IR Pedestrian Detection for Advanced Driver Assistance Systems

M. Bertozzi[1], A. Broggi[1], M. Carletti[1], A. Fascioli[1],
T. Graf[2], P. Grisleri[1], and M. Meinecke[2]

[1] Dipartimento di Ingegneria dell'Informazione
Università di Parma – Parma, I-43100, ITALY
{bertozzi,broggi,fascal,grisleri}@ce.unipr.it
[2] Electronic Research
Volkswagen AG – Wolfsburg, D-38436, GERMANY
{thorsten.graf,marc-michael.meinecke}@volkswagen.de

Abstract. This paper describes a system for pedestrian detection in infrared images implemented and tested on an experimental vehicle. A specific stabilization procedure is applied after image acquisition and before processing to cope with vehicle movements affecting the camera calibration. The localization of pedestrians is based on the search for warm symmetrical objects with specific size and aspect ratio. A set of filters is used to reduce false detections. The final validation process relies on the human shape's morphological characteristics.

1 Introduction

The capability of observing the world through visual information is a strong requirement for future driver assistance systems since their dues are getting more complex. Especially, driver assistance systems dedicated to reduce the number of fatalities and severities of traffic accidents impose several requirements on the sensorial system. One of the major and challenging tasks is the detection and classification of pedestrians.

Naturally, the use of visual cameras is a promising approach to cope with the demands of pedestrian detection. Several different image processing methods and systems have been developed in the last years, including shape-based methods [5, 6], texture- and template-based methods [7,8], stereo [9], as well as motion clues [10,11]. All these methods have to overcome the difficulties of different appearances of pedestrians in the visual domain caused mainly by e.g. clothes and illumination changes.

In order to facilitate the recognition process and to enable the detection of pedestrian in dark environments passive infrared (*IR*) cameras have come into focus [3]. Some first pedestrian detection systems [1, 2] for IR images and videos have been developed demonstrating the potential and benefits that IR cameras can provide.

In this paper we present a new pedestrian detection method employing IR cameras. This method can be divided into the following steps: (*i*) pitch angle compensation, (*ii*) localization of warm symmetrical objects with specific size and aspect ratio, (*iii*) refinement process utilizing an additional set of filters to decrease the number of false positives, and (*iv*) final validation based on human shape morphological characteristics to build the list of pedestrians appearing in the scene. Although the proposed method exploits only single images and performs no tracking, experimental results demonstrate the robustness and stability of the proposed method.

In the following section considerations on the IR domain are provided; section 3 presents the problem analysis and design choices; section 4 describes the approach and algorithm; finally section 5 discusses the results and concludes the paper with some final consideration.

2 Characterization of IR domain

Images in the IR domain convey a type of information very different from images in the visible spectrum. In the IR domain the image of an object relates to its temperature and the amount of heat it emits but is not affected by illumination changes.

Generally, the temperature of people is higher than the environmental temperature and their heat radiation is sufficiently high compared to the background. Therefore, in IR images pedestrians are bright and sufficiently contrasted with respect to the background, thus making IR imagery suited to their localization. Other objects which actively radiate heat (cars, trucks...) have a similar behavior; however people can be recognized thanks to their shape and aspect ratio.

One major point in favor of IR cameras is the independency to lighting changes: IR cameras can be used in day-time or night-time with no or little difference extending vision beyond the usual limitations of day-light cameras. Moreover, the absence of colors or strong textures eases the processing towards interpretation. Furthermore, the problem of shadows is greatly reduced.

Nevertheless, the problem of detecting humans in IR images is far from being trivial.

Conditions of high temperature or strong sun heating can decrease the temperature's differential in the scene modifying the thermal footprint of bodies. In addition, in case of strong external heat radiation, clothes that people wear can have different thermal behavior depending on their type and color, thus introducing textures to the image.

Conversely, in case of low external temperature, clothes can significantly shield the heat emission and only parts of the body can be perceivable. Another problem, even if less critical than in the visible domain, is represented by objects carried by people.

The problems mentioned above make the detection of pedestrians harder. Nevertheless, the IR domain seems to be promising and justifies deep investigation.

3 Problem analysis and design choices

Two issues have to be defined when designing the system: the *setup of the vision system*, considering physical and aesthetical automotive requirements and the *target*, i. e. the desired range of pedestrians' height and width. Moreover, the algorithm has to be designed considering that the input data are low resolution digital images.

Setup of the vision system: the mapping between image pixels and world coordinates has to be known for a correct localization. The calibration is performed on a flat stretch of road by placing markers at known distances up to 40 m; the relation between 3D coordinates of these points and the corresponding pixels in the image was used to compute camera orientation.

These parameters are then used for computing the relationships between 3D coordinates and image pixels, under the assumption of a flat road in front of the vision system. While being a strong assumption, in the area close to the vehicle (up to 20 m) it is supposed to hold even in presence of hills or bumps. In the faraway area, some errors in the calibration may occur, thus generating less confident results.

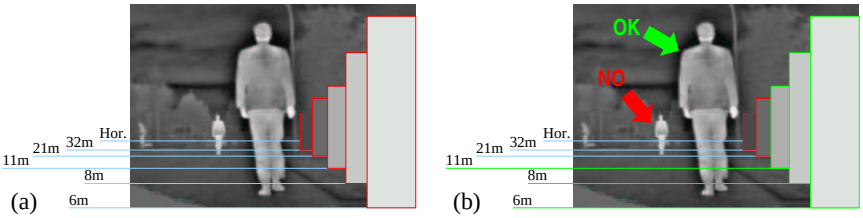


Fig. 1. Detection range: (*a*) sizes of BBs for 170 cm tall pedestrians at different distances; (*b*) the feasible distance range for a 170 cm tall pedestrian (green).

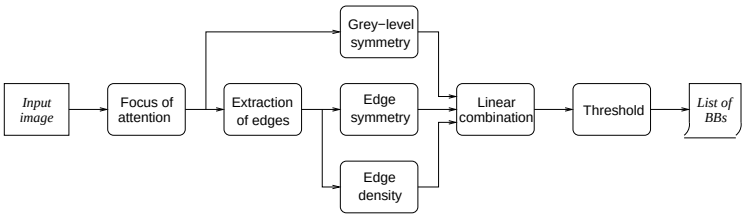


Fig. 2. Scheme of the low-level phase.

Detection distance range: the presence of a pedestrian is checked in different-sized bounding boxes (*BB*s) placed in different positions in the image.

Indeed, not all possible BBs need to be checked, mainly due to computational time and detail content. In fact, tiny BBs enclosing faraway pedestrians feature a very low information content and may be easily confused with other road participants or infrastructures that present similar thermal characteristics.

Thus, it is imperative to define a BBs' size-range for which the detection can lead to a sufficiently accurate result. In this work the considered sizes are: 12×28 pixels for the smallest BB and 42×100 for the largest one. In addition, a specific size and aspect ratio are used to define targets. The size of a pedestrian has been fixed to the following values: height 180 cm $\pm$10% and width 60 cm $\pm$10%. These two choices lead to a limited detection area in front of the vehicle, as described in the following.

Fig. 1.a shows distances from 6 m to 32 m assuming a flat road. The image displays two pedestrians at different distances from the acquisition sensor. Also sizes of BBs for a 170 cm tall pedestrian at different distances are shown. Fig. 1.b shows the size of the BBs that are examined (green), and highlights which pedestrian cannot be detected; the distance range in which a 170 cm tall pedestrian can be detected is shown as well (red). The detection distance range can be computed as a function of pedestrian height specification and the BBs height range [12]. With the setup and design choices used for this work, the distance range is 7 m $\div$ 20 m.

4 Algorithm description

The algorithm is divided into three parts: (*i*) image stabilization for pitch angle compensation, (*ii*) localization of areas of interest (attentive vision) and generation of possible candidates based on symmetry, (*iii*) filtering of candidates to remove errors, based on non-pedestrian characteristics, (*iv*) validation of candidates on the basis of a match with

(a) (b) (c)

Fig. 3. (*a*) Source image, (*b*) edges in the interesting area, (*c*) correlation between two subsequent histograms.

a simple morphological model. The last part is currently under development and its results need to be validated by further investigation.

4.1 Image Stabilization

Vehicle oscillations are generally present during driving and are due to road roughness; although filtered by the dumper system, they are perceived by the camera. Since the pedestrian recognition application relies on correct camera calibration, vertical oscillations need to be compensated. A software image stabilizer is preferable with respect to optical/mechanical ones because it requires less hardware resources.

Stabilization is regarded as a prefiltering and applied after image acquisition and before the actual image processing.

From the analysis of different video sequences, four types of movements in the images have been identified.

Perspective movements: when the vehicle is moving on a straight flat road, due to perspective, each frame resembles a zoomed version of the previous.

Horizontal translation: when the vehicle is steering, objects move horizontally.

Vertical translation: when the vehicle is driving uphill or downhill, objects move vertically. This is also the case when approaching high obstacles such as overbridges.

Vertical oscillations: road imperfections or artifacts (rough road or speed bumpers) are the main cause of vertical oscillations. A procedure devoted to compensate the last type of movements is presented.

The image stabilizer evaluates the movements of objects' horizontal borders, which generally –in absence of oscillations– move continuously and slowly towards the top of the image. Two consecutive frames are compared and the vertical offset between them is computed. The oscillation is compensated by shifting the new image in the opposite direction. In the following the main steps are presented.

Focus of Attention: in order to ignore the vertical movements of edges present in the image when approaching tall structures (e.g. overbridges) only the central stripe of the image is processed.

Edge Extraction: horizontal edges are extracted in the area of interest by means of a classical Sobel filter. The grey-level resulting image is then thresholded using an adaptive threshold.

Fig. 3.a presents a frame of a test sequence while fig. 3.b shows the result of horizontal edge detection in the interesting area.

Histogram Computation: a row-wise histogram is computed on the edge image. The histogram is checked whether appropriate information is available or not. In case of a

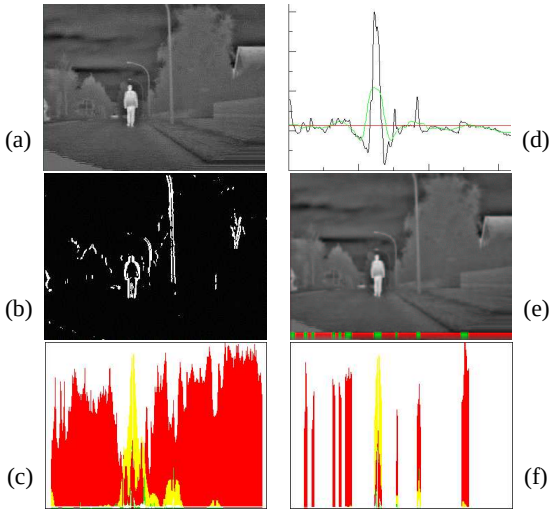


Fig. 4. Computation of symmetries and focus of attention: (*a*) original image; (*b*) vertical edges image; (*c*) symmetries of grey levels (red) and vertical edges (green), density of vertical edges (yellow), their combination (white); (*d*) histogram of grey levels together with its global (red) and local (green) average; (*e*) positions of possible vertical symmetry axes (green); (*f*) histograms are computed only in correspondence with the green dashes.

low quantity of information (e.g a uniform scene) the procedure is stopped and stabilization disabled.

Correlation Evaluation: the changes in the shape of the histogram from one frame to the following allow to determine the offset between them. This is achieved through the convolution between the current histogram and a reference one, which is dynamically updated. Fig. 3.c shows two histograms in a sequence; the red arrows point out the correlation between the highest peaks. The computation of the offset is straightforward.

A set of filters are used to remove abrupt variations and low frequency components of oscillations which generally result from perspective effects of road objects. Another filter is used to progressively reduce the offset in a sequence, bringing the horizon back to its original position.

Image Shift: the vertical offset computed so far is then used to shift the current frame before processing.

4.2 Candidates generation and filtering

The low-level part of the algorithm, depicted in fig. 2, is mainly based on the computation of symmetries. The input image is processed to focus the attention on interesting regions, then vertical edges are extracted. Both input and vertical edges images are searched for symmetrical areas, with specific aspect-ratio and size/perspective constraints. The density of edges in these areas is also considered.

The original input image is shown in fig. 4.a; fig. 4.b depicts a binary image containing the vertical edges, and fig. 4.c shows a number of histograms computed by selecting, for each vertical symmetry axis, the BB having the maximum: symmetry of grey levels

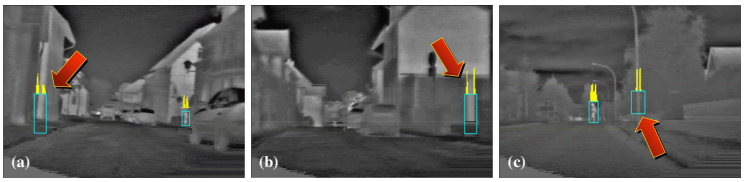


Fig. 5. Filters relying on specific features of artifacts. For each BB the vertical histogram of edges is displayed on top of it. The arrows point discarded BBs.

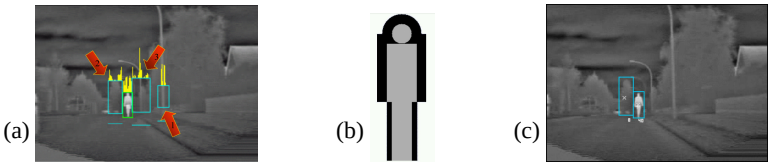


Fig. 6. (*a*) Elimination of BBs following the resize step; (*b*) the simple model encoding morphological characteristics of a pedestrian, (*c*) the match with the model allows to validate a pedestrian and to discard a false positive.

(red), symmetry of vertical edges (green), and density of vertical edges (yellow). The white histogram presents a combination of all the above; the pedestrian presents high local peaks in all histograms.

Considerations generally true for images in the IR domain permit to reduce the number of symmetry axes to be examined, thus avoiding the need of an exhaustive search. Since pedestrians are warmer than the background, a filter has been defined to eliminate symmetry axes in cold image areas. For this purpose, a histogram encoding the presence of white (hot) pixels is computed; its local average (computed on a small window) as well as its overall average are also computed (fig. 4.c). The low-pass filter is used to smooth the histogram and remove small peaks close to high peaks, while the overall average is used to mask out histogram peaks in cold areas.

Thus, vertical symmetry axes intersecting the green portions of the bottom of the image are considered, while the remaining ones (intersecting red dashes) are neglected (figs. 4.d and 4.e).

Candidates are generated by thresholding the resulting histogram. Each over-threshold peak corresponds to a BB containing the shape of a potential pedestrian. This list is then passed on to next phase of the processing which is in charge of selecting and possibly removing false positives.

The candidates are filtered on the basis of specific features of human artifacts that may have been highlighted as potential pedestrians in the previous step. In other words, specific filters have been designed to remove: BBs centered on poles, road signs, buildings, and other artifacts that present high symmetry and BBs that feature a reduced amount of edges in the upper and lower part, since pedestrians are characterized by a uniform distribution of edges.

The edges within a BB are used to compute a vertical histogram. This operation is performed for each BB. The shape of the vertical histogram is used to filter the BBs (fig. 5).

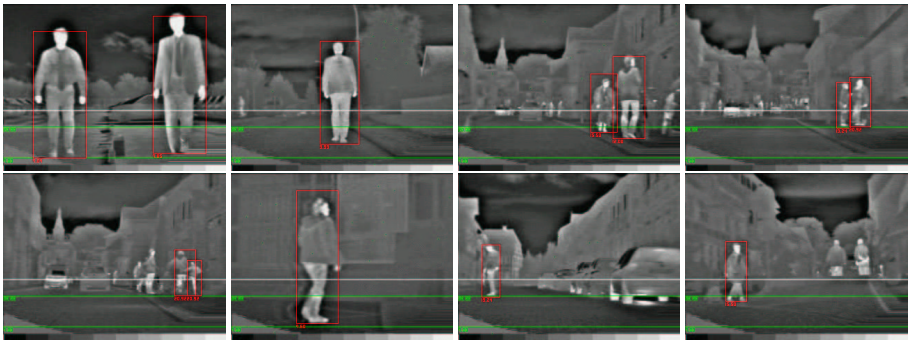
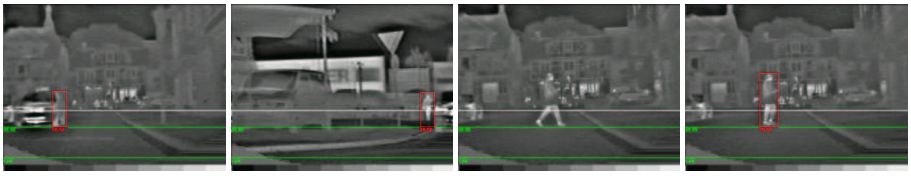


Fig. 7. Results of pedestrian detection in different situations: with complex or simple scenario or with one or more pedestrians. The distance (meters) is displayed below the boxes. The two horizontal green lines encode the range in which pedestrians are searched for, while the horizontal white line encodes the horizon.



(a) (b) (c) (d)

Fig. 8. Situation in which the algorithm fails: (*a*) a false positive is found due to a complex background, (*b*) a wrong distance is computed, (*c*) a walking pedestrian is misdetected when it does not met the aspect-ratio constraints, while it is correctly detected in the following frame (*d*).

Each surviving BB is then resized and reduced in height and width in order to fit the internal presence of edges. The BBs that have been resized too much are then removed, while the other BBs are again filtered in order to eliminate: BBs that –due to this resize operation– would represent pedestrians that are too faraway (fig. 6.a, arrow 1), BBs that do no longer meet perspective constraints (fig. 6.a, arrow 2), and BBs that do no longer meet the original assumptions on aspect ratio (fig. 6.a, arrow 3).

4.3 Validation of candidates

The following paragraph describes the preliminary results of the ongoing research.

Each surviving BB is validated through the match with a simple model encoding morphological and thermal characteristics of a pedestrian.

Fig. 6.b shows the model that is used for the match; the model is resized according to the size of the BB, and then matched with the gray level original image. A vote is given and the candidates which present a vote lower than a threshold are discarded.

In fig. 6.c the votes relative to two BBs are displayed: the highest represents an actual pedestrian, while the lowest one represents a tree. Even if the tree has a shape that resembles a human, the filter is able to discard it.

5 Discussion of results and conclusions

Fig. 7 shows a few results of pedestrian detection in IR images. The two horizontal green lines encode the detection distance range in which pedestrians are searched for ($7\ m \div 20\ m$), while the horizontal white line encodes the horizon when the flat road and no vehicle pitch assumptions are met. Results are encoded as BBs superimposed onto the original images. They highlight pedestrians within the distance range only.

The result shows that the system is able to detect one or more pedestrians even in presence of a complex background. The major critical situations, presented in fig. 8, are: (*i*) in presence of a complex background, artifacts or objects other than pedestrians are occasionally detected (fig. 8.a); (*i*) the algorithm does not miss the detection of a pedestrian but miscalculates the exact position or size of the BB, thus corrupting the distance estimation (fig. 8.b); (*iii*) walking pedestrians are occasionally not detected due to aspect ratio constraints (see fig. 8.c); anyway in the following frame of the sequence (fig. 8.d) the same pedestrian is correctly detected, thus a tracking could be used to solve this case.

The algorithm developed so far proves to be effective in different situations although it employs only single images. In order to increase the reliability and robustness of the system appropriate tracking procedures will be added and validated by carrying out extensive tests in different weather conditions.

References

1. H. Nanda and L. Davis, "Probabilistic Template Based Pedestrian Detection in Infrared Videos," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, June 2002.
2. F. Xu and K. Fujimura, "Pedestrian Detection and Tracking with Night Vision," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, June 2002.
3. Y. L. Guilloux and J. Lonnoy, "PAROTO Project: The Benefit of Infrared Imagery for Obstacle Avoidance," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, June 2002.
4. T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of Night-vision System ," *IEEE Trans. on Intelligent Transportation Systems*, vol. 3, pp. 203–209, Sept. 2002.
5. M. Bertozzi, A. Broggi, A. Fascioli, and M. Sechi, "Shape-based Pedestrian Detection," in *Procs. IEEE Intelligent Vehicles Symposium 2000*, pp. 215–220, Oct. 2000.
6. D. M. Gavrila and J. Geibel, "Shape-Based Pedestrian Detection and Tracking," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, June 2002.
7. C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking Pedestrian Recognition," *IEEE Trans. on ITS*, vol. 1, pp. 155–163, Sept. 2000.
8. M. Oren, C. Papageorgiu, P. Sihna, E. Osuna, and T. Poggio, "Pedestrian Detection using Wavelet Templates," in *Procs. IEEE Conf. on Computer Vision and Pattern Recognition*, (San Juan, Puerto Rico), pp. 193–199, 1997.
9. L. Zhao and C. Thorpe, "Stereo- and Neural Network-based Pedestrian Detection," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems'99*, pp. 298–303, Oct. 1999.
10. B. Heisele and C. Wöhler, "Motion-based Recognition of Pedestrians," in *Procs. IEEE Intl. Conf. on Pattern Recognition*, pp. 1325–1330, June 1998.
11. R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. on PAMI*, vol. 22, pp. 781–796, Aug. 2000.
12. M. Bertozzi, A. Broggi, T. Graf, P. Grisleri, and M. Meinecke, "Pedestrian Detection in Infrared Images," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, June 2003. in press.