
Pedestrian Shape Extraction by Means of Active Contours

Massimo Bertozzi¹, Alberto Broggi¹, Stefano Ghidoni¹, and Michael Del Rose²

¹ Dip. Ing. Informazione, Parma, Italy

{bertozzi,broggi,ghidoni}@ce.unipr.it

² U.S.Army TARDEC, Warren, MI, U.S.A.

mike.delrose@us.army.mil

Summary. This article presents a shape extraction and results of a preliminary validation stage for a pedestrian detection system based on the use of active contours. The complete system is based on the use of both far infrared and visible cameras to detect areas that potentially contain pedestrians; in order to validate and filter such result a refinement of the human shape by means of active contours is performed followed by a neural network based filtering.

1 Introduction

During the last years, pedestrian detection has been a key topic of the research on intelligent vehicles. This is due to the many applications of this functionality, like driver assistance, surveillance, or automatic driving systems; moreover, the heavy investments made by almost all car manufacturers on this kind of research proves that particular attention is now focused on improving road safety, especially for reducing the high number of pedestrians being injured every year. Also the U.S. Army is actively developing systems for obstacle detection, path following and anti-tamper surveillance, for its robotic fleet [9, 11].

Finding pedestrians from a moving vehicle is, however, one of the most challenging tasks in the artificial vision field, since a pedestrian is one of the most *deformable* object that can appear in a scene. Moreover, the automotive environment is often barely unstructured, incredibly variable, and apparently moving, due to the fact that the camera itself is in motion; therefore, really few assumptions can be made on the scene.

This paper describes a module for human shape filtering developed for integration into a vision-based obstacle detection system to be installed on an autonomous military vehicle. This system is able to detect all obstacles appearing in the scene and is based on the simultaneous use of two stereo camera systems: two far infrared cameras and two daylight cameras [4]. The first stages of this system provide a reliable detection of image areas that potentially contain pedestrians; following stages are devoted to refine and filter this rough results to validate the pedestrians presence. The filtering is based on a multi-vote system; several approaches are independently used to analyze areas of attention, and

each subsystem outputs a vote describing how much the obstacle is likely to be a pedestrian. Then, a final validation is done, based on all votes.

This paper describes the subsystem developed for extracting the object shape, in order to provide a vote based on it. However, the evaluation of the object shape is currently under development, therefore only results on shape extraction are detailed.

This paper is organized as follows: section 2 describes related work in pedestrian detection systems based on artificial vision. The pedestrian detection system is discussed in 3 and the module for active contours-based shape detection algorithm is detailed in section 4. Finally, section 5 ends the paper presenting few results and remarks on the system.

2 Related Work

For the U.S. Army the use of vision as a primary sensor for the detection of human shapes is a natural choice since cameras are non-invasive sensors and therefore do not emit signals.

Vision-based systems for pedestrian detection have been developed exploiting different approaches; like the use of monocular [16, 20] or stereo [17, 6] vision. Many systems based on the use of a stationary camera employ simple segmentation techniques to obtain foreground region; but this approach fails when the pedestrians has to be detected from moving platforms. Most of the current approaches for pedestrian detection using moving cameras treats the problem as a recognition task: a foreground detection is followed by a recognition step to verify the presence of a pedestrian. Some systems use motion detection [15, 6] or stereo analysis [21] as a means of segmentation.

Other systems, substitute the segmentation step with a focus-of-attention approach, where salient regions in feature maps are considered as candidates for pedestrians. In the GOLD system [1], vertical symmetries are associated with potential pedestrians. In [5] the local image entropy directs the focus-of-attention followed by a model-matching module.

For what concerns the recognition phase, recent research are often motion-based, shape-based or multi-cue based. Motion based approaches use the periodicity of human gait or gait patterns for pedestrian detection [6, 5]. These approaches seems to be more reliable than shape based ones, but they require temporal information and are unable to correctly classify pedestrians that are still or have an unusual gait pattern.

Shape-base approaches are based on pedestrians' appearance, so can detect both moving and stationary people [1, 3]. In these cases, the challenge is to model the several variations in the shapes, pose, size and appearance of humans and their backgrounds. Basic shape analysis methods consist in matching a template onto candidate foreground regions. In [10] a tree-based hierarchy of human silhouettes is constructed and the matching follows a coarse-to-fine approach. In [13, 18] probabilistic templates are used to take in account the possible variations in human shape. To accomplish the recognition step, other systems systems

exploit pattern recognition based on the use of classifiers or in combination with a shape analysis with gait detection [10, 14].

3 System Description

The active contours-based algorithm described in this work has been developed as a part of a tetra-vision based pedestrian system [2, 4]. The whole architecture is based on the simultaneous use of two far infrared and two daylight cameras. Thanks to this approach, the system is able to detect obstacles and pedestrians when the use of infrared devices is more appropriate (night, low illumination conditions...) or, conversely, in the case visible cameras are more suitable for the detection (hot, sunny environments...).

In fact, FIR images convey a type of information that is very different from those in the visible spectrum. In the infrared domain the image of an object depends on the amount of heat it emits, namely, it is generally related to its temperature (see fig. 1). Conversely, in the visible domain, objects appearance depends on how the surface of the object reflects the incident light as well as on the illumination conditions.

Since humans usually emit more heat than other objects like trees, background, or road artifacts, the thermal shape can be often successfully exploited for pedestrian detection. In such cases, pedestrians are in fact brighter than the background. Unfortunately, other road participants or artifacts emit heat as well (cars, heated buildings...). Moreover, infrared images are blurred, have a poor resolution and the contrast is low compared with rich and colorful visible images.

Consequently, both visible and far infrared images are used for reducing the search space.

Figure 2 depicts the overall algorithm flow for the complete pedestrian system. Different approaches have been developed for the initial detection in the two image domains: warm areas detection, vertical edges detection, and an approach based on the simultaneous computation of disparity space images in the two domains [2, 4].

These first stages of detection output a list of areas of attention in which pedestrians can be potentially detected. Each area of attention is labelled using a bounding box. A symmetry-based approach is further used to refine this rough result in order to resize bounding boxes or to separate bounding boxes that can contain more pedestrians.

These two steps in the processing, barely take in account specific features of pedestrians; in fact, only symmetrical and size considerations are used to compute the list of bounding boxes. Therefore, independent validators are used to evaluate the presence of human shapes inside the bounding boxes. These stages exploits specific pedestrian characteristics to discard false positives from the list of bounding boxes.

A final decision step is used to balance the votes of validators for each bounding box.



Fig. 1. Examples of typical scenarios in FIR and visible images

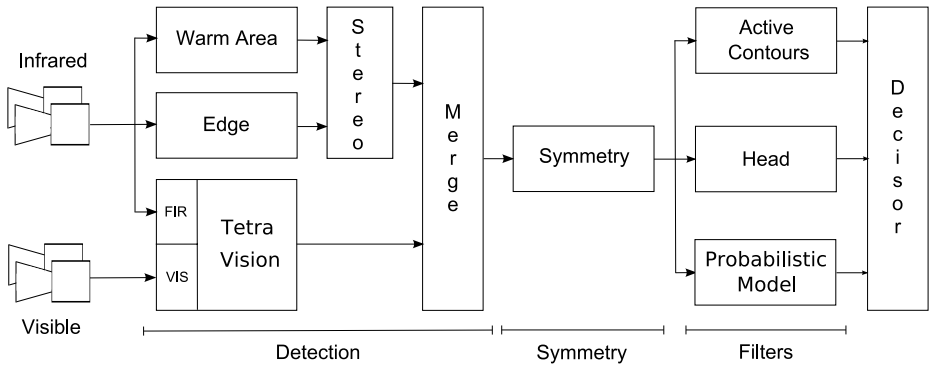


Fig. 2. Overall algorithm flow

4 Shape Extraction

As previously discussed, the pedestrian validation step is composed by several validators, each one supplying a vote, that is then provided to the final evaluation step. The validator detailed in this section is based on the analysis of a pedestrian shape, which can be extracted using the well-known active contour models, also known as snakes.

4.1 Active Contour Models

Active contour models are widely used in pattern recognition for extracting an object shape.

First introduced by Kass et al. [12], this topic has been extensively explored also in the last years. Basically, a snake is a curve described by the parametric equation $\mathbf{v}(s) = (x(s), y(s))$, where s is the normalized length, assuming values in the range $[0, 1]$. This continuous curve becomes, in a discrete domain, a set

of points, that are pushed by some energies that depend on the specific problem being addressed. Every point composing the snake reaches a local energy minimum; this means that the active contour does not find a global optimum position; rather, since it is based on local minimization, the final position strongly depends on the initial condition, i.e. the initial snake position.

Because initial stages of the pedestrian detection system provide a bounding box for each detected object, the snake initial position can be chosen as the bounding box contour; then, a contracting behavior should be impressed, to force the snake to move inside the bounding box. Other energies must also be introduced to make the snake stop when the object contour is reached.

Forces, and associated energies, can be divided into two different categories: internal and external. Internal energy only depends on the topology of the snake, and controls the continuity of the curve derivatives; it is evaluated by the equation:

$$E_{int} = \alpha(s)|\mathbf{v}_s(s)|^2 + \beta(s)|\mathbf{v}_{ss}(s)|^2 \quad (1)$$

where $\mathbf{v}_s(s)$ and $\mathbf{v}_{ss}(s)$ are, respectively, the first and second derivatives of $\mathbf{v}(s)$ with respect to s . The first contribution appearing in the sum represents the tension of the snake, that is responsible for the elastic behavior; the second one gives the snake resistance to bending; $\alpha(s)$ and $\beta(s)$ are weights.

Therefore, internal energy controls the snake mechanical properties, but is independent of the image; external energy, on the contrary, causes the snake to be attracted to the desired features, and should therefore be a function of the image.

Analytically, the snake will try to minimize the whole energy balance, given by the equation:

$$E_{snake} = \int_0^1 (E_{int}(\mathbf{v}(s)) + E_{ext}(\mathbf{v}(s))) ds \quad (2)$$

Because energies are the only way to control a snake, a proper choice of both internal and external energies should be made. In particular, the external energy depending on the image must decrease in the regions where the snake should be attracted. In the following, the energies adopted to obtain an object shape are described.

As previously said, the initial snake position was chosen to be along the bounding box contour. In this system both visible and far infrared images are available, but the latter seems much more convenient when dealing with pedestrians, due to the thermal difference between a human being and the background [4].

To extract a pedestrian shape, the Sobel filter output is a useful starting point; moreover, the edge image is needed also by previous steps of the recognition algorithm, so it is already available. A Gaussian smoothing filter is then applied to enlarge the edges, and therefore the area capable of influencing the snake behavior. The resulting image is then associated with an energy field that pushes the snake towards the edges: for this reason, the brighter a pixel in that image, the lowest the associated energy; in this way, *snavels* (the points into which the snake is discretized) are attracted by the strongest edges, see fig. 3.

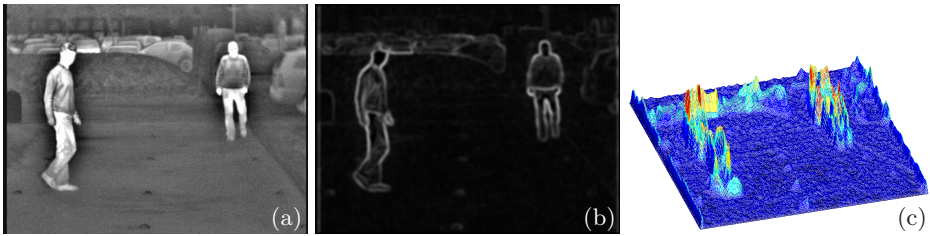


Fig. 3. Energy field due to edges: (a) original image, (b) edge image obtained using Sobel operator and gaussian smooth, (c) edge energy functional with inverted sign, to obtain a more effective graphical representation

Bright regions of the original FIR image are also considered. In fact, smoothed edges does not accurately define the object contour (mainly because they are smoothed), so snake contraction has to be arrested by bright regions in the FIR image, that can belong to a portion of a human body (see fig. 4). This method lets the snake correctly adapt to a body shape in a lot of situations, and it should also be noticed that this mechanism works only if there are hot regions inside the bounding box; a useful side effect, then, is an excessive snake contraction when there are not warm blobs inside a bounding box

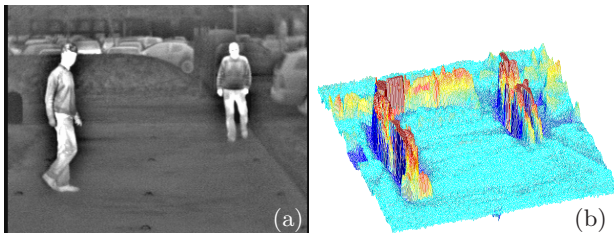


Fig. 4. Energy field due to the image: (a) original image, (b) intensity energy functional with inverted sign, to obtain a more effective graphical representation

The minimum energy location is found by iteratively moving each snaxel, following an *energy minimization algorithm*. Many of them were proposed in the literature. For this application, the greedy snake algorithm [19], applied on 5×5 neighborhood, was adopted.

During the initial iterations, the snake tends to contract, due to the elastic energy; this tendency stops when some other energy counterweights it: for instance, the presence of edges, or a light image region. While adapting to the object shape, the snake length decreases, as well as the mean distance between two adjacent snaxels. Since this mean distance is a value that affects the internal energy, to keep almost constant the elastic property also during strong contraction, the snake is periodically resampled using a fixed step; in this way some unwanted snaxels accumulation can be avoided.

5 Discussion

The developed system has been tested in different situations using an experimental vehicle equipped with the tetra-vision system. As previously said, the contour extraction technique has been developed for pedestrian validation; this means that the snake has to be further processed in order to obtain a vote, say, the probability that it contains a pedestrian. This part of the system is still under development; however, some work has already been done: the first idea was to exploit a shape-matching algorithm, like those described in [7,8], to compare the final shape of the snake with a pre-built database of typical pedestrian shapes. This technique turned out to be not very effective, because of the high number of the false positives it generates. In fact, pedestrians can be seen in a wide variety of poses and postures, so that it is difficult to describe what a *pedestrian shape* is; for this reason, a high number of different shapes have to be considered, therefore causing a lot of false positives.

The second approach chosen for pedestrian classification is based on a neural network, an approach that gives good results when the problem description turns out to be complex; some encouraging results have already been achieved, but a lot of testing is still needed in order to evaluate precise recognition statistics. Preliminary results, however, show that shape analysis based on a neural network has a particularly low false positive rate. This validator, then, will become one of the voting modules described in paragraph 3.

Some shape extraction results are now presented. In fig. 5 some example of the contraction mechanism are reported: the white lines are the snakes in the initial position, that is, on the bounding box contour, while the snakes after energy minimization are drawn in yellow. Some examples are presented for a close pedestrian, fig. 5.a, and for a distant pedestrian and a scooter, fig. 5.b. Some shape extraction issues are also presented: in fig. 5.c, pedestrians shapes are correctly found, but the contour of the pedestrian in the middle also includes a warm background object; in fig. 5.d the importance of the initial snake position is highlighted: the head is not detected because it is outside of the initial snake position (in white).

Some shape extraction results are presented when the FIR images are not optimal, like those acquired in summer, under heavy direct sunlight; in this condition, many object in the background become warm, and the assumption that a pedestrian has a higher temperature than the background is not satisfied. This causes some errors in the contraction process, so that the snake in the final position does not completely adhere to the pedestrian contour, but also include some background details (fig. 5.e); in fig. 5.f, some body parts (the legs in this case) are not included in the extracted shape, because they are colder than the background. In both cases, however, the pedestrian shape is acceptably extracted.

The above discussion focused on how snakes are able to adapt to pedestrians' shape. Looking at the images, a qualitative evaluation of the shape extraction algorithm can be performed: results seem reasonable, also in bad working conditions. Since the snake-based voting system is still under development, only



Fig. 5. Results: in (a) and (b), shape extraction of a close and distant pedestrian, respectively; the white snake represents the initial position, while the yellow one is the final configuration. In (c), a group of pedestrians is analyzed; in (d), a typical issue connected with a wrong initial snake disposition is shown: the head is outside the extracted shape because it was also outside the bounding box. In (e) and (f) some results in a difficult working condition are presented, i.e. during summer, when a lot of background objects appear bright, due to the high temperature.

preliminary quantitative results can be provided: an initial release of the neural network classifier has been tested on about 1000 frames: 73 % of all pedestrians were correctly classified, as well as 91 % of non pedestrians. These preliminary results seem promising, but the net still needs some tuning, especially to improve correct classification rate.

Acknowledgment

This work has been supported by the European Research Office of the U. S. Army under contract number N62558-05-P-0380.

References

1. Bertozzi, M., Broggi, A., Fascioli, A., Sechi, M.: Shape-based Pedestrian Detection. In: *Procs. IEEE Intelligent Vehicles Symposium 2000*, Detroit, USA, October. 2000, pp. 215–220 (2000), doi:10.1109/IVS.2000.898344.
2. Bertozzi, M., Broggi, A., Felisa, M., Vezzoni, G., Del Rose, M.: Low-level Pedestrian Detection by means of Visible and Far Infra-red Tetra-vision. In: *Procs. IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan, June 2006, pp. 231–236 (2006), doi:10.1109/IVS.2006.1689633.
3. Beymer, D., Konolige, K.: Real-time Tracking of Multiple People using Continuous Detection. In: *Procs. Intl. Conf. on Computer Vision*, Kerkyra (1999)
4. Broggi, A., Bertozzi, M., Felisa, M., Grisleri, P., Ghidoni, S., Vezzoni, G., Hilario Gómez, C., Del Rose, M.: Pedestrian Detection by means of Far-infrared Stereo Vision. *Computer Vision and Image Understanding* 106(2), 194–204 (2007)
5. Curio, C., Edelbrunner, J., Kalinke, T., Tzomakas, C., von Seelen, W.: Walking Pedestrian Recognition. *IEEE Trans. on Intelligent Transportation Systems* 1(3), 155–163 (2000)
6. Cutler, R., Davis, L.S.: Robust Real-time Periodic Motion Detection, Analysis and Applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 22(8), 781–796 (2000)
7. Dao, M.-S., G., F., Natale, B.D., Massa, A.: Edge potential functions and genetic algorithms for shape-based image retrieval. In: *Procs. IEEE Intl. Conf. on Image Processing (ICIP 2003)*, Barcelona, Spain, September 2003, vol. 2, pp. 729–732 (2003)
8. Dao, M.-S., Natale, F.G.B.D., Massa, A.: Efficient Shape Matching Using Weighted Edge Potential Function. In: *Procs. 13th Intl. Conf. on Image Analysis and Processing (ICIAP 2005)*, Cagliari, Italy (September 2005)
9. Del Rose, M., Frederick, P.: Pedestrian Detection. In: *Procs. Intelligent Vehicle Systems Symposium*, Traverse City, USA (2005)
10. Gavrilu, D.M.: Pedestrian Detection from a Moving Vehicle. In: *Procs. of European Conference on Computer Vision*, vol. 2, pp. 37–49 (2000)
11. Kania, R., Del Rose, M., Frederick, P.: Autonomous Robotic Following Using Vision Based Techniques. In: *Procs. Ground Vehicle Survivability Symposium*, Monterey, USA (2005)
12. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models. *Intl. Journal of Computer Vision* 1(4), 321–331 (1988)

13. Nanda, H., Davis, L.: Probabilistic Template Based Pedestrian Detection in Infrared Videos. In: *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France (June 2002)
14. Philomin, V., Duraiswami, R., Davis, L.: Pedestrian Tracking from a Moving Vehicle. In: *Procs. IEEE Intelligent Vehicles Symposium 2000*, Detroit, USA, October 2000, pp. 350–355 (2000)
15. Polana, R., Nelson, R.C.: Detection and Recognition of Periodic, Non-rigid Motion. *International Journal of Computer Vision* 23(3), 261–282 (1997)
16. Shashua, A., Gdalyahu, Y., Hayun, G.: Pedestrian Detection for Driving Assistance Systems: Single-frame Classification and System level Performance. In: *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy (June 2004)
17. Shimizu, H., Poggio, T.: Direction Estimation of Pedestrian from Multiple Still Images. In: *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy (June 2004)
18. Stauffer, C., Grimson, W.E.L.: Similarity Templates for Detection and Recognition. *Procs. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* 1, 221–228 (2001)
19. Williams, D.J., Shah, M.: A Fast Algorithm for Active Contours and Curvature Estimation. *CVGIP: Image Understanding* 55(1), 14–26 (1992)
20. Zhao, L.: Dressed Human Modeling, Detection, and Parts Localization. Ph.D. dissertation, Carnegie Mellon University (2001)
21. Zhao, L., Thorpe, C.: Stereo and neural network-based pedestrian detection. *IEEE Trans. on Intelligent Transportation Systems* 1(3), 148–154 (2000)