Pedestrian Detection for Driver Assistance Using Multiresolution Infrared Vision

Massimo Bertozzi, *Associate Member, IEEE*, Alberto Broggi, *Associate Member, IEEE*, Alessandra Fascioli, *Member, IEEE*, Thorsten Graf, and Marc-Michael Meinecke

***Abstract*—This paper describes a system for pedestrian detection in infrared images, which has been implemented on an experimental vehicle equipped with an infrared camera. The proposed system has been tested in many situations and has proven to be efficient and with a very low false-positive rate. It is based on a multiresolution localization of warm symmetrical objects with specific size and aspect ratio; anyway, because road infrastructures and other road participants may also have such characteristics, a set of matched filters is included in order to reduce false detections. A final validation process, based on human shape's morphological characteristics, is used to build the list of pedestrian appearing in the scene. Neither temporal correlation nor motion cues are used in this first part of the project: the processing is based on the analysis of single frames only.**

***Index Terms*—Infrared imagery, machine vision, multiresolution, pedestrian detection.**

I. Introduction

THE development of in-vehicle assistance systems dedicated to reducing the number of fatalities and the severity of traffic accidents is an important and active research field. Since pedestrian accidents represent the second largest source of traffic-related injuries (annually, more than 200 000 pedestrians are injured and approximately 9000 are killed in traffic accidents in the European Union), systems that are capable of reducing the number or effects of traffic accidents involving pedestrians are of major interest.

The tasks of such driver-assistance systems are extremely complex; the use of vision sensors and image-processing methods provides a promising approach.

Several different image-processing methods and systems dedicated to detecting and classifying pedestrians have been developed in the last years, including shape-based [5], [6], texture-based [7], stereo [8], and motion [9] methods. An approach that combines motion and appearance information is presented in [10]. All of these approaches have to overcome the difficulties of different appearances of pedestrians in the visual domain, mainly caused by clothes, carryons, illumination changes, and—indeed—different postures.

Manuscript received July 21, 2003; revised March 5, 2004 and May 13, 2004. This work was supported by Volkswagen AG.

M. Bertozzi, A. Broggi, and A. Fascioli are with the Dipartimento di Ingegneria dell'Informazione, Università di Parma, Parma I-43100, Italy (e-mail: bertozzi@ce.unipr.it; broggi@ce.unipr.it; fascioli@ce.unipr.it).

T. Graf and M.-M. Meinecke are with Electronic Research, Volkswagen AG, Wolfsburg D-38436, Germany (e-mail: thorsten.graf@volkswagen.de; marc-michael.meinecke@volkswagen.de).

Digital Object Identifier 10.1109/TVT.2004.834878

Only recently, thanks to the decreasing cost of infrared devices, the benefits and advantages of using infrared cameras have been actually considered (e.g., [3] and [11]). Some first pedestrian-detection systems have been developed, showing that infrared images can facilitate the recognition process [1], [12].

In this paper, we present a new pedestrian-detection method employing far infrared images. It is based on the following:

1) localization of warm symmetrical objects with specific aspect ratio and size;
2) filtering process to avoid a number of false positives;
3) final validation procedure based on human shape and thermal characteristics.

The result is a list of pedestrians appearing in the scene, each detected by position, angle of view, height, and an approximate posture.

The process is iterated at different image resolutions in order to detect both close and faraway pedestrians. The following assumptions have been made:

1) pedestrians are not occluded;
2) complete shape of the pedestrian appears in the image;
3) a number of pedestrians appear simultaneously in the image but they do not occlude each other.

Although the proposed method does not perform tracking, experimental results demonstrate its robustness and effectiveness.

In Section II, considerations on the infrared domain are provided. Section III shows how design choices affect the detection range and Section IV describes the approach and algorithm. Finally, Section V discusses the results and concludes the paper with some final considerations.

The images produced by our software have been modified by hand in order to adapt to grayscale printing.

II. Characterization of the Infrared (IR) Domain

Images in the IR domain convey a type of information that is 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 IR 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 IR, images pedestrians belong to the upper range in the gray-level scale and are sufficiently contrasted with respect to the surroundings, thus

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

making IR imagery particularly suited to pedestrians localization. Obviously, other objects that 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 IR cameras is their independence of light conditions: they can be used in the daytime or nighttime with little or no difference, extending vision beyond the usual limitations of daylight cameras. Moreover, the absence of colors or textures eases the processing toward 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 IR 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 IR systems.

Moreover, conditions of high temperature and strong sun heating can decrease the difference of temperature between pedestrians and other objects. In fact, objects that have a passive heat radiation behavior, such as traffic signs, barriers, trees, buildings, and road markings, may be strongly heated by the sun, making the scene more complex or even causing heat radiations or reflections. In addition, in the 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 the 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 more difficult. Nevertheless, the IR domain seems to be promising and justifies deep investigation.

III. Design Choices and Detection Range

Two issues have to be defined when designing the system:

- setup of the vision system, considering physical and aesthetical automotive requirements;

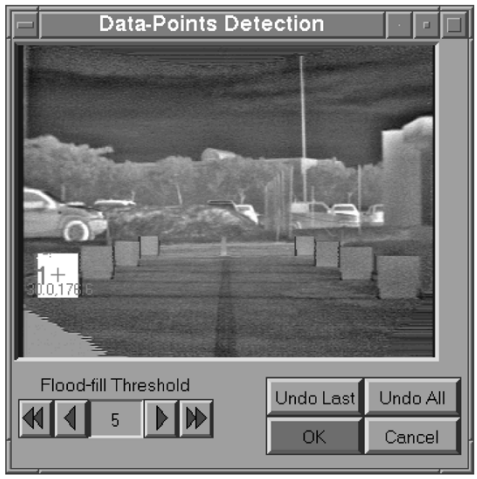

Fig. 2. Window of the graphical calibration tool showing the calibration setup.

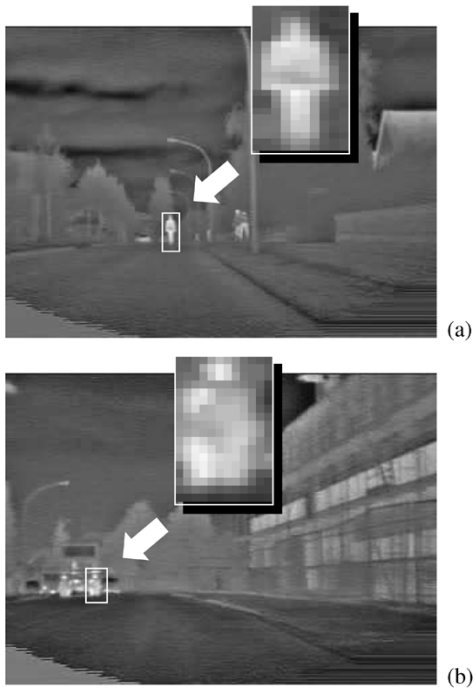

Fig. 3. Two small bounding boxes, enclosing (a) a faraway pedestrian and (b) a fake pedestrian.

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

Moreover, the algorithm has to be designed considering that the input data are low resolution (320×240) digital images. All these design choices influence the performance of the system in terms of the distance range of the detection.

A. Setup of the Vision System

The camera position is fixed by physical constraints and aesthetical choices (see Fig. 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 m (see Fig. 2); the relation between three-dimensional (3-D) coordinates of these

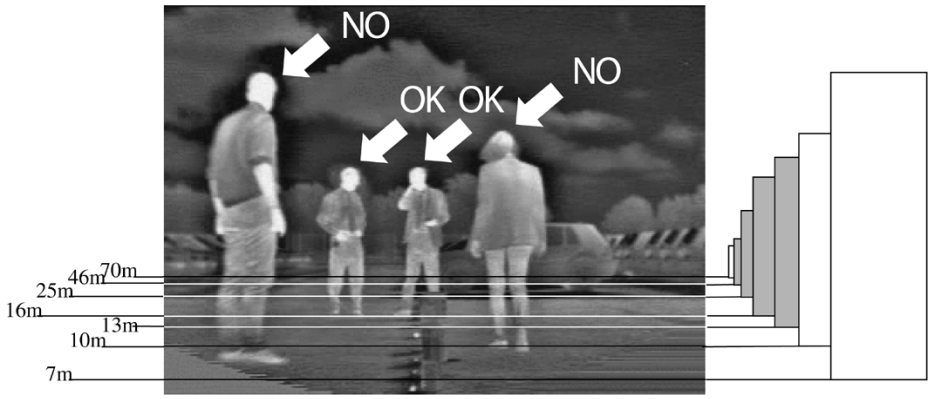

Fig. 4. Pedestrians of different heights standing at different distances and a bounding box containing a 170-cm-tall pedestrian at different distances; in white, the feasible detection range for a 170-cm-tall pedestrian.

points and the corresponding pixels in the image is used to compute camera extrinsic parameters.

The computed parameters are then used for all future relationships between 3-D 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 can be supposed to hold in the area close to the vehicle (up to 20 m) even in the presence of hills or bumps. Conversely, in the faraway area (more than 20 m), less confident results may be obtained. To reduce these errors, a software image-stabilization procedure has been developed [13].

B. Definition of the Target

Specific size and aspect ratio are used to define targets. The size of a pedestrian is chosen as follows: 1) height: 160 cm $\div$ 200 cm and 2) width: 40 cm $\div 80$ cm. The large tolerance on the width takes into account different pedestrian postures (e.g., the typical walking positions of pedestrians crossing the observer's trajectory). Actually, only the combinations of height and width satisfying specific limits on aspect ratio are considered (a range of 2.4–4.0 is assumed for the height/width ratio).

C. Detection Range

The presence of a pedestrian is checked for in different-sized bounding boxes placed at different positions in the image. In the assumption of a flat road, perspective constraints allow us to limit the search, decreasing computational time.

Moreover, since this attentive technique relies on symmetry and morphological characteristics, not all bounding boxes need to be checked due to detail content. In fact, too large bounding boxes may contain a too detailed shape, showing too many disturbing small details. In other words, the presence of texture (not only caused by clothing) and the many different human postures that must be taken into account would make the detection difficult.

On the other hand, very small bounding boxes feature a very low information content. In these situations, it is easy to obtain false positives, since many road participants (other than pedestrians) 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 Fig. 3.

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

- smallest bounding box is 28×7 pixels;
- largest bounding box: 100×40 pixels.

The limits on the bounding box height (28 and 100 pixels) were experimentally determined, while the limits on the bounding box width (7 and 40 pixels) were computed using the limit values for the target height and width.[1] Indeed, this choice leads to a limited detection area in front of the vehicle, as described in the following.

Assuming a flat road, the calibration is used to fix the correspondence between

- distances in the 3-D world and lines of the image
- size of 3-D targets and the size of bounding boxes in the image.

Distances from 7 to 70 m are considered in Fig. 4 as an example. For reference purposes, the image also shows the bounding box corresponding to a 170-cm-tall pedestrian at the different distances (the farther, the smaller).

Fig. 4 shows in gray the bounding boxes that comply with the above specifications on the bounding box size. The distance range in which the detection of a 170-cm-tall pedestrian can take place $(13 \text{ m} \div 46 \text{ m})$ is also shown in white. As can be seen, not all pedestrians can be detected, due to their size in the image. Indeed, the extension of the search to a 160–200-cm height range for the target would further narrow the detection range.

The graph in Fig. 5 shows the working area of the system. The minimum distance, given by the setup, at which pedestrians can be completely seen is represented by the vertical dashed line. On the other hand, the specifications 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 additional considerations, deriving from the definition of the bounding box size, need to be made in order to localize the region of the graph that represents the actual

[1] $7 = 28/(160/40), 40 = 100/(200/80)$.

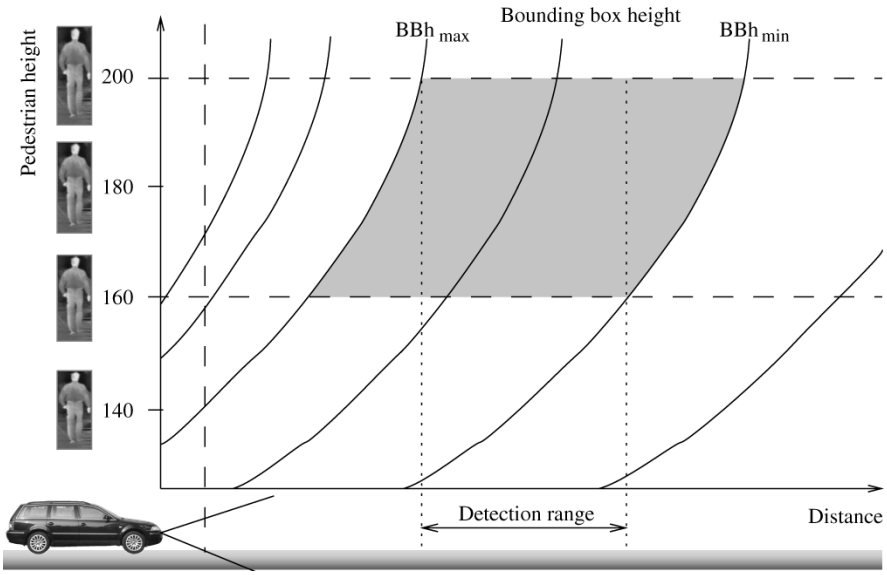

Fig. 5. Detection range.

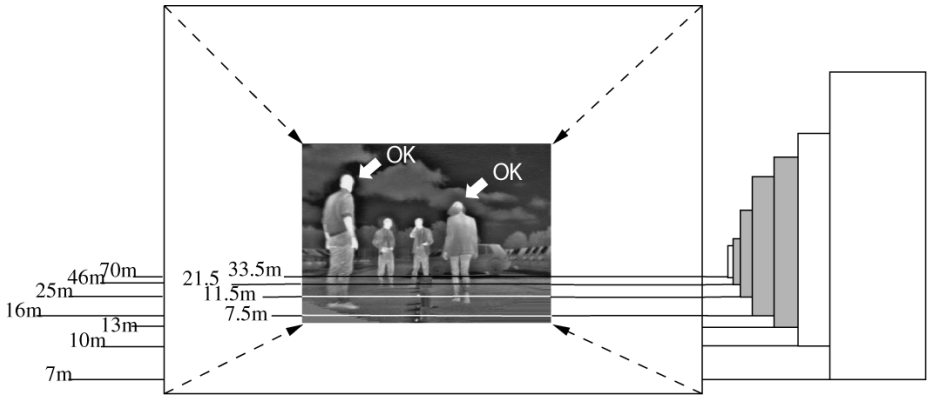

Fig. 6. After subsampling, close pedestrians fall in the detection range.

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 a bounding box with a given height in pixels. Given the range of bounding boxes height $\mathrm{BBh}_{\min} \div \mathrm{BBh}_{\max}$, the working range of the systems is depicted as the intersection of the search area described above with the area that extends between the two iso-bounding box mappings corresponding to $\mathrm{BBh}_{\max}$ pixels and $\mathrm{BBh}_{\min}$ pixels, shaded in Fig. 5.

In order to be sure that for a given distance all pedestrians in the height range (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 detection range. Assuming all the values given before, the resulting detection range is 15 m $\div$ 43.5 m.

Considerations may be made on the behavior of the detection range with the increment or decrement of the target height range; in other words, extending the target height range to include children would shorten the system detection range.

D. A Multiresolution Approach for an Extended Detection Range

As mentioned before, processing the original image does not allow for the detection of all pedestrians; conversely, only pedestrians in a specific detection range can be localized. While the low information content for too distant pedestrians cannot be compensated for, a subsampling of the image can extend the detection range to include close pedestrians. Namely, after subsampling, the original image the size of bounding boxes enclosing close pedestrians falls within the limits imposed by the algorithm on maximum bounding box size. The subsampling process also requires a new mapping between pixels and 3-D world (see Fig. 6).

Thus, in order to extend the detection range to a closer region, processing is performed on a smaller version of the original image. Actually, the image is first subsampled and processed to look for pedestrians in a close distance range, then processed again at the original resolution to search for pedestrians in a farther distance range (as justified in Section IV-D).

In the processing of the subsampled image, the size of the investigated bounding boxes is the same used for the original image. Given that the image is now smaller by a factor s ($1 : s$ subsampling), the use of the same bounding box size brings to the localization of pedestrians that in the original image are contained into bounding boxes s times larger and wider than the predefined size range. In other words, the system is now able to detect larger—and, thus, closer—pedestrians.

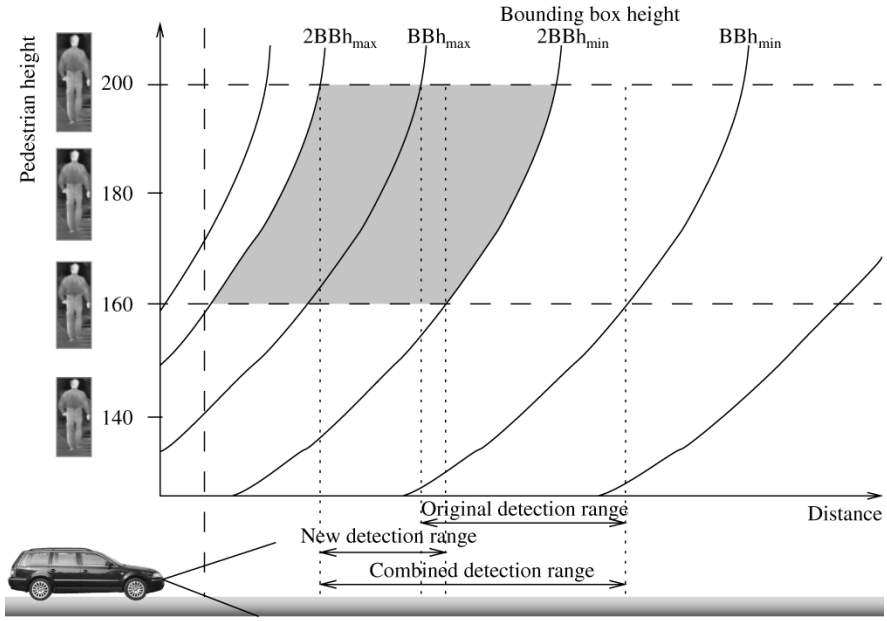

Fig. 7. Extended detection range.

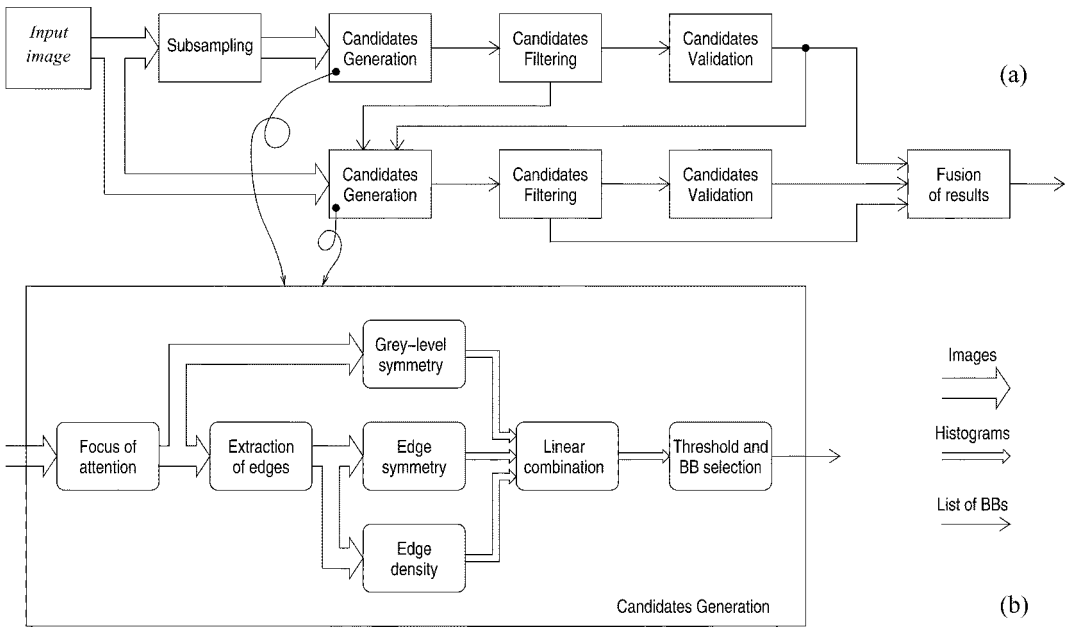

Fig. 8. (a) Block diagram of the algorithm and (b) detailed flow chart of candidates' generation.

For example, Fig. 7 shows the new detection range when using a 1:2 subsampled image (which is equivalent to use the range of bounding boxes height $2 \times \mathrm{BBh_{min}} \div 2 \times \mathrm{BBh_{max}}$ on the original image). The graph shows the two detection ranges for the original and subsampled images. In general, they have the following characteristics:

- the higher the subsampling rate, the closer the new detection range and the shorter it gets;
- the two detection ranges can overlap.

With the current setup and design choices, the distance explored when the original image is used ranges from 15 to 43.5 m, while the detection range investigated when using a 1:2.15 subsampled image is $7\ \mathrm{m} \div 20\ \mathrm{m}$. The subsampling rate has been computed so as to push the minimum explored distance to the limit imposed by the setup constraints (7 m). The two areas overlap and, thus, one search area needs to be reduced in order to avoid duplicate analysis. Therefore, the search for distant pedestrians is actually performed from 20 to 43.5 m.

IV. Algorithm Description

As mentioned in Section III-D, the core of the algorithm is repeated for two different image resolutions [see Fig. 8(a)]. It is divided into the following parts:

1) localization of areas of interest (focus of attention) and generation of possible candidates based on symmetry;
2) candidates filtering to remove errors, based on non-pedestrian characteristics;

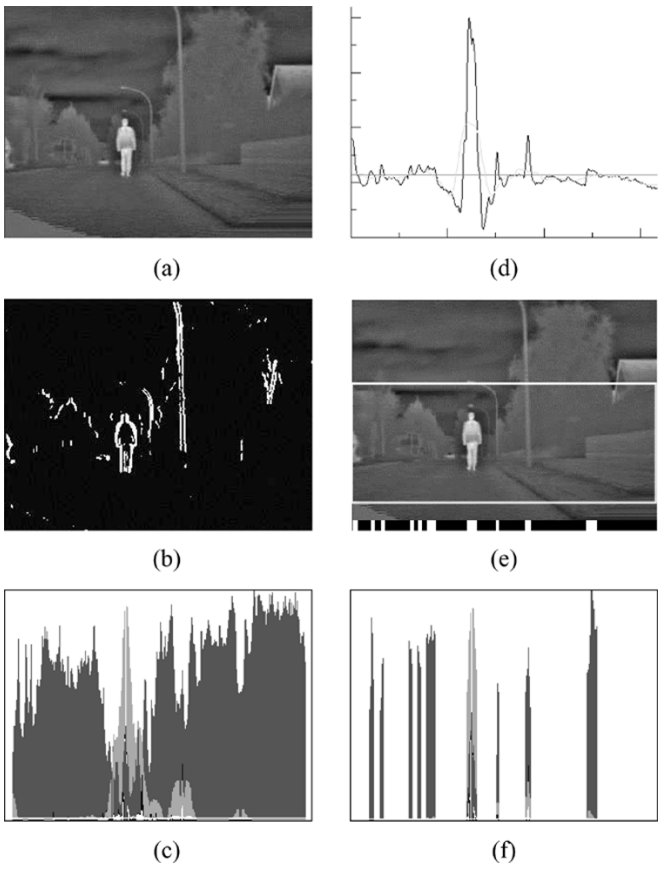

Fig. 9. Computation of symmetries and focus of attention. (a) Original image; (b) vertical edges image; (c) symmetry of gray levels (dark gray), symmetry of vertical edges (black), density of vertical edges (light gray), and a combination (white); (d) histogram of gray levels together with its global average and local average; (e) positions of possible vertical symmetry axes (in white) and search stripe; (f) histograms are computed only in correspondence to the white dashes shown in the bottom of (e).

3) candidates validation on the basis of a match with a model of a pedestrian;
4) fusion of the results of the two iterations.

A. Candidates Generation

The low-level part of the algorithm, depicted in Fig. 8(b), is mainly based on the computation of symmetries. First, 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. These areas need to match specific aspect-ratio and size constraints that are typical of a pedestrian shape, also taking into account perspective issues. The density of edges in these areas is also considered.

Fig. 9 shows as an example: the original input image [Fig. 9(a)], a binary image containing its vertical edges [Fig. 9(b)], and a number of histograms [Fig. 9(c)] computed by maximizing, for each vertical symmetry axis as follows:

- symmetry of gray levels (dark gray);
- symmetry of vertical edges (black);
- density of vertical edges (light gray)

among the different bounding boxes centered on the same axis. The white histogram presents a combination of all the above; it can be observed that the pedestrian presents high local peaks in all histograms and in their combination, as well.

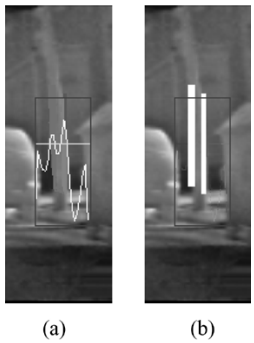

Fig. 10. Bounding box framing a tree. (a) Vertical edges histogram and the threshold value and (b) vertical contours.

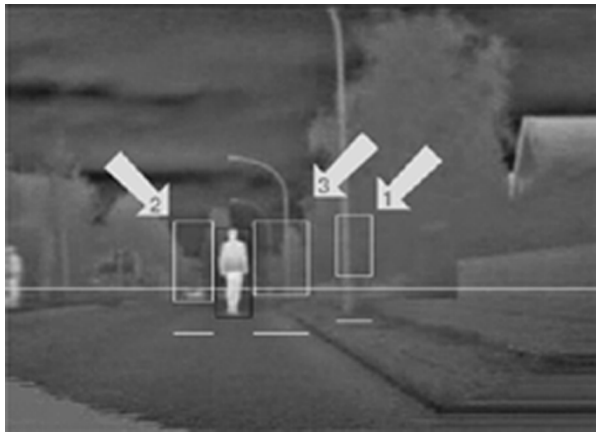

Fig. 11. Elimination of bounding boxes after the resize step: for each bounding box, the original base is displayed with a segment while the horizontal line indicates the horizon. The black box is not modified by the resize step.

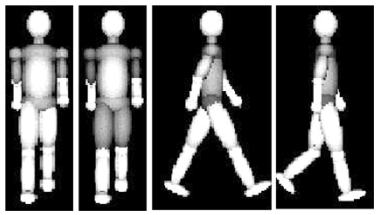

Fig. 12. Models representing different clothings, postures, and points of view.

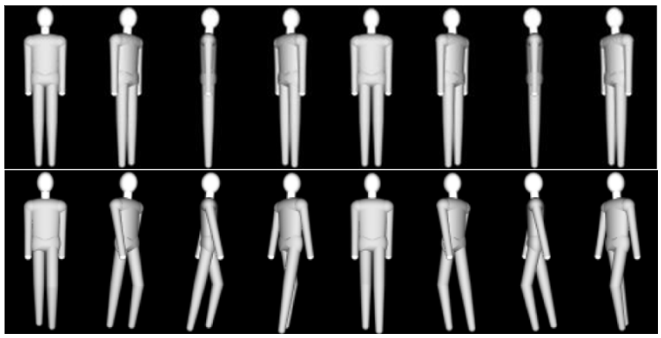

Fig. 13. Examples of eight points of view for a standing and walking pedestrian.

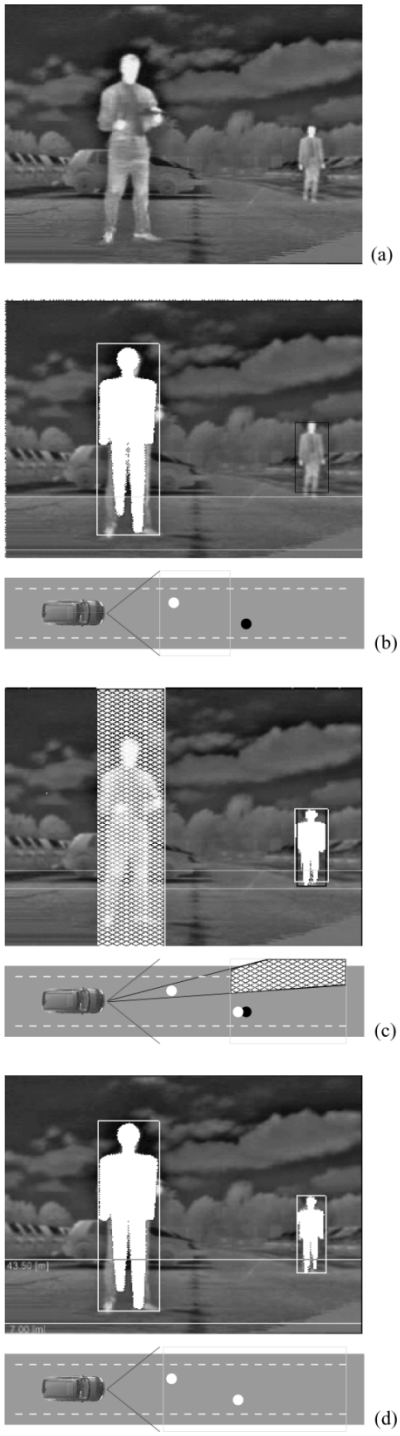

Fig. 14. Example of the fusion of the results achieved working at the two resolutions. (a) Input image, (b) results of low resolution processing, (c) results of original resolution processing, and (d) final results.

Candidates are generated by thresholding the resulting histogram. Each over-threshold peak corresponds to a bounding box containing the shape of a potential pedestrian.

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. Perspective constraints limit the search to a stripe of the image [see Fig. 9(e)]. Moreover, considerations that are generally true for images in the IR domain permit to reduce the number of symmetry axes to be examined: a filter has been defined to select symmetry axes in warm image areas only. 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 to remove small peaks close to high peaks, while the overall average is used to mask out histogram peaks in cold areas. Fig. 9(d) shows the histogram, its average and its low-pass filtered version. As explained before, assuming that 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. As an example, as shown in Fig. 9(e), vertical symmetry axes intersecting the white portions of the bottom of the image are considered, while the remaining ones (intersecting black dashes) are neglected. Fig. 9(f) shows the actual histograms computed in correspondence to the white dashes only. This technique improves both the detection (false positives are reduced in number) and computational time.

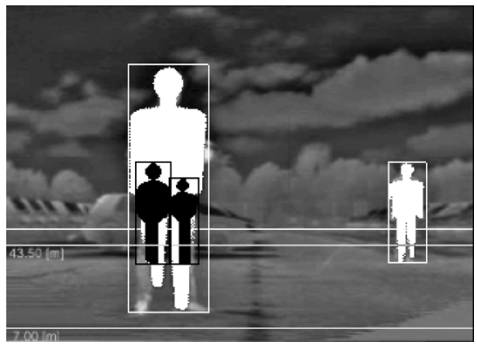

Fig. 15. False-positives result if the area covered by the close pedestrian is not eliminated when looking for far pedestrians.

The bounding box list is then passed on to the next phase, which is in charge of removing false positives.

B. Candidates Filtering

Unfortunately, artifacts featuring strong vertical edges are likely to confuse the bounding boxes generation phase. A specific filter has been designed to discard such false positives.

The vertical binarized edges inside each bounding box are considered. Actually, the edges above and below the bounding box are also considered, since objects (e.g., poles, columns, trees, road signs, edges of building, etc.) can extend outside the box. A vertical histogram is computed using vertical edges; the peaks of the histogram higher than a given threshold indicate the positions of a significant amount of vertical edges. These areas are further investigated to detect the exact position of vertical contours by building chains of contiguous edges. Short contours are discarded. If the bounding box is centered on a high amount of vertical contours, it is discarded as a false positive.

An example of the application of this filter is shown in Fig. 10. This figure shows a bounding box framing a tree trunk; Fig. 10(a) displays the vertical edges histogram and the threshold value, while Fig. 10(b) shows the vertical contours.

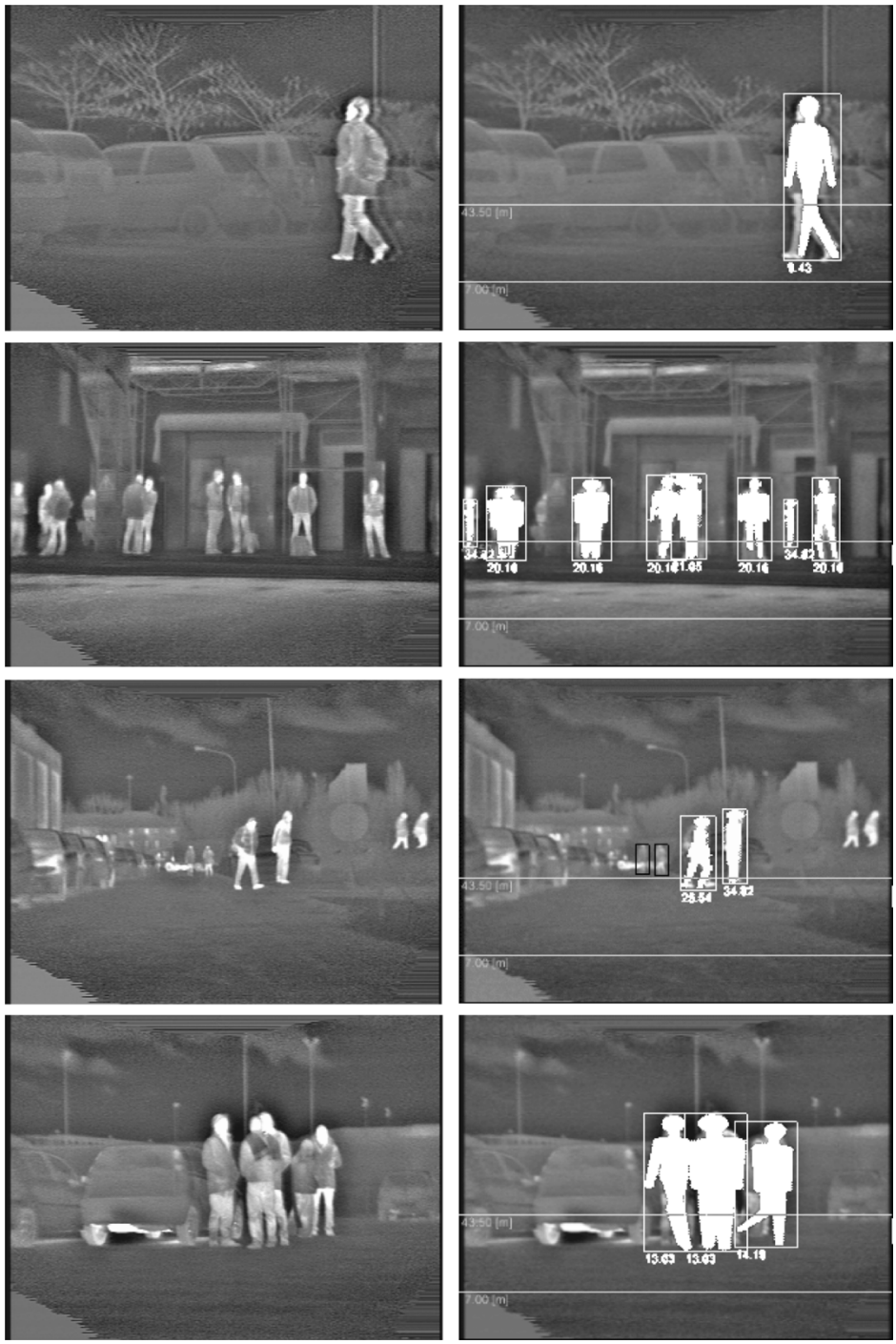

Fig. 16. Results of pedestrian detection in different situations: with complex or simple scenarios or with one or more pedestrians. The distance (in meters) is displayed below the boxes; the two horizontal lines encode the range for which pedestrians are searched.

Other criteria based on the analysis of the vertical histogram of edges are used to eliminate false positives [13]. A bounding box is removed in the following cases.

- When the center of the histogram is empty: This is true for large poles, pylons, and columns, even if they are not perfectly vertical.
- When more than half of the histogram is empty: This is true for large vertical poles, pylons, and columns.
- 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 histogram is concentrated in two small areas that contain more than 80% of the contributions: This is true for thin vertical poles, pylons, and columns.

Each surviving bounding box is then reduced in height and width in order to fit the internal presence of edges. The bounding boxes that have been resized too much, due to the absence of edges in their border regions, are removed, since pedestrians are characterized by a uniform distribution of edges.

The surviving bounding boxes are further examined in order to eliminate bounding boxes that

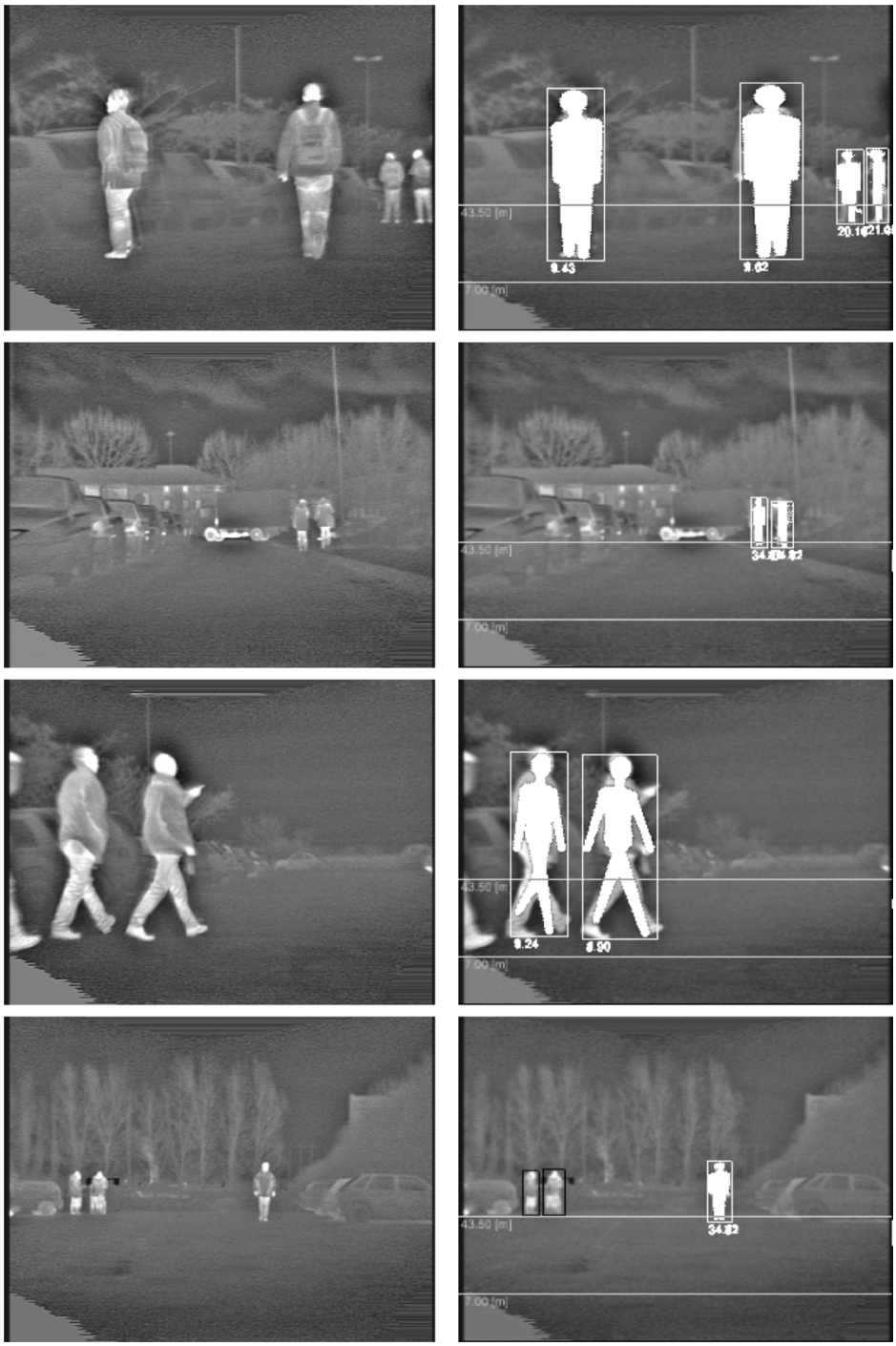

Fig. 17. Other results of pedestrian detection in different situations: with complex or simple scenarios or with one or more pedestrians. The distance (in meters) is displayed below the boxes; the two horizontal lines encode the range for which pedestrians are searched.

- due to this resize operation are completely over the horizon (arrow 1 in Fig. 11);
- no longer meet perspective constraints (arrow 2 in Fig. 11);
- no longer meet the original assumptions on aspect ratio (arrow 3 in Fig. 11).

The resize operation can move the base of the box. After this operation, some boxes may lie beyond the actual search area. These boxes are considered to be *guesses* and are not passed on to the validation step.

C. Candidates' Validation

Each surviving bounding box is validated through a match with a 3-D model of the human shape. This filter, based on shape and/or thermal patterns, is used to remove candidates that do not present a human shape. The 3-D models represent different postures and viewing angles of the human shape.

The idea of generating the models at run-time and performing an exhaustive search for the best configuration has been discarded, since it is time consuming and does not fit real-time

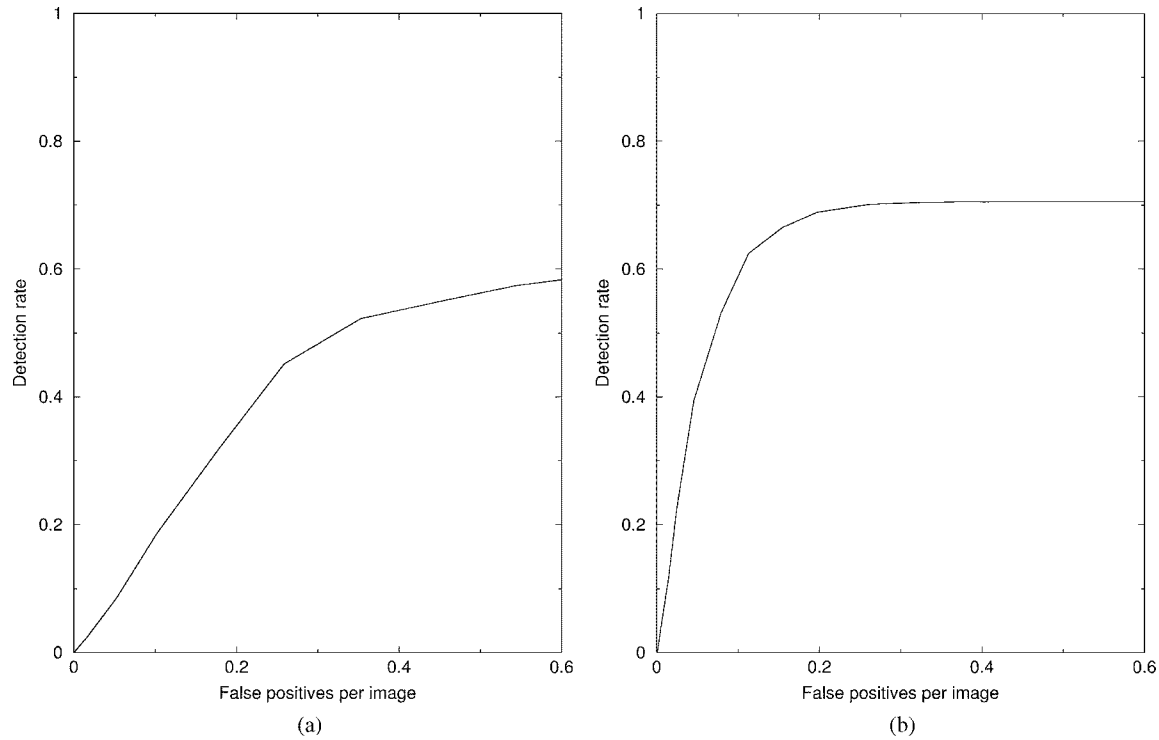

Fig. 18. ROC curve on (a) test sequence 1 and (b) test sequence 2.

criteria. A selection of precomputed configurations has been chosen.

The possibility of adapting the models to real images attributing different gray values to the body parts in order to encode different body temperatures has also been considered and tested. In fact, generally, the head and hands are not covered by clothes and, thus, more heat escapes from them with respect to the trunk or limbs both in winter and summer. Fig. 12 shows some examples of models representing different clothing. Anyway, detailed investigations about models encoding thermal differences have not be made so far. Instead, most of the investigation is focused on using a large number of different shapes.

Two degrees of freedom are sufficient to obtain a good match in most situations and are used to generate the complete matching set: postures and point of view. A third degree of freedom (size) is implicit in the match process. A first set of eight configurations obtained combining four points of view with two positions were initially tested but demonstrated to be not sufficiently reliable. A new set of 72 configurations were finally chosen. They were obtained by combining eight different points of view with nine positions (one standing and eight walking). Fig. 13 shows the 72 configurations generated using a smoother model, also taking into account the actual viewing angle, orientation, and height of the camera on the test vehicle.

Each model is scaled to the bounding box size and overlapped to it using different displacements to cope with small errors in localization of the box. The matching is implemented through a simple and fast cross-correlation function. The result is a percentage rating the quality of the match. A threshold is applied for the final evaluation.

This filter has proven to be effective in most cases, both in the identification of pedestrians and in the exclusion of bounding boxes that do not contain humans.

Anyway, the localization of pedestrians is difficult in some situations, such as bikers, running people, or when the bounding box is not precise.

D. Fusion of the Results

The results obtained by separately processing the undersampled image and the original-sized one need to be fused together. Indeed, even if the two detection ranges are disjoint, they are contiguous anyway. Therefore, a trivial joining of these two results may lead to double detections and a method has been devised to join the two results more effectively, which is based on the following considerations.

First, a correct detection in the close area eliminates the need to perform the search in the same direction in the faraway area. For this reason, as mentioned before, the close-range processing is performed first and the position of close pedestrians found is considered to limit the search area in the second far range phase [see close pedestrians to be used as mask in Fig. 8(a)].

Second, as explained in Section IV-B, due to the resize step, the candidate filtering can output some results that are beyond the actual close search area and, thus, are not passed on to the validation step. These low confidence results are worth being propagated anyway to the far-range processing [see Fig. 8(a)] for a validation using correct criteria, which are only available during the search in the faraway area.

An example is shown in Fig. 14. As can be seen in Fig. 14(b), the close-range processing detects a near pedestrian (the white box with the model superimposed) and a guess just beyond the search area (the black box). A bird's eye view of the results is also sketched.

Following the above considerations, the results of the first phase (low resolution) are taken into account to limit the search area in the second phase (original resolution). More specifically, no further search for symmetries is performed in the image area where the close pedestrian was found. Furthermore, the guesses attained in the first phase is passed on to the second phase and added to the list of new candidates generated by the search for symmetries. Together with the new candidates, the guess will be filtered and resized, and possibly validated. Fig. 14(c) displays the results of the far detection range processing. Two boxes are visible for the far pedestrian: the black one corresponds to the approximate localization deriving from the guess [the black box in Fig. 14(b)], while the white one corresponds to the new correct detection. Both boxes have been validated as representing a pedestrian by the 3-D models. As displayed in the bird's eye view, the area covered by the close pedestrian is not investigated in the second phase. Besides speeding up the search, this avoids false detections that may originate by misinterpretations of parts of the close pedestrian (see Fig. 15).

As in this example, generally pairs of similar bounding boxes may be generated when a guess generated in the first stage is validated in the second stage and, at the same time, a new detection is obtained in the second stage for the same pedestrian. An extra step devoted to the fusion of similar bounding boxes is needed. In case two, bounding boxes are overlapped (similar in position and size) and the selection is based on their detection confidence and match with the 3-D model.

- Whether one of them was rated as a guess and the other as a correct result, the guess will be dropped and the correct result maintained.
- In the case that both bounding boxes received the same confidence, two criteria are adopted: the larger is preferred if both are guesses, while the vote assigned by the match with the 3-D models is used to decide which one should be kept and which one should be discarded when both are validated boxes.

Fig. 14(d) shows the final result of the discussed example after the fusion procedure.

V. Discussion of Results and Conclusion

Figs. 16 and 17 show a few results of pedestrian detection in IR images in a number of different situations. The two horizontal lines encode the detection range in which pedestrians are searched for $(7\ \text{m} \div 43.5\ \text{m})$. In correspondence to a detected pedestrian, the image shows a white bounding box and the model that best matches the pedestrian. The guesses out of the detection range are also displayed using a black bounding box. Please note that, as mentioned in the introduction, the images produced by our software have been modified by hand in order to adapt to grayscale printing. The result shows that the system is able to detect one or more pedestrians, even in the presence of a complex background.

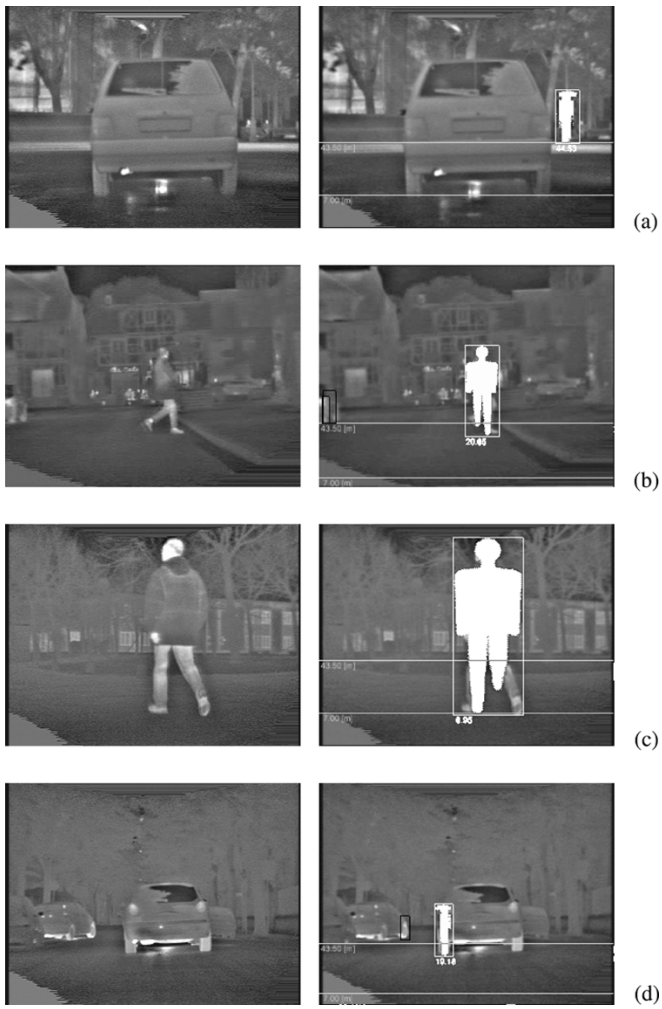

Fig. 19. Critical situations. (a) and (d): The algorithm finds false positives due to a noisy background. (b) and (c): a wrong 3-D model is matched against a correct detection due to the limited number of 3-D models used.

The two major critical situations, exemplified in Fig. 19, are

- in the presence of a complex background, artifacts or objects other than pedestrians are occasionally detected [see Fig. 19(a)];
- the algorithm does not miss the detection of a pedestrian, but a wrong model is matched [see Fig. 19(b)].

Since there is no widely available image test set, a direct comparison to other systems is not possible. However, a quantitative performance analysis was carried out to measure performance and to allow our approach to be compared to other systems for which quantitative results were published (such as [6] and [10]), even if these studies do not refer to the IR domain.

Ground truth was collected for a test set of 4111 images containing 5082 pedestrian instances. A graphical tool specifically developed to annotate pedestrians and assess statistics about results was employed [14]. Test images were acquired in different times of day, weather, and scenarios and include very complex situations with many pedestrians, occlusions, and groups of people. A quantitative evaluation of results is given by means of ROC curves. We run the system over the test set for which

ground truth is available and computed the number of correct detections and false positives. Results were considered correct if they adequately overlapped with the annotated boxes (see [14]).

The detection and false positives rates depend on the parameters and thresholds used in each stage of the system. Since the execution flow chart of our system is complex (stages are repeated twice and there is a feedback of the output from the first low resolution run to the second high resolution run, see Fig. 8), a separate analysis of the different stages (candidates generation, filtering, and validation) poses severe problems. We decided to evaluate the behavior of the whole system with a single ROC curve obtained by varying the correlation threshold used in the matching with the models. In fact, due to the high complexity of the system, optimizing the thresholds one at a time would have been prohibitive.

The graph in Fig. 18(a) presents the ROC curve obtained for the complete test set. The false-positive rate is very low, but the detection rate is affected by the very high number of complex scenes with groups of people.

The graph in Fig. 18(b) shows the curve obtained running the system over a part of the test sequence: complex scene were eliminated and significantly occluded pedestrians maintained, in order to reflect the assumptions upon which the algorithm was designed. The same choice was adopted in other studies ([6] and [10]); therefore, this second sequence represents a testing condition similar to theirs. Moreover, this is a reasonable assumption, since the most dangerous pedestrian is the closest one, who is least occluded. This new test set is composed of 2152 images with 1496 pedestrian samples. A high number of frames without pedestrians was left to particularly challenge the system with respect to false alarms. In this case, a detection rate of 70% is achieved with 0.2 false positives per image. This very low false-alarm rate is a good result in order to not to flood the driver with too many unnecessary warnings, which could decrease the driver's confidence in the system.

Currently, the system is based on the processing of single shots only; one of the most important enhancements will be the integration of a tracking procedure that will allow us to improve the final results, both reducing temporary misdetections caused by noise or occlusions and decreasing the number of false alarms. In addition, tracking would permit us to speed up the processing; in fact, due to the similarity between two subsequent images, only a subset of the models can be used in the correlation.

Moreover, in the case of walking pedestrians, the sequence of 3-D models to be used in the correlation may suggest the pedestrian moving direction. Furthermore, improvements may be obtained by using a more representative set of 3-D models to reproduce the average pedestrian appearance with more accuracy. In fact, we realized that the models described in this work (see Fig. 13) are too thin compared to a dressed person.

Concerning time performance, the system has been tested on a 1.8 GHz Athlon XP (FSB 266 MHz) with 512 MBytes DDR@400 MHz. The pedestrian detector proved to be very efficient: the average time required for the processing of a frame is 127 ms (correspondent to a frame rate of about 8 frames/s). Indeed, the actual frame-processing time depends on the number of pedestrians.

The algorithm developed so far proves to be effective in different situations. Extensive tests are being carried out in different seasons (winter/summer). The results are promising, though the system is not ready for deployment. We believe that significant improvements could be achieved by using stereo IR vision and making the system more robust by fusing visual information with radar data.

REFERENCES

[1] H. Nanda and L. Davis, "Probabilistic template based pedestrian detection in infrared videos," presented at the IEEE Intelligent Vehicles Symposium 2002, Paris, France, June 2002.

[2] F. Xu and K. Fujimura, "Pedestrian detection and tracking with night vision," presented at the Proc. IEEE Intelligent Vehicles Symposium 2002, Paris, France, June 2002.

[3] Y. L. Guilloux and J. Lonnoy, "PAROTO project: The benefit of infrared imagery for obstacle avoidance," presented at the Proc. IEEE Intelligent Vehicles Symposium 2002, Paris, France, June 2002.

[4] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of night-vision system," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, pp. 203–209, Sept. 2002.

[5] M. Bertozzi, A. Broggi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proc. IEEE Intelligent Vehicles Symp.*, Detroit, MI, Oct. 2000, pp. 215–220.

[6] D. M. Gavrila and J. Geibel, "Shape-based pedestrian detection and tracking," presented at the Proc. IEEE Intelligent Vehicles Symposium, Paris, France, June 2002.

[7] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking pedestrian recognition," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, pp. 155–163, Sept. 2000.

[8] L. Zhao and C. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Trans. Intell. Transport. Syst.*, vol. 1, pp. 148–154, Sept. 2000.

[9] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 781–796, Aug. 2000.

[10] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Procs. IEEE Int. Conf. Computer Vision*, Nice, France, Sept. 2003, pp. 734–741.

[11] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki, "Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection," in *Proc. IEEE Intelligent Vehicles Symp.*, Columbus, OH, June 2003, pp. 505–510.

[12] M. Bertozzi, A. Broggi, T. Graf, P. Grisleri, and M. Meinecke, "Pedestrian detection in infrared images," in *Proc. IEEE Intelligent Vehicles Symp.*, Columbus, OH, June 2003, pp. 662–667.

[13] M. Bertozzi, A. Broggi, M. Carletti, A. Fascioli, T. Graf, P. Grisleri, and M. Meinecke, "IR pedestrian detection for advanced driver assistance systems," *Lecture Notes Comp. Sci.*, vol. 2781, pp. 582–590, 2003.

[14] M. Bertozzi, A. Broggi, P. Grisleri, A. Tibaldi, and M. D. Rose, "A tool for vision based pedestrian detection performance evaluation," in *Proc. IEEE Intelligent Vehicles Symp.*, Parma, Italy, June 2004, pp. 784–789.

Massimo Bertozzi (S'95–A'98) received the Dr.Eng. (M.S.) degree in electronic engineering and the Ph.D. degree in information technology, both from the Università di Parma, Parma, Italy, in 1994 and 1997, respectively. His thesis was on the implementation of simulation of Petri nets on the CM-2 massive parallel architecture.

He is a Researcher in the Dipartimento di Ingegneria dell'Informazione, Università di Parma. His research interests focused mainly on the application of image processing to real-time systems and to vehicle guidance, the optimization of machine code at assembly level, and parallel and distributed computing.

Dr. Bertozzi chaired the local IEEE student branch from 1994 to 1997.

Alberto Broggi (S'93–A'96) received the Dr.Eng. degree in electronic engineering and the Ph.D. degree in information technology from the Università di Parma, Parma, Italy, in 1990 and 1994, respectively.

From 1994 to 1998, he was an Associate Researcher in the Dipartimento di Ingegneria dell Informazione, Università di Parma. From 1998 to 2001, he was a Professor of Artificial Intelligence in the Dipartimento di Informatica e Sistemistica, Università di Pavia. He is now again with the Università di Parma, where he has obtained full Professor recognition. He has authored more than 120 refereed publications in international journals, book chapters, and conference proceedings and has delivered invited talks at many international conferences His research interests mainly include real-time computer vision approaches for the navigation of unmanned vehicles and visual perception for intelligent vehicles.

Dr. Broggi is the Editor-in-Chief of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and is a Member of the IEEE ITS Council Executive Committee.

Alessandra Fascioli (S'97–M'99) received the Dr.Eng. (M.S.) degree in electronic engineering and the Ph.D. degree in information engineering from the Università di Parma, Parma, Italy. Her M.S. thesis was on stereo vision-based obstacle localization in automotive environments.

Alessandra Fascioli is currently temporary researcher at the University of Parma. Her research interests focus on Real-Time Computer Vision and Computer Architectures for Automatic Vehicle Guidance. She is also interested in image-processing techniques based on the Mathematical Morphology computational model.

Dr. Fascioli chaired the local IEEE student branch from 1997 to 1999.

Thorsten Graf received the diploma (M.Sc.) degree in computer science and the Ph.D. degree (his thesis was on "Flexible Object Recognition Based on Invariant Theory and Agent Technology") from the University of Bielefeld, Bielefeld, Germany, in 1997 and 2000, respectively.

In 1997, he became a Member of the "Task Oriented Communication" graduate program, University of Bielefeld, funded by the German research foundation DFG. In June 2001, he joined Volkswagen Group Research, Wolfsburg, Germany. Since then, he has worked on different projects in the area of driver assistance systems as a Researcher and Project Leader. He is the author or coauthor of more than 15 publications and owns several patents. His research interests include image processing and analysis dedicated to advanced comfort/safety automotive applications.

Marc-Michael Meinecke received the Dipl.Ing. degree in electronical engineering from the Technical University of Braunschweig, Braunschweig, Germany, in 1997 and the Ph.D. degree from the Technical University of Hamburg-Harburg, Hamburg, Germany, in 2001 (his Ph.D. dissertation was on "Optimized Waveform Design for Automotive Radars.")

In January 2001, he joined the Volkswagen Group Research, Wolfsburg, Germany. Since then, he has worked on several research projects in the area of driver assistance electronics as a Project Leader, where he is responsible for radar technology, precrash detection, and pedestrian recognition systems. Currently, he is involved in the SAVE-U research project (precrash and pedestrian recognition), which is funded by the European Commission. He is the author or coauthor of more than 25 publications and ows about 30 patents.

Dr. Meinecke was awarded the Volkswagen Research Prize in 2002.