuerstenberg, K.C., Dietmayer, K.J., Willhoeft, V., "Pedestrian recognition in urban traffic using a vehicle based multilayer laserscanner." *In.Proc. IEEE Intelligent Vehicles Symposium* France,Jun.2002.

[2] Sashua, A., Gdalyahu, Y., Hayun, G., "Pedestrian detection for driving assistance systems: single frame classification and system level performance" *In Proc. IEEE Intelligent Vehicles Symposium* pp1-6,Italy,2004.

[3] Broggi, A., Bertozzi, M., Fascioli, A., Sechi, M., "Shape-based Pedestrian Detection" *In Proc. IEEE Intelligent Vehicles Symposium* USA,Oct.2000.

[4] Bertozzi, M., Broggi, A., Fascioli, A., Lombardi, P., "Vision-based Pedestrian Detection: will Ants Help?" *in Proc. IEEE Intelligent Vehicles Symp*, vol1,pp1-7,Jun.2002.

[5] Bertozzi, M., et al. "Shape-based pedestrian detection and localization" *In Proc. IEEE Intelligent Transportation Systems*, vol1,pp328-333,Oct.2003.

[6] Lefe, D., Mousset, S., Bensrhair, A., Bertozzi, M., "Cooperation of Passive Vision Systems in Detection and Tracking of Pedestrians" *In Proc. IEEE Intelligent Vehicles Symposium* Italy,2004.

[7] Sotelo, M.A., Parra, I., Fernández, D., Naranjo, E. "Pedestrian Detection using SVM and Multi-feature Combination", *In Proc. IEEE Intelligent Transportation Systems C.*,, Canada,Sep.2006.

[8] Zhao, L., Thorpe, C., "Stereo and neural network-based pedestrian detection" *In Proc. IEEE Intelligent Transport Systems,* vol1.No3,Sep.2000.

[9] Gavrila, D.M., Giebel, J., Munder, S., "Vision-Based Pedestrian Detection: The PROTECTOR System" *In Proc. IEEE Intelligent Vehicles Symposium,* Italy,2004.

[10] Gavrila, D.M., Philomin, V. "Real-time object detection for smart vehicles" *In Proc. Seventh IEEE International Conference Computer Vision,* vol1,pp87-93,Greece,1999.

[11] Grubb, G., Zelinsky, A., Nilsson, L., Rilbe, M., "3D Vision Sensing for Improved Pedestrian Safety" *In Proc. IEEE Intelligent Vehicles Symposium* Italy,2004.

[12] Labayrade, R., Aubert, D., Tarel, J.P, "Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through V-disparity Representation" *In Proc. IEEE Intelligent Vehicle Symposium,* France,June 2002.

[13] Bertozzi, M., et al. "Pedestrian detection in infrared images" *In Proc. IEEE Intelligent Vehicle Symposium,* pp662-667,Columbus,June 2003.

[14] Xu, F., Liu, X., Fujimura, K., "Pedestrian detection and tracking with night vision" *In IEEE Transactions on Intelligent Transportation Systems,* vol6,No1,March 2005.

[15] Liu, X., Fujimura, K., "Pedestrian detection using stereo night vision" *In IEEE Transactions on Vehicular Technology,* vol53,No6,Nov 2004.

[16] Bertozzi, M., Broggi, A., Lasagni A., Del Rose, M., "Infrared Stereo Vision-based Pedestrian Detection" *In Proc. IEEE Intelligent Vehicle Symposium,* USA,Jun. 2005.

[17] Chiu, S., "Fuzzy model identification based on cluster estimation" *in J. of Intelligent and Fuzzy Systems,* Vol2,No3,pp267-278,1994.

[18] Nedevschi, S., et al. "High Accuracy Stereovision Approach for Obstacle Detection on Non-Planar Roads" *in IEEE Inteligent Engineering Systems,*, pp.211-216,Romania 2004.

[19] Trucco, M., Verri, A. "Introductory Techniques for 3-D Computer Vision", *Prentice Hall*, 1998