Proceedings of the
IEEE Intelligent Vehicles Symposium 2000
Dearbon (MI), USA • October 3-5, 2000

Shape-based Pedestrian Detection*

A. Broggi[1] M. Bertozzi[2] A. Fascioli[2] M. Sechi[1]

[1]Dip. Informatica e Sistemistica
via Ferrata, 1
27100, Pavia
e-mail: alberto.broggi@unipv.it

[2]Dip. Ing. dell'Informazione
Parco area delle Scienze, 181A
43100, Parma
e-mail: {bertozzi,fascioli}@ce.unipr.it

Abstract

This paper presents the method for detecting pedestrian recently implemented on the ARGO vehicle. The perception of the environment is performed through the sole processing of images acquired from a vision system installed on board of the vehicle: the analysis of a monocular image delivers a first coarse detection, while a distance refinement is performed thanks to a stereo vision technique.

1 Introduction

For an on-board vision system to be able to assist the driver not only during motorway driving but also within urban environments, besides the need for recognizing obstacles, such as sedans and trucks, the ability of detecting pedestrians is essential to avoid dangerous traffic situations. This work presents the method for detecting pedestrians implemented on the ARGO vehicle.

ARGO is an experimental autonomous vehicle equipped with vision systems and featuring automatic steering capability [1]. The main target of the ARGO Project is the development of an active safety system which can also act as an automatic pilot for a standard road vehicle. ARGO is able to determine its position with respect to the lane, to compute road geometry, to detect generic obstacles on the path, to localize a leading vehicle, and has recently been enhanced with the ability of detecting pedestrians.

Vision-based detection of moving pedestrians is a non-trivial task in case of a non-stationary camera. In fact, the observer's ego-motion entails additional motion in the background, and illumination changes are to be taken into account as well. These problems are critical for approaches which detect human movements by subtracting subsequent frames [2]. Also optical flow techniques are difficult to be applied due to the non-rigid motion of pedestrians.

Several approaches base the recognition process which discriminates pedestrians from other objects on motion cues. In [3] an integration of texture and contour information extracted from the images, along with motion patterns of limb movements, sustain the generation of hypothesis. The quasi-rigid part of the body is tracked and the final classification is based on a temporal analysis of the walking process. In [4] a real-time stereo algorithm is used for the detection and tracking of image region possibly containing a pedestrian; a time delay neural network is then used for a classification based on the typical motion patterns of a pedestrian's leg. These approaches generally assume that the pedestrian's leg are visible and are limited to walking people only.

Conversely, shape-based techniques allow the recognition of both moving and stationary pedestrians. The main difficulties for these methods are given by the wide variety in pedestrian appearance, since they rely on shape features. In [5] an SVM classifier is trained with local wavelet features deriving from a set of training examples. There is a trade-off for this system between the accuracy of the classification and the processing speed. In [6] a stereo-based segmentation algorithm is used to extract objects from the background, followed by a neural network-based recognition. Stereo vision allows short and middle distance detection and the method proves to be robust.

In this paper a novel approach is described which exploits the morphological characteristics and the strong vertical symmetry of the human shape for the detection and recognition of pedestrians. This method allows to identify pedestrians in various poses, positions and clothing, and is not limited to walking people. Section 2 presents a detailed description of the algorithm's steps, section 3 discusses the experimental results, while section 4 summarizes and concludes the paper.

2 Pedestrian Detection

The Pedestrian Detection functionality is aimed at sensing and localizing the objects with a human shape.

*This work was partially supported by the Italian National Research Council (CNR) in the framework of the MADESS2 Project.

Pedestrians are detected through a search for objects featured by specific characteristics, using a single monocular image sequence. Stereo vision is however exploited in certain steps of the algorithm where the understanding of the objects' distance is concerned. This allows strengthening the localization while maintaining low computational complexity.

The Pedestrian Detection algorithm relies on the following hypothesis: a pedestrian is featured by:

- mainly vertical edges with a strong symmetry with respect to the vertical axis,
- size and aspect ratio satisfying specific constraints,
- and is generally placed in a specific region.

Given these assumptions, the localization of pedestrians proceeds as follows: first an area of interest is identified on the basis of perspective constraints and practical considerations (the detection of a person so close to the vehicle to appear only partially in the image cannot be of any use). Then the vertical edges are extracted and, after eliminating the objects belonging to the background, the areas which present high vertical symmetry are considered. Too uniform areas are recognized by evaluating the edges' entropy and immediately discarded, while for the remaining candidates a rectangular bounding box is determined by finding the object's lateral and bottom boundaries and localizing the head through the match with a simple model encoding a pedestrian's head. Distance assessment is then performed: the evaluation deriving from the position of the bounding box' bottom border is refined thanks to a simple stereo vision technique. Finally the pedestrian candidates are filtered: only the ones which satisfy specific constraints on the size and aspect ratio and present sufficiently differentiated (i. e. non uniform) composition are selected and labelled as pedestrians. Temporal correlation is taken into account only in certain steps by using the results from the previous frame to correct and validate the current ones. Next sections detail the steps involved in the processing.

2.1 Preliminary processing

The acquired grey-level image is down-sampled to a 256×288 pixels size and a specific central region is considered as the area of interest where pedestrians are more likely to be found and the detection is more useful in order to apply avoiding strategies. Objects edges' modulus and phase are extracted by means of a Sobel operator and, considering the edges' phase, two separate binary maps for vertical and horizontal edges are obtained (see figure 1).

At this stage, prior to the analysis of the edges' vertical symmetry, a specific stereo vision based procedure is undertaken aimed at eliminating objects belonging to the background. Since two stereo views of the scene are already available in the system [1], by shifting either of the two by a fixed offset (which can be determined from the system calibration) it is possible to obtain the overlapping of objects which lie at such a distance from the cameras that can be considered infinite. Therefore background objects can be eliminated from the vertical edges map by a logical bitwise AND with a binary mask. This mask is obtained from the application of a positive threshold to the signed difference between the left image and a properly shifted version of the right image, hence it presents black areas in correspondence to the background and white spots in correspondence to near objects in the left image (see figure 2). It is actually computed for the upper part of the image only, since far objects are supposed to be in that area, and it is enhanced by applying a morphological operator which filters out small sized disjunct details while amplifying dense areas (the results of this step is shown in figure 3.b).

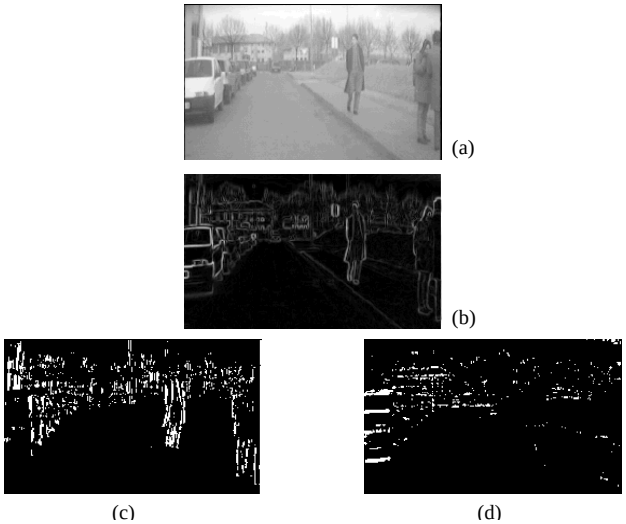

Figure 1: Edges extraction: *(a)* original image; *(b)* edges' modulus; *(c)* vertical edges; *(d)* horizontal edges.

After removing the edges deriving from distant objects and background patterns, the vertical edges map is further processed with a morphological operator aimed at enhancing groups of pixels while discarding single pixels: vertical neighborhood is enforced by marking selected areas with fixed length vertical segments (see figure 3.c).

2.2 Symmetry detection and candidates selection

In order to evaluate vertical edges' symmetry with respect to vertical axes, bi-dimensional maps are computed for different values of the axis' horizontal position within the search area (horizontal axis) and of the symmetry window's width (vertical axis), as shown in figure 3.d. The

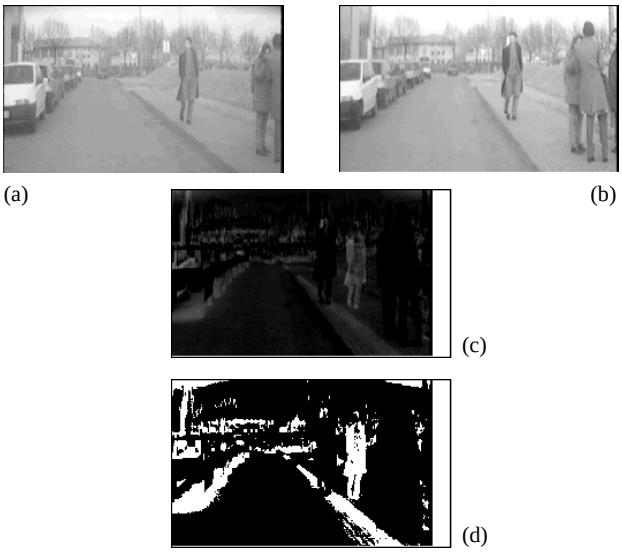

Figure 2: Computation of the map for background elimination: *(a)* and *(b)* left and right images; *(c)* signed difference between left image and shifted right image (negative values are set to zero); *(d)* binarization through a positive threshold.

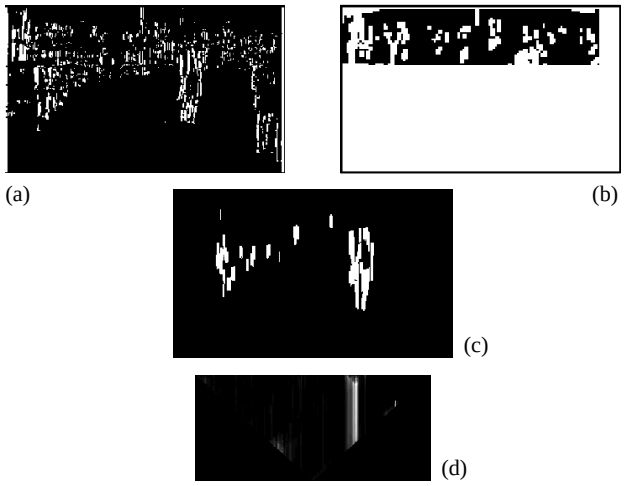

Figure 3: *(a)* Vertical edges map; *(b)* background objects mask; *(c)* amplification of remaining vertical edges; *(d)* vertical symmetry map.

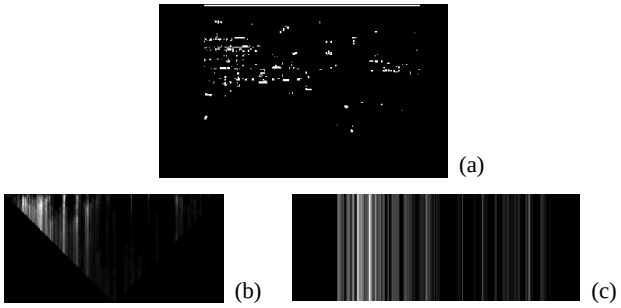

Figure 4: Vertical symmetry *(b)* and linear density maps *(c)* for horizontal edges *(a)*; the high values on the left are given by the presence of vehicles.

maps' triangular shape is due to the limitation in scanning large windows for peripheral vertical axes. Bright points encode the presence of high symmetries.

These maps provide significative indication concerning the position of potential pedestrians, but they may highlight other symmetrical objects too. Pedestrian can however be discriminated thanks to the following considerations: while human shapes exhibit few or null horizontal components, several other objects different from pedestrians (e. g. vehicles) –besides a strong vertical symmetry– present lots of horizontal edges as well. For this reason from the analysis of the horizontal edges two further maps are produced, the former deriving from the evaluation of their vertical symmetry, the latter counting the number of border pixels for each column, both shown in figure 4. These two maps are combined with the previous one to form a single symmetry map using negative experimentally estimated coefficients. Figure 5 shows all component maps and the outcome given by the linear combination.

The resulting map is scanned in order to extract the columns possessing the highest values. Selected columns closer than a given interval are compared and only the dominant one is kept. In fact, very near or partially overlapped pedestrians can be considered as a single individual without diminishing the algorithm's effectiveness. Moreover the positions of the pedestrians detected in the previous frame are also taken into account in the selection. The areas surrounding the axes which have been picked out and ordered are then considered as the candidates that are most likely to represent pedestrians.

Before going on to delimit these areas, a preliminary filtering is performed to get rid of scarcely significant regions that may have been chosen by error. The image entropy is estimated for a stripe surrounding each axis, measuring the density of over-threshold pixels in the previously computed edges' modulus map (as shown in figure 6). This allows to drop symmetrical but highly uniform areas which very likely do not comprise human shapes.

2.3 Bounding box detection

Once found out that a given image area contains a sufficiently structured and symmetrical object, it is necessary to detect its size to determine whether it may represent a pedestrian based on its height, width and aspect ratio. For each object the lateral and bottom borders and the head position are searched, and a rectangular bounding box en-

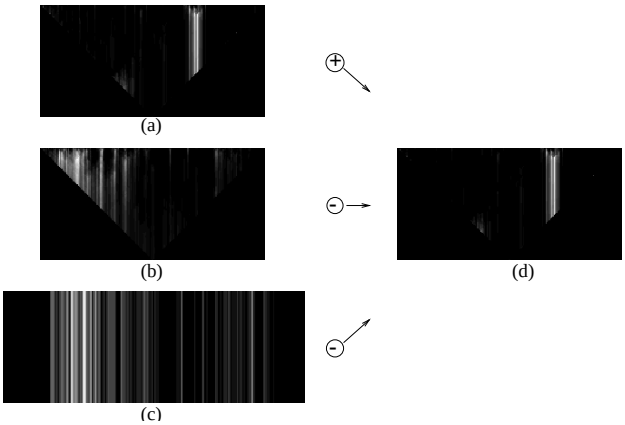

Figure 5: Linear combination of the contribution from vertical and horizontal edges: *(a)* vertical edges' symmetry; *(b)* horizontal edges' symmetry; *(c)* horizontal edges' linear density; *(d)* resulting symmetry map.

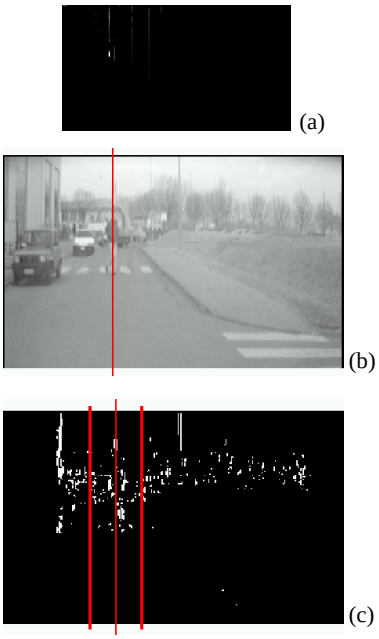

Figure 6: Evaluation of the entropy for an area surrounding the symmetry axis: *(a)* symmetry map; *(b)* original image with symmetry axis; *(c)* search area for entropy computation.

closing the object is built.

The left and right borders are separately searched in the vertical edges map scanning the columns to the left or right of the axis and counting the number of pixels (see figure 7). Each column is given a weight inversely proportional to the distance from the axis in order to avoid the detection of too large boxes in case the object is close to another one. Due to the use of the vertical edges map, lateral borders

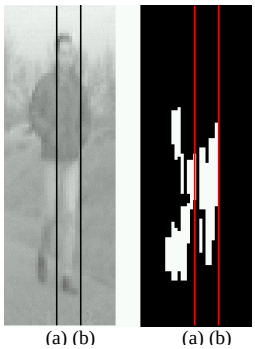

Figure 7: Lateral boundaries determination: *(a)* symmetry axis and *(b)* right boundary.

correspond to the pedestrian's arms only if these are nearly parallel to the body, otherwise the trunk is detected. In either case it can be considered a significant measure of the pedestrian's width.

The pedestrian's head is looked for in the edges' modulus map through the match with a model. A specific binary pattern representing the head and shoulders is matched (at least partially) and localized by a simple correlation function. This reference image is obtained from the selection of the features common to a great deal of sample human shapes, and is scaled to three different sizes (see figure 8). The search area is delimited by the previously found lateral borders and by two fixed horizontal limits deriving from considerations about the position in the image that a standing person's head can reasonably occupy.

Figure 8: Binary model used to localize the pedestrian's head: different sizes are considered.

The bottom boundary is again detected in the vertical edges map: a bottom-ward search is performed, starting from the row containing the maximum number of pixels within a given distance from the axis, and ending with the first empty row.

The distance to the supposed pedestrian can now be computed from the position of the bottom border thanks to the knowledge of the camera calibration. In order to refine this measurement, which is of basic importance for the decisions to be taken by the automatic pilot, an adjustment step is performed taking advantage of a stereo technique. Starting from the distance value estimated from the left image,

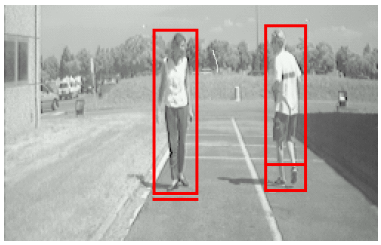

Figure 9: Examples of bounding boxes: the horizontal segment corresponds to the distance computed before the refinement process.

a portion of the right image is searched for a pattern similar to the one enclosed into the bounding box. Once the pattern is detected, the offset between the bounding boxes in the two images is used to compute the distance to the pedestrian. Since not only the search pattern is known, but an estimate of the rough distance –and therefore of the offset as well,– the search is performed in a reduced region of the image only and therefore this step is not as computation intensive as traditional stereo techniques. This procedure relies onto the assumption that the pedestrian can be approximated as a vertical plane; this is acceptable in particular for frontal and rear views of pedestrians. Moreover, the luminance differences in the two stereo views are considered negligible. Figure 9 shows two examples of bounding boxes enclosing pedestrians; the bottom border position has been corrected thanks to the stereo technique.

2.4 Candidates filtering and final selection

Pedestrian candidates can now be filtered to select the objects which show the best affinity with the human shape's morphological characteristics. The bounding box' height and width are checked against experimentally determined limit values. The maximum value is chosen to be linearly dependent on the object's distance to the camera and the minimum value is proportional to the maximum. The aspect ratio is compared to the expected range of values for a standing human shape. This allows to get rid of the objects whose aspect ratio does not belong to a typical range.

A further filtering is performed to discard the areas which are too homogeneous to represent a pedestrian. Since a bounding box has been defined, the measure of the object's entropy is now more significative than before, and is used as the final decisive criterium to judge the survived candidates, taking into account the results of the previous frame as well. If one of the objects under examination can be associated to a pedestrian detected in the previous frame thanks to the axes' closeness, it is also possible to verify that its size, position and entropy are compatible with the values assumed in the previous frame.

3 Results and discussion

Figure 10 presents the results of the pedestrian detection algorithm in a number of different situations, ranging from simple scenes where pedestrians are the only objects, to more complex urban environments where several objects (such as cars, motorbikes, trees and buildings) appear both near and in the background, or even cases where adjacent or partially overlapped pedestrians are present (as in figure 10.h and 10.k).

The discrimination of pedestrian from other objects based on the vertical edges' prevalence and symmetry proved to be robust. Moreover the localization of human shapes standing at different distances shows a good precision thanks to the refinement of the stereoscopic technique (see figure 10.a, 10.b, 10.c, and 10.l).

The algorithm's weakest point regards the bounding box construction, which is sometimes not very accurate, in particular concerning the head's position (as shown in figure 10.j). An improvement in the lateral borders and head localization procedure would permit to shrink the currently quite broad size limits in the selection phase, thus making more robust the distance refinement procedure. In fact, when the bounding box does not completely or accurately delimit the pedestrian, the window used to compute the stereo correlation does not enclose the features of interest only, but also the background. This leads to a less precise distance computation (e. g. in figure 10.f and 10.g).

Wrong results are rare and could be eliminated thanks to a more thorough use of the temporal correlation between subsequent frames. For example in figure 10.i, even if a strong vertical symmetry was correctly detected, a low entropy measure led to discard the candidate.

Furthermore, when the pedestrian is very close to the camera a multiple detection may occur due the symmetry widening and fragmentation (see figure 10.d and 10.e). However the algorithm's effectiveness does not go invalidated.

4 Conclusions

In this paper a vision-based algorithm aimed at the detection of pedestrian has been presented. It has been integrated in the ARGO vehicle and tested in urban environments. The algorithm is based on the localization of human shapes, based on symmetry, size, ratio, and shape. It requires the whole pedestrian to be present in the image (even if it has proven to work also when the pedestrian is partly occluded by other pedestrians), at a distance ranging from 10 to 40 meters. Initially, a coarse detection of pedestrians is computed through the processing of a single image. Then, a distance refinement is performed using a simplified stereo-vision technique.

Figure 10: Pedestrian detection results: images show the symmetry search area (delimited by black corners) and the detected pedestrians, each with the symmetry axis (in white) and bounding box (in black). The horizontal black segment shows the distance obtained from the monocular analysis, prior to stereo refinement. The trapezoidal white line delimits the road surface portion visible by both cameras.

The algorithm has been tested both in laboratory and on board of ARGO. Preliminary results show that Pedestrian Detection is quite accurate even in the case of complex scenarios. An extension to the discussed algorithm to fully exploit the high temporal correlation of consecutive frames in sequences and, therefore, enhance the robustness and reliability is currently under development.

References

[1] A. Broggi, M. Bertozzi, A. Fascioli, and G. Conte, *Automatic Vehicle Guidance: the Experience of the ARGO Vehicle*. World Scientific, Apr. 1999. ISBN 981-02-3720-0.

[2] K. Rohr, "Towards Model-based Recognition of Human Movements in Image Sequences," *CVGIP: Image Understanding*, vol. 59, pp. 94–115, Jan. 1994.

[3] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking Pedestrian Recognition," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems'99*, (Tokyo, Japan), pp. 292–297, Oct. 1999.

[4] C. Wöhler, J. K. Aulaf, T. Pörtner, and U. Franke, "A Time Delay Neural Network Algorithm for Real-time Pedestrian Detection," in *Procs. IEEE Intelligent Vehicles Symposium'98*, (Stuttgart, Germany), pp. 247–251, Oct. 1998.

[5] C. Papageorgiou, T. Evgeniou, and T. Poggio, "A Trainable Pedestrian Detection System," in *Procs. IEEE Intelligent Vehicles Symposium'98*, (Stuttgart, Germany), pp. 241–246, Oct. 1998.

[6] L. Zhao and C. Thorpe, "Stereo- and Neural Network-based Pedestrian Detection," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems'99*, (Tokyo, Japan), pp. 298–303, Oct. 1999.