

Pedestrian Detection in Infrared Images

M. Bertozzi, A. Broggi, P. Grisleri

Dipartimento di Ingegneria dell'Informazione
Università di Parma
Parma, I-43100, Italy
{bertozzi,broggi,grisleri}@ce.unipr.it

T. Graf, M. Meinecke

Electronic Research
Volkswagen AG
Wolfsburg, D-38436, Germany
{thorsten.graf,marc-michael.meinecke}@volkswagen.de

Abstract— This paper describes an approach for pedestrian detection in infrared images. The developed system has been implemented on an experimental vehicle equipped with an infrared camera and preliminarily tested in different situations.

It is based on the localization of warm symmetrical objects with specific size and aspect ratio; since also road infrastructures and other road participants may have such characteristics, a set of matched filters was added in order to reduce false detections. A final validation process, based on the human shape's morphological characteristics, is used to build the list of pedestrian appearing in the scene.

No temporal correlation, nor motion cues are used in this first part of the project.

Keywords— Pedestrian Detection, Machine Vision, IR Imagery.

I. INTRODUCTION

The development of advanced driver assistance systems is an important and active research area. Especially the development of systems that are capable of reducing the number or the severity of traffic accidents involving pedestrians is of major interest. Pedestrian accidents represent the second largest source of traffic-related injuries (right after accidents involving car passengers). For example, yearly more than 200,000 pedestrians are injured and about 9,000 are killed in traffic accidents in the EU.

The use of video sensors and image processing methods dedicated to detect and classify pedestrians provides a promising approach for the development of such driver assistance systems

In the last years research has mainly focused on visual cameras. Many different methods and systems have been developed including e.g. shape-based methods [5, 6], texture-based methods [7], stereo [8], and motion [9]. All these methods have to cope with the difficulties of different appearances of pedestrians in the visual domain caused by e.g. clothes and additional equipment.

Recently, the benefits and advantages of using passive infrared cameras have been considered (i.e. [3]). Some first pedestrian detection systems [1, 2] for infrared images and videos have been developed showing that infrared images can facilitate the recognition process.

In this paper we present a new pedestrian detection method employing infrared images. It is based (i) on the localization of warm symmetrical objects with specific size and aspect ratio, (ii) a refinement process utilizing an additional set of filters to decrease the number of false positives, and (iii) a final validation procedure based on human shape's 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.

II. CHARACTERIZATION OF IR DOMAIN

Images in the infrared domain convey a type of information very different from images in the visible spectrum. Basically, in the visible spectrum the image of an object depends on the amount of incident light on its surface and on how well the surface reflects it. On the other hand, in the infrared domain the image of an object relates to its temperature and the amount of heat it emits.

Generally, the temperature of people is higher than the environment temperature and their heat radiation is sufficiently high compared to the background. Therefore, in infrared images pedestrians belong to the upper range in the grey-level scale and are sufficiently contrasted with respect to the surroundings, thus making infrared imagery particularly suited to their localization. Obviously, other objects which actively radiate heat, such as automobiles, trucks, busses, and motorcycles, have a similar behavior; however people can be recognized thanks to their shape and aspect ratio.

One major point in favor of infrared cameras is their independence of light conditions: they 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 textures eases the processing towards interpretation. Furthermore, the problem of shadows is greatly reduced. In fact, even if persistent shadows are still present in IR images – due to the different temperatures caused by shadows themselves – incidental shadows, which do not modify the temperature of bodies, are not perceivable.

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

Weather conditions, such as heavy fog or rain, can modify the thermal footprint of bodies, limiting the effectiveness of infrared systems.

Moreover, conditions of high temperature and strong sun heating can decrease the difference of temperature between pedestrians and other objects. In fact, objects which have a passive heat radiation behavior, such as traffic signs, barriers, trees, buildings, and road markings, may be strongly heated by the

This work was funded by Volkswagen AG.

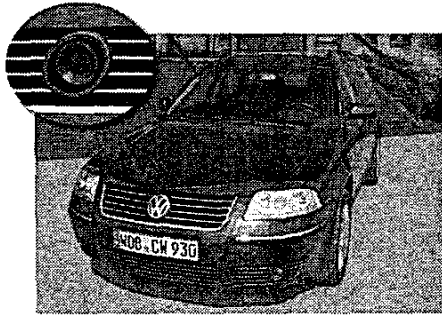


Fig. 1 The position of the infrared camera on the VW test vehicle.

sun, making the scene more complex or even causing heat radiations or reflections. 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 adding texture to the image.

Conversely, in case of low external temperature, clothes can significantly shield the heat emission and only parts of the body (such as head or hands) 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.

III. 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;
- the target, i. e. the desired range of pedestrians' height and width.

The algorithm has to be designed considering that the input data are low resolution digital images.

A. Setup of the vision system

The camera position is fixed by physical constraints and aesthetical choices (see figure 1).

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 meters (see figure 2); the relation between 3D coordinates of these points and the corresponding pixels in the image was used to compute camera orientation.

The computed parameters are then used for all future relationships between 3D world coordinates and image pixels, under the assumption of a flat road in front of the vision system and negligible vehicle pitch. Indeed, these strict assumptions may be violated, but in the area close to the vehicle (up to 20 meters) they are supposed to hold even in presence of hills or bumps. In the faraway area (more than 20 meters) some errors in the calibration may occur, thus generating less confident results.

B. Definition of target

A specific size and aspect ratio was used to define targets. The size of a pedestrian was fixed to the following values: *i* height: 180 cm +/- 10% and *ii* width: 60 cm +/- 10%.

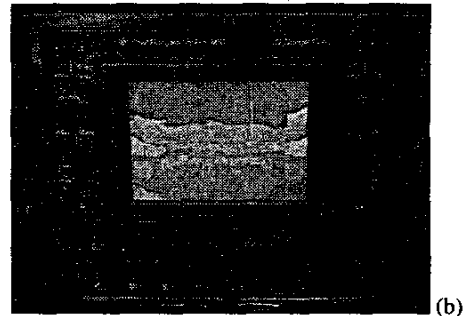
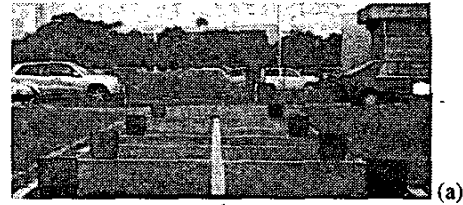


Fig. 2 (a) The calibration setup (b) internal view.



Fig. 3 A small bounding box enclosing a faraway pedestrian.

C. Detection distance range

The presence of a pedestrian is checked in different-sized bounding boxes placed in different positions in the image. Indeed, not all possible bounding boxes need to be checked, mainly due to computational time and detail content.

In fact, very small bounding boxes enclosing faraway pedestrians feature a very low information content. In these situations it is easy to obtain false positives since many road participants (other than pedestrians), other objects, and even road infrastructures may present morphological characteristics similar to a human shape. An example of the low information content in small bounding boxes is shown in figure 3.

The largest bounding box is limited by the vertical resolution of the image.

It is therefore imperative to define a range of reasonably-sized bounding boxes in which the detection may lead to a sufficiently accurate result. In this work the considered size is:

- smallest bounding box: 12 × 28 pixels,
- largest bounding box: 42 × 100 pixels.

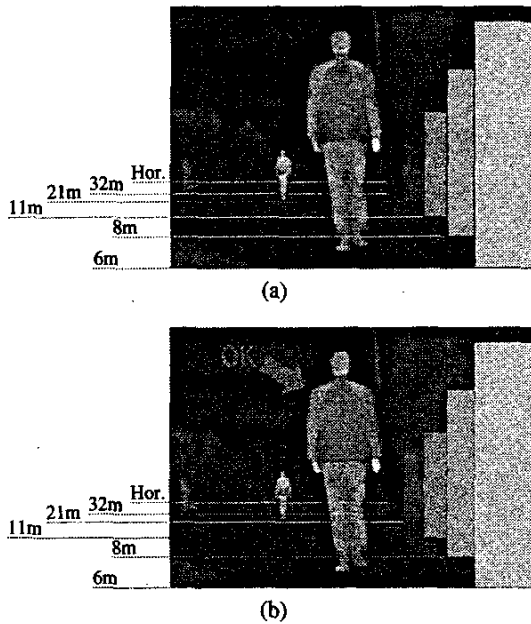


Fig. 4 (a) Pedestrians of different height standing at different distances and a bounding box containing a 170 cm tall pedestrian at different distances; (b) in green the feasible distance range for a 170 cm tall pedestrian.

Indeed, this choice leads to a limited detection area in front of the vehicle, as described in the following.

The calibration is used to fix the correspondence between lines of the image and distances in the 3D world, assuming a flat road. Distances from 6 m to 32 m are considered in figure 4.a as an example. The image displays two pedestrians at different distances from the acquisition sensor. For reference purpose, the image also shows the size of a bounding box containing a 170 cm tall pedestrian at different distances.

Given the above specifications on bounding boxes size, figure 4.b shows in green the size of the bounding boxes that are examined, and highlights which of the two pedestrians cannot be detected due to his/her size in the image. The distance range in which the detection of a 170 cm tall pedestrian can take place is also shown.

The graph in figure 5 shows the working area of the system. The minimum distance at which pedestrians can be completely seen, given by the setup, is represented by the vertical dashed line. On the other hand, the specification about pedestrian height determine the limits represented by the two horizontal dashed lines. Therefore, the search area extends to the right of the vertical dashed line and between the two horizontal dashed lines.

Moreover, some extra considerations, deriving from the definition of bounding box size, need to be made in order to localize the region of the graph which represents the actual working area of the system. The additional curves on the diagram represent the iso-bounding box mappings: each curve describes the relationship between the distance and height of objects enclosed by bounding boxes with the same height in pixels. Given the range of bounding boxes height (BBh_{max} , BBh_{min}), the working range of the systems is depicted as the intersection of the search

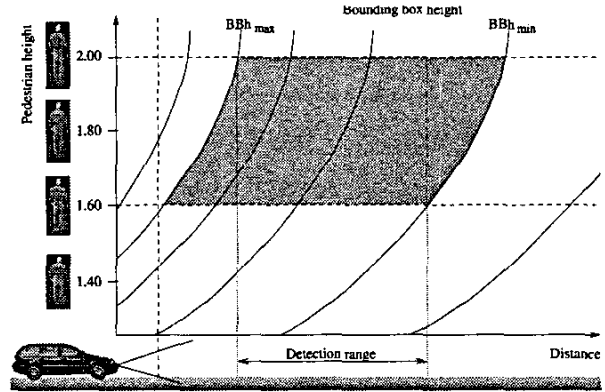


Fig. 5 The detection distance range.

area described above with the area which extends between the two iso-bounding box mappings corresponding to BBh_{max} pixels and BBh_{min} pixels, shaded in figure 5.

In order to be sure that for a given distance all pedestrians (from the shortest to the tallest) can be detected, the working area has to be further limited to the portion of the shaded area delimited by the two vertical dotted lines. The arrow highlights the actual distance range. With the current setup and design choices the distance range is 7 m ÷ 20 m.

Considerations may be made on the behavior of the distance range with the increment or decrement of the target height; in other words, extending the target height range will shorten the distance range in which all pedestrians are detected.

IV. ALGORITHM DESCRIPTION

The algorithm is divided into three parts:

- localization of areas of interest (attentive vision) and generation of possible candidates based on symmetry,
- filtering of candidates to remove errors, based on non-pedestrian characteristics,
- validation of candidates on the basis of a match with a simple morphological model of a pedestrian.

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

A. Candidates generation

The low-level part of the algorithm, depicted in figure 6, 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 the input image and the image containing vertical edges are searched for symmetrical areas, with specific aspect-ratio and size constraints matching a normal pedestrian shape, taking also into account perspective issues. The density of edges in these areas is also considered.

Figure 7 shows the original input image (figure 7.a), a binary image containing its vertical edges (figure 7.b), and a number of histograms computed by selecting, for each vertical symmetry axis, the bounding box having the maximum (figure 7.c):

- symmetry of grey levels (red),
- symmetry of vertical edges (green),
- density of vertical edges (yellow).

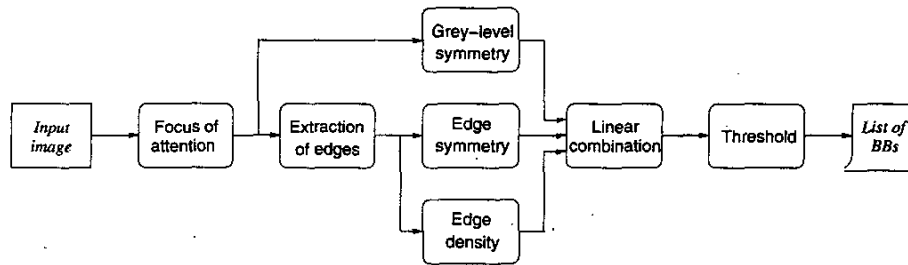


Fig. 6 Scheme of the low-level phase.

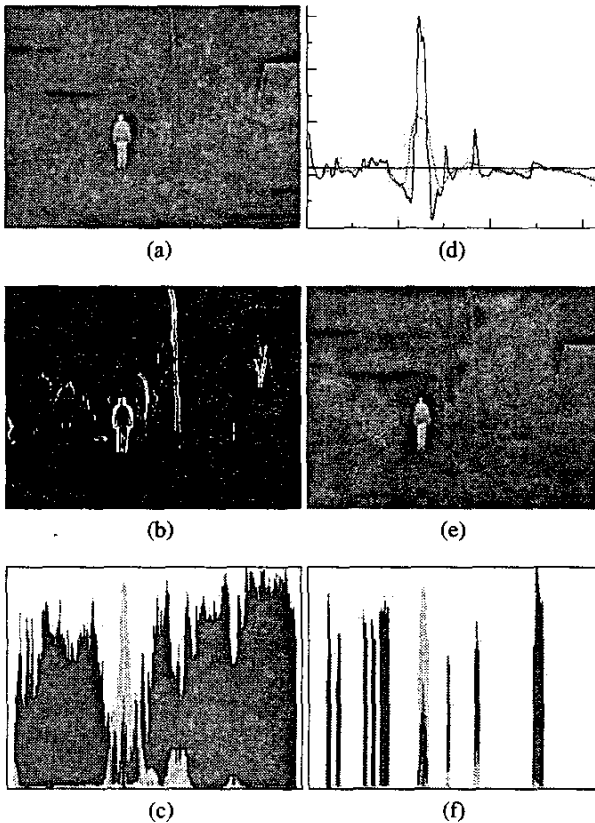


Fig. 7 Computation of symmetries and focus of attention: (a) original image; (b) vertical edges image; (c) symmetry of grey levels (red), symmetry of vertical edges (green), density of vertical edges (yellow), combination (white); (d) histogram of grey levels together with its global average (in red) and local average (in green); (e) positions of possible vertical symmetry axes (in green); (f) histograms are computed only in correspondence with the green dashes.

The white histogram presents a combination of all the above; the pedestrian presents high local peaks in all histograms.

Instead of performing an exhaustive search, which would definitely take a long time and consume a great amount of computational resources, specific areas of interest are determined. Considerations generally true for images in the infrared domain permit to reduce the number of symmetry axes to be examined. In other words, 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. 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.

Figure 7.d shows the histogram, its average (in red) and its low-pass filtered version (in green); assuming a pedestrian is hotter than its background, the symmetries are computed only in the areas in which the histogram presents values larger than the overall average and the local average. More specifically, as shown in figure 7.e, vertical symmetry axes intersecting the green portions of the bottom of the image are considered, while the remaining ones (intersecting red dashes) are neglected. Figure 7.f shows the actual histograms computed in correspondence with the green dashes only. This improves both the detection (false positives are reduced in number) and the computational time.

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

B. Candidates filtering

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:

- bounding boxes centered on poles, road signs, buildings, and other artifacts as well as road infrastructures that present high symmetry,

- bounding boxes that feature a reduced amount of edges in the upper and lower part, since pedestrians are characterized by a uniform distribution of edges in the bounding box.

The edges within a bounding box are used to compute a vertical histogram. This operation is performed for each bounding box. The shape of the vertical histogram is used to filter the bounding boxes. Three criteria have been developed.

Filter no.1: The bounding box is removed when the center of the histogram is empty. This is true for large poles, pylons, columns, even if they are not perfectly vertical (see figure 8.a).

Filter no.2: The bounding box is removed when more than half of the histogram is empty. This is true for large vertical poles, pylons, columns (see figure 8.b).

Filter no.3: The bounding box is removed when the histogram is confined to the central part of the bounding box, namely when the left and/or right parts are empty, or when the

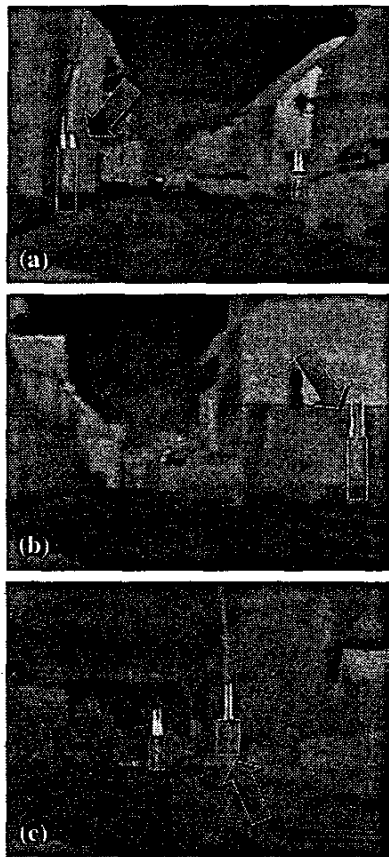


Fig. 8 Filters relying on specific features of artifacts. For each bounding box the vertical histogram of edges is displayed in yellow on top of it. The bounding boxes pointed by the arrows are discarded.

histogram is concentrated in two small areas that contain more than 80% of the contributions. This is true for thin vertical poles, pylons, columns (see figure 8.c).

Each surviving bounding box is then resized and reduced in height and width in order to fit the internal presence of edges. The bounding boxes that have been resized too much are then removed, while the other bounding boxes are again filtered in order to eliminate:

- bounding boxes that –due to this resize operation– would represent pedestrians that are too faraway (see arrow 1 in figure 9),
- bounding boxes that do no longer meet perspective constraints (see arrow 2 in figure 9),
- bounding boxes that do no longer meet the original assumptions on aspect ratio (see arrow 3 in figure 9).

C. Validation of candidates

The following paragraph describes the current status of the ongoing research. The results depicted here are preliminary.

Each surviving bounding box is validated through the match with a simple model encoding morphological characteristics of a pedestrian. This filter, based on shape, is used to remove candidates that do not present a human shape, or are not as hot as a pedestrian is expected to be.

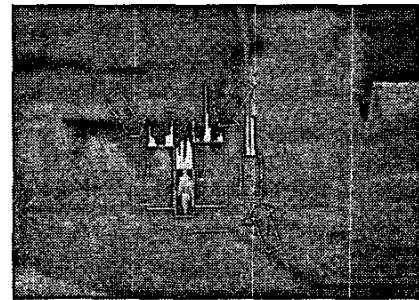


Fig. 9 Elimination of bounding boxes following the resize step.

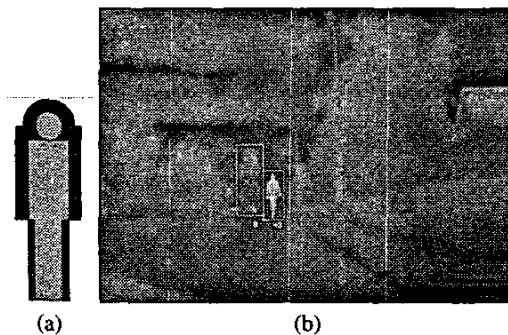


Fig. 10 (a) The simple model encoding morphological characteristics of a pedestrian; (b) the match with the model allows to validate a pedestrian and to discard a false positive.

Figure 10.a shows the model that is used for the match; the model is resized according to the size of the bounding box, 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 figure 10.b the votes relative to two bounding boxes 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.

V. DISCUSSION OF RESULTS AND CONCLUSIONS

Figure 11 shows a few results of pedestrian detection in infrared 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 bounding boxes 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. 12, are:

- in presence of a complex background, artifacts or objects other than pedestrians are occasionally detected (see fig. 12.a, 12.b, and 12.c);
- the algorithm does not miss the detection of a pedestrian but miscalculates the exact position or size of the bounding box, thus corrupting the distance estimation (see fig. 12.d);
- walking pedestrians are occasionally not detected due to as-

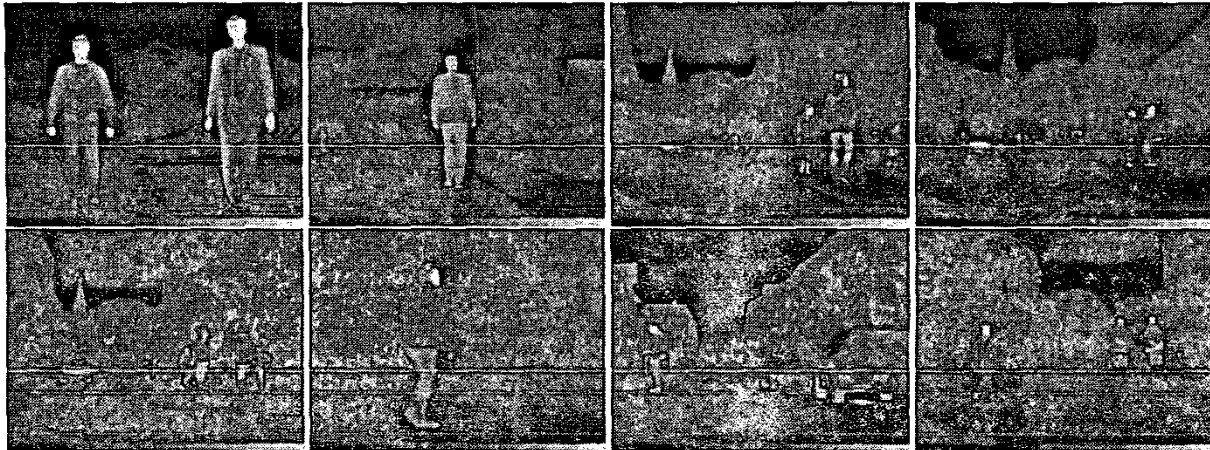


Fig. 11 Results of pedestrian detection in different situations: with complex or simple scenario or with one or more pedestrians. The distance (meters) is displayed below the boxes. The two horizontal green lines encode the range in which pedestrians are searched for, while the horizontal white line encodes the horizon.

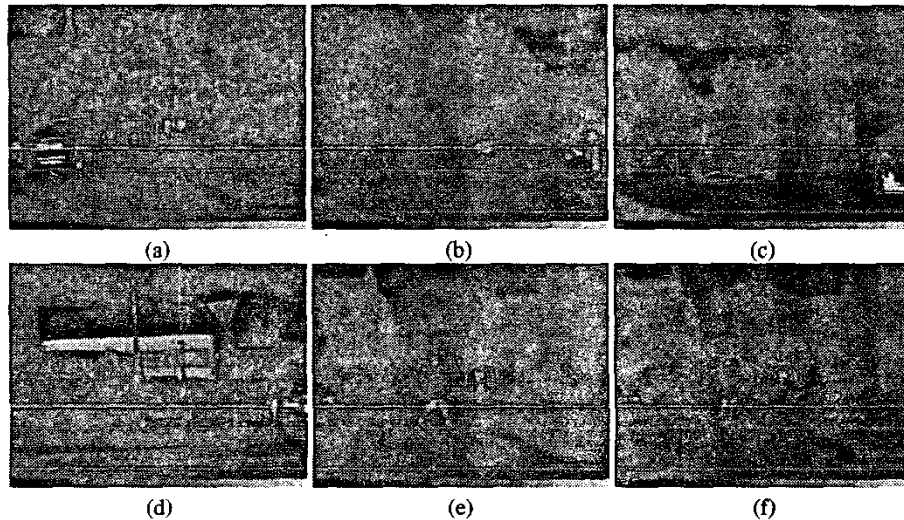


Fig. 12 Situation in which the algorithm fails: (a), (b), and (c) the algorithm finds false positives due to a complex background, (d) the algorithm computes a wrong distance, (e) a walking pedestrian is misdeteched when not meeting the aspect-ratio constraints, while he/she is correctly detected in the following frame (f).

pect ratio constraints (see fig. 12.e); in the following frame of the sequence (figure 12.f) the same pedestrian is still correctly detected, thus tracking could be used to solve this particular case.

The algorithm developed so far proves to be effective in different situations. Extensive tests are being carried out in different weather conditions. Future research steps include the strengthening of the last part of the algorithm with the use of a 3D model.

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*, (Detroit, USA), 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 Intelligent Transportation Systems*, vol. 1, pp. 155–163, Sept. 2000.
- [8] 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.
- [9] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 781–796, Aug. 2000.