

# Pedestrian Detection for Intelligent Vehicles based on Active Contour Models and Stereo Vision

C. Hilario, J. M. Collado, J. M<sup>a</sup> Armingol, A. de la Escalera

Intelligent Systems Lab. Universidad Carlos III de Madrid. Leganes, Madrid, 28911. Spain  
{chilario, jcollado, armingol, escalera}@ing.uc3m.es

**Abstract.** Recently, the focus of safety systems for intelligent vehicles has been on researching and developing Advanced Driver Assistance Systems (ADAS). Most efforts have been concentrated at the driver, not taking into account the protection of the most vulnerable road users. This paper describes a pedestrian detection algorithm based on stereo vision. The use of visual information is a promising approach to cope with the different appearances of pedestrians and changes of illumination in cluttered environments. Active contour models are used to detect and track people from the images taken by an on-board vision system, performing contour extraction in sequential frames.

## 1 Introduction

### 1.1 Motivation

Over the past 20 years, the high rate of road-accidents all over the world has motivated the development of intelligent vehicles. The researchers community, the automotive industry and several organizations, have been actively involved in improving road safety through the development of ADAS[1]. However, work has been focussed on the driver, whilst the protection of pedestrians has been relegated [2].

Projects that have dealt with this case are quite recent, as it has been pointed out at the Fifth Framework Programme [3]. A possible reason for it could be the fact that detecting pedestrians with an artificial system is a difficult task. The main challenges are the high degree of variability of the human appearance, the cluttered backgrounds and the changing lighting conditions. Moreover, the applications to protect pedestrians define hard real time requirements. An open issue is which sensors are best to address this complexity. Distance sensors, like radar or laser, have the advantage of giving a direct distance measurement. Among the disadvantages stand out their lower resolution and their tendency to interfere each other if they are in closeness. On the other hand, computer vision gives a richer description of the environment, although the information is more difficult to process. Even if other sensors as lasers or radars can detect pedestrians, vision is the unique that can comprehend their motion and predict their movements. For the reason that diverse sensors could be complementary, some approaches have decided to integrate them.

The methods to detect pedestrians based on computer vision can be classified in three main groups. Those that try to find simple features that define a person are at the lowest level. Their main drawback is that if one of those features is not enough present in the image, the pedestrian is lost. Besides, they are prone to false tracks. On the other hand, there are methods that include some kind of learning. Generally, they are based on neural networks. That type of methods require a lot of time to be trained. Model-based approaches, take advantage of the two previous. Usually, a model of the person is built, so they are more robust than feature based algorithms, but slightly slower.

## 1.2 Previous work

Papageorgious and Poggio [4] presented a pedestrian detection system based on wavelet analysis and Support Vector Machines. However, the system was computationally expensive as it had to scan the whole image at multiscales. Gavrilin and Philomin [5] developed a real time pedestrian detection algorithm based on distance transforms. This method performs a coarse-to-fine template matching. But the template hierarchy cannot capture the variety of human shapes. Zhao and Thorpe [6] developed a robust algorithm for detecting pedestrians in cluttered scenes through stereo-based segmentation and neural network-based recognition. Broggi *et al.*[7] also used stereo vision, combining it with a verification technique based on symmetry properties. Both systems got deceived by objects similar to humans. Recently there has been an increasing interest in using infra-red sensors [8]. Although they can detect pedestrians by the heat their bodies emit, pedestrians are not the only sources of heat in a traffic environment.

## 2 The Pedestrian Detection Module

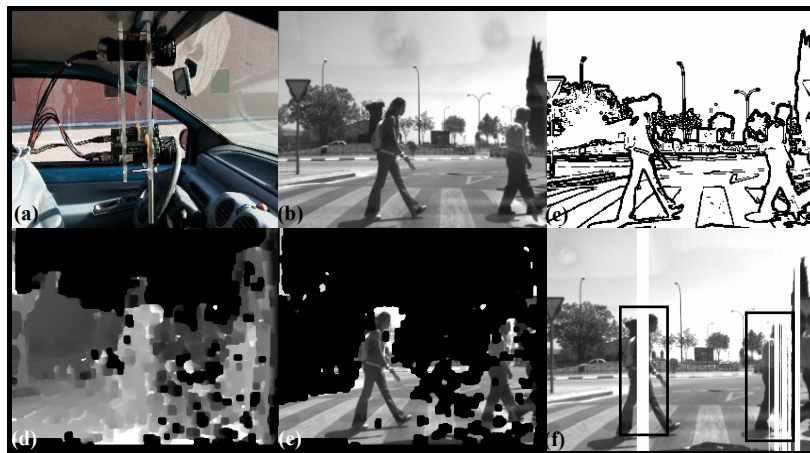
Active contour models or “*snakes*” were proposed by Kass *et al* [9] in 1988 as a segmentation scheme. Its ability to extract contours, even in presence of gaps or occlusions, together with its dynamic behavior, makes this approach adequate for the detection and tracking of non-rigid objects. The main drawback is their high sensibility to the initial position. In order to overcome this limitation, a stereo module is integrated to guide the location of active contours.

### 2.2 Active Contour Models initialization

The motivation for using stereo vision is manifold. When dealing with images taken by a non-static camera, most of the segmentation techniques used for non-moving camera fail due to the movement of the camera. Among the advantages of using stereo vision, it allows occlusion analysis, is robust to illumination changes and can detect both moving and motionless objects.

In the system developed, stereo vision is used to generate a disparity map of the scene (Fig. 1-d). As the pedestrians can appear in the scene at very diverse distances, the use of range-information allows filtering the images based on distance measures.

Therefore, regions that are not at the desired distance are eliminated (Fig. 1-c), performing subsequent calculations only on the filtered areas. Hence, two advantages are obtained; On one hand, the algorithm is less time consuming. On the other hand, the task of initializing the snakes is eased because only the filtered area is considered. Since regions with high vertical symmetry are potential candidates for an active contour initialization, vertical symmetries are looked for. With that aim, the vertical gradient component of the filtered image is found and only pixels with high response are taken. Then, pairs of pixels on the same line vote for central pixels as their symmetry axis. An active contour is initialized in a symmetry axis, if the number of pixels that vote for that axis is over a given threshold (Fig. 1-f).



**Fig. 1.** (a) A detail of the stereo system. (b) A right image taken by the stereo camera. (c) Gradient image. (d) Disparity map (e) and the filtered image based on distance measures. (f) Both vertical symmetries and snakes initialization region enclosed by bounding boxes are shown

### 2.3 Active Contour Model

Active contour models are proposed as energy minimization splines that, from an initial position, are deformed by external and internal forces, until they reach an equilibrium state. The major reason for their success is the possibility to integrate physical and topological knowledge into the segmentation process. Our approach follows the explicit contour representation proposed by Kass *et al.*[9], because it allows an efficient processing and its physical properties can be controlled in a very intuitive manner. In their seminal paper, Kass *et al* model a contour as a rubber band under the influence of image forces and elastic forces. Image forces are due to external energies associated to a potential field that attracts the snake. On the other hand, elastic forces counteract strong expansion and bending of the deformable model. They represent the internal energy of the contour as a weighted combination of membrane and thin plate energy. It is used to regularize the contour and hence to avoid strange effects. The

evolution of the contour is governed by the minimization of both internal and external energies.

The internal energies used in this proposal extend the ones used by Williams and Shah [10]. Their formulae maintain the points in the snake more evenly spaced than Kass, so the natural tendency to shrink of the snakes is mitigated.

$$E_{\text{int}}^*(v(s)) = \alpha(s) \left( \overline{\text{dist}} - \left| \frac{dv}{ds} \right|^2 \right) + \beta(s) \left| \frac{d^2 v}{ds^2} \right|^2. \quad (1)$$

In order to avoid shrinkage, a new internal force is included to control the shape of the deformable model. This regularizing force prevents the shrinking effect of the snake, as it is based on higher degrees of smoothness than the membrane and the thin-plate energy terms, which are based on the first and second derivatives respectively.

$$E_{\text{int}}(v(s)) = E_{\text{int}}^*(v(s)) + \theta(s) \left| \frac{d^4 v}{ds^4} \right|^2. \quad (2)$$

This term is based on the fourth derivatives along the contour and it looks for segments presenting no change on its center of curvature, and therefore are prone to correspond to head and feet areas of a pedestrian. Once those segments are localized, the amount of stretching and bending of them is modified locally.

While the classic active contour model is non-adaptive with respect to the underlying image data, in this algorithm the elasticity and bending properties of the contour are related to the underlying image structure. Firstly, curvature of the model is calculated and depending on its value, the elasticity and bending weights are modified. In general, bending of the snake is not too much allowed. Next, for those segments in the curve that present a slight curvature, the new energy term is calculated. Therefore, the snake is constrained to deform in a particular way.

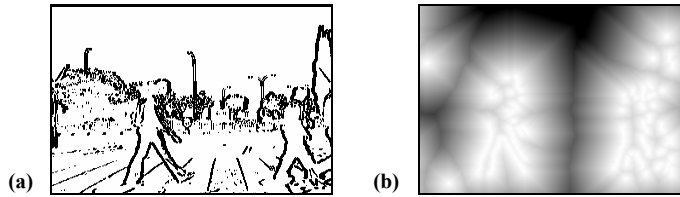


Fig. 2. (a) Vertical gradient and (b) its distance map

For the external forces, a new potential field which smoothly extends over a long distance is defined. So, the snake is affected not only by surrounding features. The fact that pedestrians have a strong vertical symmetry is exploited to construct the potential field. The same idea was used to decide where to put an active contour. Therefore, a distance map of the symmetry axes obtained from that stage is constructed. In order to avoid the snake shrinking to the axis, movement is allowed until it reaches a vertical edge. Besides this potential field, the image gradient (Fig. 1-c) and distances to vertical borders (Fig. 2-b) are also considered.

The deformable model proposed extends the greedy algorithm of Williams and Shah as it is a stable, fast and flexible optimization technique. This approach is adequate for non-rigid objects detection and tracking, performing contour extraction in sequential frames. Once the snake is initialized on an object contour in the first frame, it will automatically track the contour from frame to frame. This method requires small deformation and movement of an object between frames.

Some points in the snake are still prone to errors, like getting trapped into the shadow of the pedestrian (Fig. 3-d). Besides, if the external forces are not strong enough, the snake tends to shrink (Fig. 3-b and 3-e). These problems are a side-effect of the representation used. As the model is only evaluated at some discrete points, these have to be uniformly spaced. Otherwise, the elasticity, curvature and concavity terms are inaccurate. A possible solution could be using splines, as the model is evaluated not only at its control points, but also along the contour.



Fig. 3. From (a) to (f) a complete sequence of frames is shown

## 4 Conclusions and Results

A system based on computer vision for the detection of pedestrians has been presented. It is based on a deformable contour model using a parametric representation. The greedy algorithm is implemented to reach the minimum energy. The evolution of the contour is driven by a new potential based on distances to vertical symmetries and vertical borders. Besides, a regularization term is included in its internal energy, which aims to control the contour shape without producing any shrinkage.

The quality of the segmentation is improved by the information provided by the stereo module. Initial segmentation is performed in the images, filtering them with the data from disparity maps. Regions that are not at the desired distance are eliminated from the images, easing the active contour models initialization and the successive processing.

This algorithm has been tested on images taken by a stereo camera mounted on the Ivvi (Intelligent Vehicle based on Visual Information) vehicle (Fig. 1-a), which is an experimentation platform for researching and developing Advance Driver Assistance System based on computer Vision. The pedestrian detection module is part of this ADAS.

## Acknowledgements

This work was supported in part by the Spanish government under CICYT grant TRA2004-07441-C03-01.

## References

1. McDonald, J., Markham, C., McLoughlin, S.: Selected problems in automated vehicle guidance. Tech. Report NUIM/SS/2001/05 Signals and Systems Group, National University of Ireland (2001)
2. Gavrilu D.M., Kunert M., Lages U.: A multi-sensor approach for the protection of vulnerable traffic participants-the PROTECTOR project. IEEE Instrumentation and Measurement Technology Conference, Vol.3. (2001) 2044-2048
3. Information Society Technologies for Transport and Mobility. Achievements and Ongoing Projects from the Fifth Framework Programme: Office for Official Publications of the European Communities, 2003. <<http://europa.eu.int>>
4. Papageorgiou, C., Evgeniou, T., Poggio, T.: A trainable pedestrian detection system. Proc. of Intelligent Vehicles (1998) 241-246
5. Gavrilu, D.M.: Philomin, V.: Real-time object detection for "smart" vehicles. Proc. of IEEE Intl. Conf. On Computer Vision (1999) 87-93
6. Zhao, L., Thorpe, C.E.: Stereo- and neural network-based pedestrian detection. IEEE Transactions on Intelligent Transportation Systems, Vol. 1. (2000) 148-154
7. Broggi, A., Bertozzi, M., Fascioli, A., Sechi, M.: Shape-based pedestrian detection. IEEE Intelligent Vehicles Symposium (2000) 215-220
8. Meis, U., Oberländer, M., Ritter, W.: Reinforcing the reliability of pedestrian detection in far-infrared sensing. IEEE Intelligent Vehicles Symposium (2004) 779-783
9. Kass, M., Witkin, A., Terzopoulos D.: Snakes: Active Contour Models. Int. J. Comp. Vision, Vol. 1. (1988) 321-331
10. Williams, D.J., Shah, M.: A Fast Algorithm for Active Contours and Curvature Estimation. CVGIP: Image Understanding, Vol. 55. (1992) 14-26