2004 IEEE Intelligent Vehicles Symposium
University of Parma
Parma, Italy • June 14-17, 2004

A Multi-resolution Approach for Infrared Vision-based Pedestrian Detection

A. Broggi, A. Fascioli, M. Carletti
Dipartimento di Ingegneria dell'Informazione
Università di Parma
Parma, I-43100, Italy
{broggi,fascioli,carletti}@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 presents the improvements of a system for pedestrian detection in infrared images.**

The system is based on a multi-resolution localization of warm symmetrical objects with specific size and aspect ratio; the multi-resolution approach allows to detect both close and far pedestrians. A match against a set of 3D models encoding human shape's morphological and thermal characteristics is used as a validation process to remove false positives.

No temporal correlation, nor motion cues are used for the processing that is based on the analysis of single frames only.

I. Introduction

Yearly, in the EU zone, more than 200,000 pedestrians are injured and about 9,000 are killed in accidents; namely, the second largest source of traffic-related injuries in the EU is due to accidents that involve pedestrians.

Therefore, it is not surprising that a primary goal for several research groups is the the development of in-vehicle assistance systems for reducing the effects or the number of such traffic accidents.

Among the others, the use of vision sensors and image processing methods provides a promising approach; in the last years, several different vision-based systems dedicated to detect and classify pedestrians have been developed, including shape-based methods [1]–[3], neural nets-based methods [4], texture-based methods [5], stereo [6], and motion [7], [8]. These systems must face the fact that pedestrians in the visible domain can feature different appearances mainly caused by clothes, carry-ons, illumination changes, and different postures.

Recently, thanks to the decreasing cost of infrared (*IR*) devices, the advantages and benefits of using far IR cameras have been considered (e.g. [9], [10]). Only few systems exploiting far IR cameras have been developed so far, showing that IR images can facilitate the recognition process [11]–[13].

In this paper the improvements of the system described in [12] for the detection of pedestrians in far IR images are presented. The process is now iterated at different image resolutions in order to detect both close and far away pedestrians; in addition, a match with a set of 3D models encoding shape and thermal patterns of pedestrians is used to remove candidates that do not present a human shape.

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

In the following section considerations on how design choices affect the detection range are made and the multi-resolution approach able to extend this range is presented; section III briefly sketches the algorithm and discuss the improvements; finally section IV presents the results and concludes the paper with some final consideration.

II. Detection range

Two issues have been defined when designing the system described in [12]: (*i*) the desired target, i.e. the interval of pedestrians' height and width, and (*ii*) the position and orientation of the IR camera, considering aesthetical and physical automotive requirements. Moreover, the algorithm has to cope with low resolution (320 × 240) images provided by IR sensors. All these design choices impacts on the system's capabilities and, particularly, on the distance range of the detection.

A. Setup of the vision system

The algorithm is based on monocular IR vision, therefore the distance of objects is computed using a mapping between image pixels and world coordinates. This approach strongly rely on a precise calibration of the vision system; the calibration is performed using few markers placed in known position (up to 40 meters) on flat stretch of road (see figure 1); the relation between 3D coordinates of these 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 3D world coordinates and image pixels, under the assumption of a flat road in front of the IR camera and negligible vehicle pitch. Unfortunately, vehicle oscillations due to road roughness are generally present during driving; although filtered by the dumper system, they are perceived by the camera. While the flat road assumption can be supposed to hold in the area close to the vehicle (up to 20 meters) even in presence of hills or bumps, in the faraway area (more than 20 meters) less confident results

The research described in this work was funded by Volkswagen AG.

may be obtained. Therefore, vertical oscillations need to be compensated and a specific software image stabilizer able to reduce these effects has been developed [14].

B. Definition of target

Targets are characterized by size and aspect ratio constraints. For the pedestrians height a 180 cm ÷ 200 cm range has been chosen while for the pedestrians width a 40 cm ÷ 80 cm interval is considered acceptable. The large tolerance on the width takes into account different pedestrian postures (e. g. a walking pedestrians crossing the observer's trajectory). Moreover, an additional aspect ratio constraint (2.4 ÷ 4.0) furtherly reduces the possible combinations of height and width satisfying those limits.

C. Detection range

Differently-sized bounding boxes are placed in different positions in the image. The presence of a pedestrian inside those bounding boxes is checked for.

Perspective constraints and the assumption of a flat scene allow to decrease the computational time limiting the search area. Moreover, not all bounding boxes need to be checked due to unsuitable 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 clothings) and the many different human postures that must be taken into account would make the detection hard.

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 other road participants than pedestrians and even road infrastructures may present morphological characteristics similar to a human shape (see figure 2).

Thus, it is mandatory to define a range for bounding box sizes that may lead to sufficiently accurate results. In this work a range 28×7 ÷ 100×40 pixels for bounding box size has been considered. 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 range values derived from the target height and width limits.

Unfortunately, this choice limits the detection area as described in the following.

Assuming a flat road, the calibration is used to fix the correspondence between: (*i*) distances in the 3D world and lines of the image, and (*ii*) the size of 3D targets and the size of bounding boxes in the image.

As an example, distances from 7 m to 70 m are considered in figure 3 that also shows the bounding box corresponding to a 170 cm tall pedestrian at different distances (the farther, the smaller). Green bounding boxes 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 m ÷ 46 m) is also shown in green. Due to their size in the image, not all pedestrians can be detected.

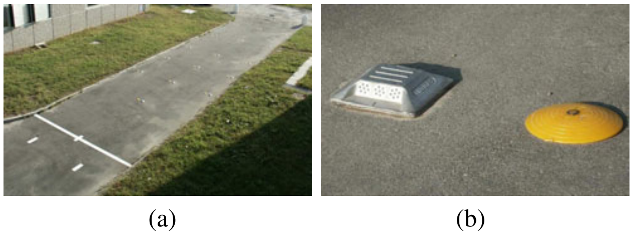

Fig. 1. Vision system calibration: (*a*) stretch of road used for the vision system calibration equipped with (*a*) near and far IR markers.

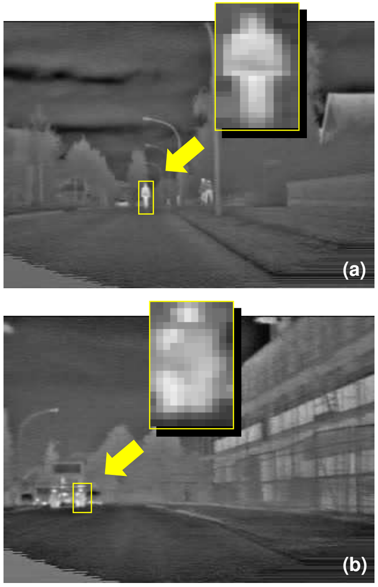

Fig. 2. Two small bounding boxes, enclosing (*a*) a faraway pedestrian and (*b*) a human shaped pattern.

The graph in figure 4 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 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 a bounding box with a given height in pixels. Given the range of bounding boxes height $BBh_{min} \div BBh_{max}$, the working range of the systems is depicted as the intersection of the search 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 4.

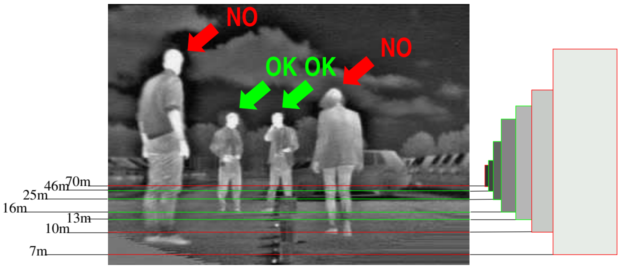

Fig. 3. Pedestrians of different heights standing at different distances and a bounding box containing a 170 cm tall pedestrian at different distances; in green the detection range for a 170 cm tall pedestrian.

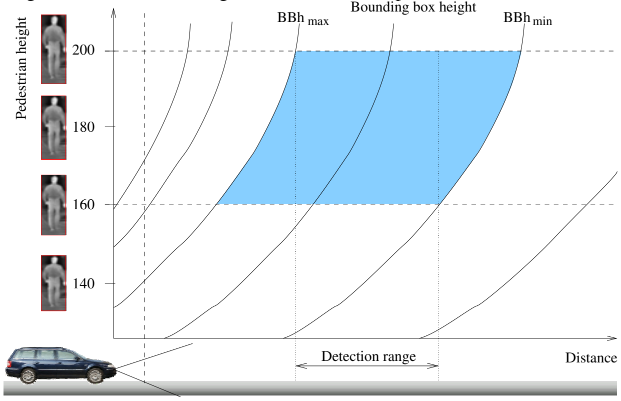

Fig. 4. The detection range.

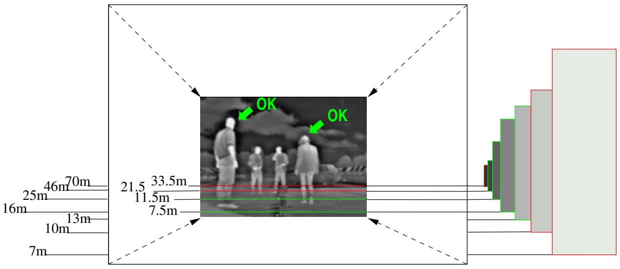

Fig. 5. After subsampling close pedestrians fall in the detection range.

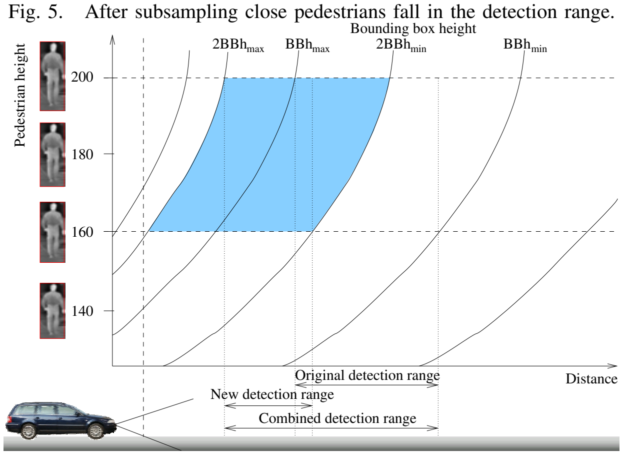

Fig. 6. The extended detection range.

In order to be sure that, for a given distance, all pedestrians in the height range can be detected, the working area has to be further limited by the two vertical dotted lines. Assuming all the values given before, the resulting detection range is 15 m ÷ 43.5 m.

The detection range varies according to the increment or decrement of the target height range. Unfortunately, extending this range to include children further narrows the detection range.

D. A multi-resolution approach for an extended detection range

As discussed in the previous paragraph, not all pedestrians can be detected due to a limiting detection range. While the low information content for too distant pedestrians cannot be compensated for, a multi-resolution approach allows an extension of the detection range and thus the detection of close pedestrians. The original image is subsampled in order to bring the size of close pedestrians to match the limits imposed by the algorithm on maximum bounding box size. The subsampling process also requires a new mapping between pixels and 3D world (see figure 5).

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 III-B).

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.

As an example, figure 6 shows the new detection range when using a $1:2$ subsampled image (that is equivalent to use the range of bounding boxes height $2 \times BBh_{min} \div 2 \times BBh_{max}$ on the original image). The graph shows the two detection ranges for the original and subsampled images. 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 m ÷ 43.5 m, while the detection range investigated when using a 1:2.15 subsampled image is 7 m ÷ 20 m. The subsampling rate has been computed so to push the minimum explored distance to the limit imposed by the camera setup constraints, namely the closest distance for which the road surface is still visible (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 m ÷ 43.5 m.

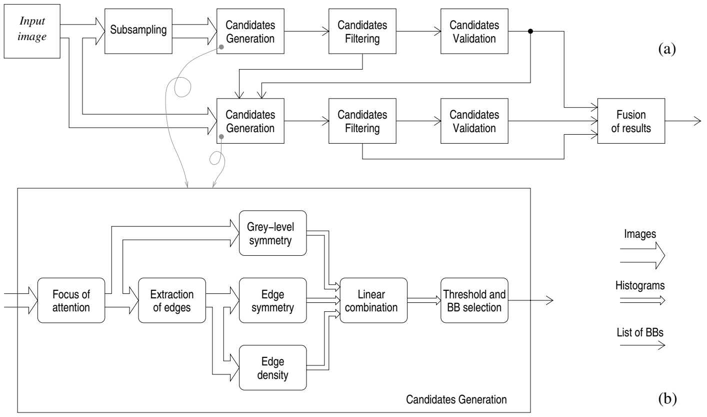

Fig. 7. (*a*) block diagram of the algorithm and (*b*) detailed flow chart of candidates generation.

III. The algorithm

As mentioned in section II-D, the core of the algorithm is repeated for two different image resolutions (see figure 7.a). It is divided into the following parts: (*i*) localization of areas of interest and generation of possible pedestrian candidates, (*ii*) candidates filtering to remove errors, (*iii*) candidates validation on the basis of a match with a model of a pedestrian, and (*iv*) fusion of the results of the two iterations.

The following assumptions have been made: the pedestrians are not occluded, the complete shape of the pedestrian appears in the image, a number of pedestrians appear simultaneously in the image but they do not occlude each other.

The part that has been noticeably improved with respect to the system presented in [12] is the validation of candidates. It was based on a match with a very simple shape model, while now each surviving bounding box is validated through a match with a 3D model of the human shape.

Moreover, a fusion phase has been added for fusing the results of the processing at the two different resolutions.

A. Candidates validation

The validation match is based on shape and/or thermal patterns and is used to remove candidates that do not present a human shape.

The 3D models represent different postures and viewing angles of the human shape and are generated from different points of view to achieve a better adaptability to real situations.

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 criteria. A selection of pre-computed configurations has been chosen.

The possibility to adapt the models to real images attributing different grey 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 are warmer than the trunk or limbs both in winter and summer. Figure 8 shows some examples of models representing different clothings. Anyway, detailed investigations about models encoding thermal differences have not be made so far. Instead, most of the investigation is focused on using a high number of different shapes.

Two degrees of freedom suffice 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 8 configurations obtained combining 4 points of view with 2 positions were initially tested but demonstrated not sufficiently reliable. A new set of 72 configurations were finally chosen. They were obtained combining 8 different points of view with 9 positions (one standing and 8 walking). Figure 9 shows the 72 configurations generated using a smoother model, taking also 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. A cross-correlation function is used for the match; the result is a percentage rating the quality of the match.

This filter has been 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.

B. Fusion of results for different resolutions

In this phase the results obtained processing the undersampled image and the original-sized image are fused together. Even if the two detection ranges are contiguous, a trivial joining of these two results may lead to double detections; thus, the following considerations have been used.

(*i*) a detection in the close area eliminates the need to

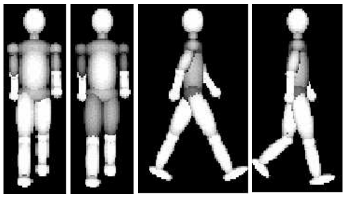

Fig. 8. A few models representing different clothings, postures, and points of view.

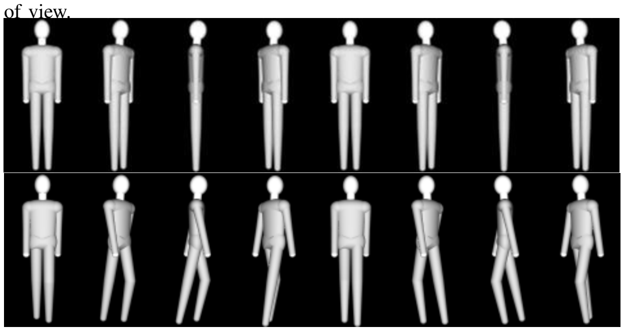

Fig. 9. Examples of 8 points of view for a standing and walking pedestrian.

perform the search in the same direction in the faraway area. Thus, the close range processing is performed first and the results of this phase are taken into account to limit the search area for far pedestrians. More specifically, no further search is performed in the image area where the close pedestrian was found. Besides speeding up the search, this avoids false detections that may originate by misinterpretations of parts of the close pedestrian (see figure 10).

(*ii*) during the processing of subsampled image, some results beyond the actual search area can be produced (*guesses*) [12]. These bounding boxes are not considered as valid at this step of the processing but are anyway worth to be propagated to the far range processing for further investigations and passed on to the second phase to be considered as potential pedestrians.

(*iii*) 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), the selection is based on their detection confidence and match with the 3D model: (*a*) whether one of them was rated as guess and the other as correct result, the guess will be dropped and the correct result maintained; (*b*) in case both bounding boxes got the same confidence two criteria are adopted: the larger is preferred if both are guesses, while the vote assigned by the match with the 3D models is used to decide which one should be kept and which one should be discarded when both are validated boxes.

IV. Discussion

The result shows that the system is able to detect one or more pedestrians even in presence of a complex background in a 7 m÷43.5 m range. Figure 11 shows a few results of pedestrian detection in infrared images. The two horizontal green lines encode the detection range. In correspondence to a detected pedestrian, the image shows a red bounding box and the 3D model which best matches the pedestrian. A yellow bounding box is used to signal the guesses out of the detection range.

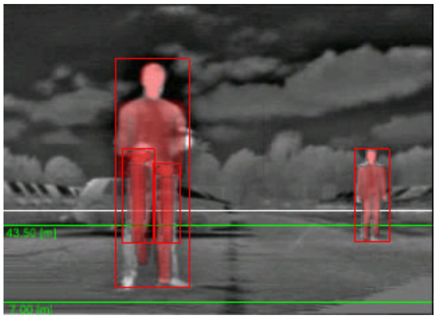

Fig. 10. False positives result if the area covered by the close pedestrian is not eliminated when looking for far pedestrians.

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 both to improve the final results and to speed up the processing. Moreover, in case of walking pedestrians, the sequence of 3D models to be used in the correlation may suggest the pedestrian moving direction. Furthermore, improvements may be obtained by using an extended set of 3D models.

The system has been tested on a 1.8 GHz Athlon XP (FSB 266 MHz) with 512 MBytes DDR@400 MHz; the time required for the whole processing is 127 ms.

References

[1] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape-based pedestrian detection and localization," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, Shangai, China, Oct. 2003, pp. 328–333.

[2] D. M. Gavrila and J. Geibel, "Shape-Based Pedestrian Detection and Tracking," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

[3] H. Elzein, S. Lakshmanan, and P. Watta, "A Motion and Shape-Based Pedestrian Detection Algorithm," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 500–504.

[4] H. Nanda, C. Benabdelkedar, and L. Davis, "Modelling Pedestrian Shapes for Outlier Detection: a Neural Net based Approach," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 428–433.

[5] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking Pedestrian Recognition," *IEEE Trans. on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 155–163, Sept. 2000.

[6] L. Zhao and C. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Trans. on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 148–154, Sept. 2000.

[7] 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, no. 8, pp. 781–796, Aug. 2000.

[8] Z. Qiu, D. Yao, Y. Zhang, D. Ma, and X. Liu, "Detecting Pedestrian and Bicycle using Image Processing," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, Shangai, China, Oct. 2003, pp. 340–345.

[9] Y. L. Guilloux and J. Lonnoy, "PAROTO Project: The Benefit of Infrared Imagery for Obstacle Avoidance," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

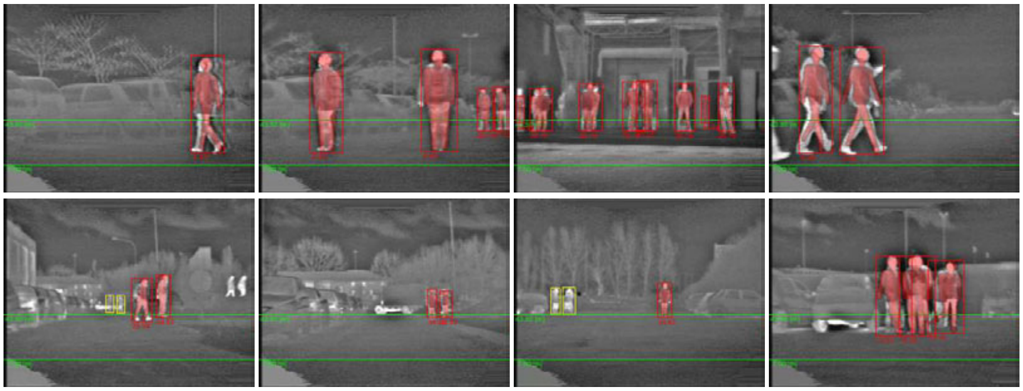

Fig. 11. Results of pedestrian detection in different situations: with complex or simple scenarios or with one or more pedestrians. The distance (in meters) is displayed below the boxes; the two horizontal green lines encode the range in which pedestrians are detectable.

[10] 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 *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 505–510.

[11] H. Nanda and L. Davis, "Probabilistic Template Based Pedestrian Detection in Infrared Videos," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

[12] M. Bertozzi, A. Broggi, T. Graf, P. Grisleri, and M. Meinecke, "Pedestrian Detection in Infrared Images," in *Procs. IEEE Intelligent Vehicles Symposium 2003*, Columbus, USA, June 2003, pp. 662–667.

[13] X. Liu and K. Fujimura, "Pedestrian Detection using Stereo Night Vision," in *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, Shangai, China, Oct. 2003, pp. 334–339.

[14] 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 in Computer Science*, vol. 2781, pp. 582–590, Sept. 2003.