Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

A Night Vision Module for the Detection of Distant Pedestrians

M. Bertozzi, A. Broggi, S. Ghidoni
VisLab – Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Parma, Italy
{bertozzi, broggi, ghidoni}@ce.unipr.it
www.vislab.it

M. M. Meinecke
Electronic Research
Volkswagen A.G.
Wolfsburg, D-38436, Germany
marc-michael.meinecke@volkswagen.de

***Abstract*— This paper presents a monocular night vision system specifically developed for detecting very distant pedestrians. The focus of the system is the recognition of pedestrians that are between 40 and 100 m away from the camera. The system is intended to integrate with an existing system, which is capable of detecting pedestrians at distances less than 40 m. At very large distances, pedestrians appear at low resolution, and this requires a specific detection algorithm, rather than an adaptation of an existing one. The presented system performs best in rural environments, where it can locate pedestrians at such great distances, that the pedestrians are hardly visible even to a human driver.**

I. Introduction

Pedestrian detection remains a key topic in the artificial vision field, because of its many applications. In particular, it can be exploited in the automotive field to protect the thousands of pedestrians being injured every year in road accidents. Many systems have been developed, based on different approaches, as described in [1]. The majority of pedestrian recognition vision systems uses far-infrared (FIR) imagery, because the heat emitted by the human body causes pedestrians to appear as bright objects at these frequencies. This facilitates the image segmentation process, and, above all, makes it possible to detect pedestrians in the absence of any natural or artificial lighting, and at great distances, even at night. However, FIR images are less useful during summer, because high ambient temperatures cause many objects in the environment to appear as bright as human bodies.

A. Related work

Some examples of night vision systems can be found in [2] and [3] (based on a stereo camera pair) and in [4], which is based on monocular vision. All these systems aim to detect all pedestrians in the image, therefore they deal with objects whose size can vary over a wide range. To cope with this, in [4] each bounding box found in the image is resized to a fixed size for the final validation. Moreover, in [5] it was observed that a multi-resolution approach can be useful for restricting the variability of pedestrians size in the image. This suggests that the best performance can be obtained if all possible pedestrians sizes are in a small range.

B. System goal

This paper presents a new algorithm capable of extending the detection range of the system described in [5] from 7-43.5 m to distances up to 100 m. Hence the focus is on pedestrians further than 40 m from the camera. The system works on low-resolution (320×240) 8-bit greyscale images, as it often happens in the FIR domain; therefore a pedestrian that is further than 40 m from the camera is sensibly smaller than those usually detected by other night vision systems, thus requiring a specific processing. Table I reports the values of a pedestrian's height in pixels at various distances for a 1.80 m tall person. In [6] the detection of very small pedestrians is considered. However, that system misses some pedestrians at long distances, since it is required to find also pedestrians near the camera.

TABLE I
Values of pedestrian heights in the image acquired by the system. The pedestrian is 1.80 m tall.

Distance from the camera (m)	Pedestrian height (pixels)
20	66
40	30
60	18
80	15
100	13

As previously mentioned, the algorithm does not consider pedestrians up to 40 m from the camera. Moreover, the task of the algorithm is really challenging, since it should detect even those pedestrians appearing as a small group of pixels. For this reason, it is intended to work in rural environments, where the number of warm objects is smaller than downtown. This is acceptable, because in a urban environment cars drive at relatively slow speeds, so that it suffices to detect pedestrians at 40 m to avoid collisions. On the other hand, in rural environments, cars travel much faster, and the absence of public lighting makes a pedestrian almost invisible until it is in the range of car lights. A system capable of detecting pedestrians at distances up to 100 m can therefore increase the probability of spotting a distant pedestrian in conditions of poor visibility (be it at night or in fog), and therefore also increase road safety.

The paper is organized as follows: the system structure, based on a two-step algorithm, is described in section II, while some recognition examples are reported in section III, and section IV discusses the overall system performance,

Fig. 1. Input image to the system. A distant pedestrian highlighted by the cyan oval appears in the scene.

together with some open issues that are still present.

II. System Structure

A typical scene containing a distant pedestrian that should be detected is shown in Fig. 1. The environment is suburban, and the pedestrian appears as a small hot spot, that is, a small group of bright pixels (80 in the example shown). Unfortunately, other small bright objects are also present in the scene, so a classification method capable of working on very small objects is needed.

The distant pedestrians detection algorithm can be divided into two main steps: the former locates regions of interest (ROI) in the image, while the latter validates them to select only those containing a pedestrian.

A. ROI Selection

For this application, a region of interest is an area containing an object that is warmer than the background. The whole image is first scanned to analyze contrast and brightness. Then, exploiting this data, a hot spot detector looks for the regions covered by the brightest pixels, or those with a lower grey-level, but that are connected to a bright region. This is done in order to expand each bright region to include all pixels that seem to belong to the same object. Such low-level segmentation is much more helpful than thresholding, since pedestrians do not present the same grey-level for the whole body, and some body portions would be discarded by a simple threshold.

For each selected region, a column-wise histogram is then computed to select only hot spots that have a significant vertical component, as it happens when they contain a pedestrian. This method is similar to that in [7], with some modifications to adapt it to detect even small objects. Because of the small size of the bounding boxes found so far, it is impossible to perform a check on symmetry properties.

At this level, a communication with the near pedestrian detector system is present. Since it runs prior to the distant pedestrian detector, it is possible to exclude all the regions in which a near pedestrian has already been found. This helps to avoid some false positives that appear on some parts of the human body, like closed hands, that in some cases present a shape similar to a head.

B. ROI Validation

A further analysis is needed to select, among all hot spots found so far, only those that contain a pedestrian; actually, this is the most challenging part. The ROI selection algorithm generates a great number of hot spots in each frame. Therefore validation should be efficient in discarding many of them in a quick manner, and then applying more complicated filterings only on the few remaining hot spots.

1) Position classification: The first quick check is about hot spots position. Recall that the goal of the system is to detect pedestrians in the range between 40 m and 100 m. Thanks to a precise camera calibration, needed also by the close pedestrians detection system that this algorithm integrates, and assuming the ground to be flat, it is possible to roughly know where distant pedestrians will appear, if present in the scene. Therefore, some regions are selected in the image, as it can be seen in Fig. 2: a pedestrian should be inside regions labeled as A and B, and its baseline should lay inside region B. Of course, when dealing with a camera fixed on a running vehicle, some oscillations and pitch variations have to be considered, therefore both regions A and B are a little larger than they should be, in order to gain some tolerance to calibration errors and camera movements[1].

All hot spots outside these areas are automatically discarded, as well as those that are inside A and B, but also overlap with regions C. This last observation is particularly useful because it helps filtering a lot of trees, poles, and tall symmetrical objects, that are one of the major issues for pedestrian detection systems. The need to verify if a hot spot expands also outside region A justifies the fact that the ROI selection is done in regions A, B, and C, and not only where hot spots may be found (that is, A and B only).

2) Size classification: Hot spots that are in a valid position are further investigated. Their size is considered, according to values in Table I, with some tolerance, but it is impossible to verify if the size is compatible with the pedestrian distance to the camera, because this last value is unavailable. In fact, monocular systems evaluate the distance of objects by only looking at where their baselines are placed, thanks to the knowledge of the geometry of the ground (that is almost always considered to be a plane). But in the low-resolution images obtained by a FIR camera placed on the lower part of the vehicle (on the bumper at 65 cm above the ground), the difference between the baseline of a pedestrian at 40 m and another one at 100 m is of few pixels only, so it is impossible to obtain a reliable measurement – and, if ever, the quantization would be unacceptable. This means that no strong checks can be made on the hot spots size.

After size check, the aspect ratio is analyzed: this can cause a group of pedestrians to be discarded, but it is unlikely

[1]The actual values used as size of areas A, B, C, and all other parameters, are kept confidential.

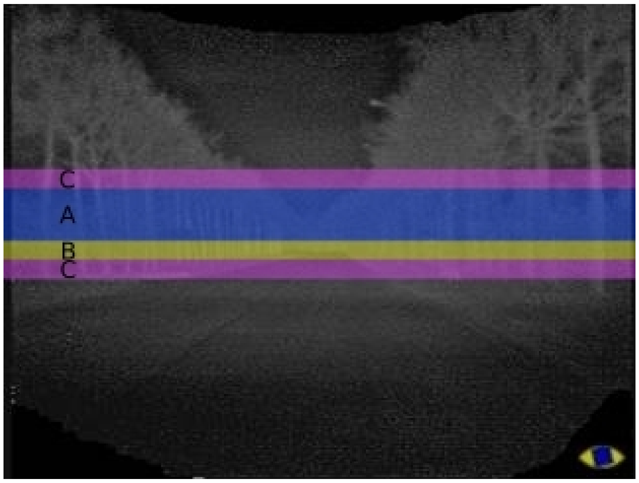


Fig. 2. Detection regions. A bounding box containig a distant pedestrian should be inside areas A and B. Bounding boxes overlapping also on C are discarded.

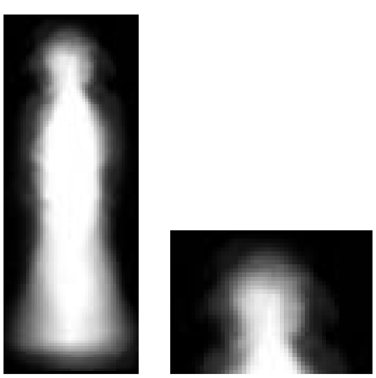
to happen that a large group of pedestrians walk on an rural road. This simple check can discard a large amount of hot spots that do not contain pedestrians.

3) Template matching: The last filtering is the most effective, but also the most computationally expensive one, therefore it is the last one to be applied. It is based on a probabilistic template model of a pedestrian, an approach that appears to be promising especially with very small pedestrians, like it is discussed in [8] for a FIR images-based system, and in [9] when dealing with visible imagery. During system development, both a match with a whole body template and a template of the head only (shown in Fig. 3) were considered: finally, the latter was chosen because it gave better results.

The match is performed using a simple correlation matching function. The correlation coefficient is obtained as:

$$\text{Match} = \frac{\sum_{i,j}\left(d_{\text{model}}(i,j)\cdot d_{\text{image}}(i,j)\right)}{\sqrt{\sigma^2_{\text{model}}\cdot\sigma^2_{\text{image}}}} \quad , \tag{1}$$

where $d_{\text{image}}(i,j)$ represents the difference between the image pixel at position (i,j) and the mean value of all pixels



(a) (b)

Fig. 3. Probabilistic model templates: (a) whole body, (b) head only.

of the image involved in the correlation process (this last value being μ_{image}); $d_{\text{model}}(i,j)$ is the same computation applied to the probabilistic model. The values of σ are:

$$\sigma^2_{\text{image}} = \sum_{i,j}\left(\text{image}(i,j)-\mu_{\text{image}}\right)^2 \quad , \tag{2}$$

$$\sigma^2_{\text{model}} = \sum_{i,j}\left(\text{model}(i,j)-\mu_{\text{model}}\right)^2 \quad . \tag{3}$$

From (1) it comes that the matching value may also become negative: this is the case, for example, of an image that is matched with its negative.

The probabilistic model of the head is first rescaled to match the hot spot width, and the match with its upper part is evaluated. Then, the model is rescaled and matched again to investigate if there is another size that provides a better matching, depending on the position of the pedestrian inside the bounding box. After the best match has been found, it is saved, and then compared to the other matching values obtained shifting the probabilistic head both horizontally and vertically. It was observed that if the hot spot is generated by a pedestrian, the matching value reaches a maximum in the initial position (or nearby), and then sensibly decreases when moving the model upwards, while it does not decrease too much if the model is moved downwards. This happens because, in the last case, the head is matched with the body, which turns out to be quite bright, even if not as much as the head. Since the same behavior is not observed for other obstacles, this technique based on multiple matches can be used to distinguish between pedestrians and other objects.

Match values computed when the head model is horizontally shifted are also considered. Indeed, such values should decrease, otherwise the analyzed area is too uniform, and is unlikely to contain a pedestrian. After all matchings are performed, a "pattern of matching values" is avaliable to the final validator, that chooses if the hot spot is a pedestrian or not. The decision is based on the absolute value of the best and worst matching regions, and on the pattern of correlation values.

To perform this processing, the correlation has to be computed several times, but the compared areas are so small, that the computational time does not sensibly increase.

C. ROI Merging

It can happen, especially during cold seasons, that a human body appears in a FIR image as a set of disjoint hot spots. Head, legs and feet are usually bright, while the trunk can appear quite dark. In this case, the low-level image segmentation in the ROI selection phase does not join together all the blobs of a pedestrian, because a relatively large and dark region is present in the middle. For this reason, after the ROI validation phase, discarded hot spots are considered again, and a merging process takes place.

When a pedestrian is split into different hot spots, it is assumed that at least the head and the legs appear, so the hot spots are vertically aligned, and close to each other. For these reasons, bright blobs are merged if they are nearby and if they are almost vertically aligned. After the merging, ROI validation is again performed on the new hot spots to verify if they contain pedestrians.

It was observed that the merging algorithm increases the correct detection rate in winter scenes, while not sensibly affecting the false detection rate.

III. Results

The described algorithm was tested in real traffic scenes both in urban and rural environments, even if it is intended to be applied only to the latter one. Tests were performed using a vehicle equipped with a FIR camera mounted over the bumper, see Fig. 4.

Some examples of the algorithm output are presented. In Fig. 6 (a) a scene with two distant pedestrians is shown: even if they are small (the height is 18 pixels), and a number of hot objects is present, they are both recognized (b). Hot spots containing a pedestrian are shown with a yellow rectangle with a red circle around, to ease visualization. In Fig. 7 (a) the system is working in a suburban environment, in (b) in a rural one: the capability of recognizing distant pedestrians is good, even if they appear really small in the image (20 pixels only). In Fig. 8 (a), a pedestrian appears with a dark trunk, and is therefore recognized as two separate hot spots, that are both discarded in the filtering phase. When the merging process is applied, however, the two hotspots are merged together (b), and the resulting box is not filtered away. In Fig. 9 (a), both a pedestrian and a cyclist are recognized; in (b) two spots can be seen: the one on the left is a pedestrian, while the other one is a vehicle headlight and tyre. This is actually the biggest issue of the system, because the shape of a light resembles a head, and other warm parts of the vehicle can get confused with the body.

Precise quantitative performance evaluation turns out to be difficult to obtain: in fact, in common systems it is done by comparing the system output with ground truth; however, also for a human operator it is extremely difficult to locate pedestrians that are really small, therefore the collection of ground truth itself was difficult. In general, a good behavior is observed in suburban and rural roads. Statistics are based on 2938 frames, taken in different conditions: the correct detection rate is computed as CD/(CD+FP), where CD is the number of pedestrians that are correctly detected, and FP is the number of objects that have been classified as pedestrians, even if they are not; these values have been computed frame by frame. The correct detection rate is 69.2 %, while only 0.036 false positives per frame are found; moreover, false positives are not persistent; however, they sensibly increase in images taken in complicated scenarios inside a city. Interesting results were also obtained with the hot spot merging, that allows to reliably detect also bicyclists.

The major issue of the system is the misrecognition of vehicles headlights, that are often merged with warm tyres and the surrounding area, causing most of the false positives. This is due to the fact that the headlights shape, at a great distance, is circular, so that the matching with the probabilistic template produces a correlation pattern similar to that obtained when a human body is matched. Other false positives appear on buildings, but are rather rare, thanks to the correlation pattern method.

The computational load of the algorithm is important, since, as already said, this system should be integrated in a more complex one without decrementing too much the processing rate. Average execution times are about 3 ms on a Pentium 4 machine working at 3 GHz.

IV. Conclusions and Open Issues

The most critical problem affecting this system is due to the dependency on calibration. For example, when driving on a speed bumper, the camera orientation is subject to a temporary but heavy change, as can be seen in Fig. 5. In these conditions the system performance is limited, even when some tolerance is considered in defining all system parameters. Such tolerance lets anyway the system work properly when the camera is subject to the common oscillations of a running car. The calibration problem could be substantially solved employing a stabilization algorithm, so that the areas of interest of the algorithm are always in a correct position.

Fig. 4. Vehicle equipped with a FIR camera used for testing the algorithm.

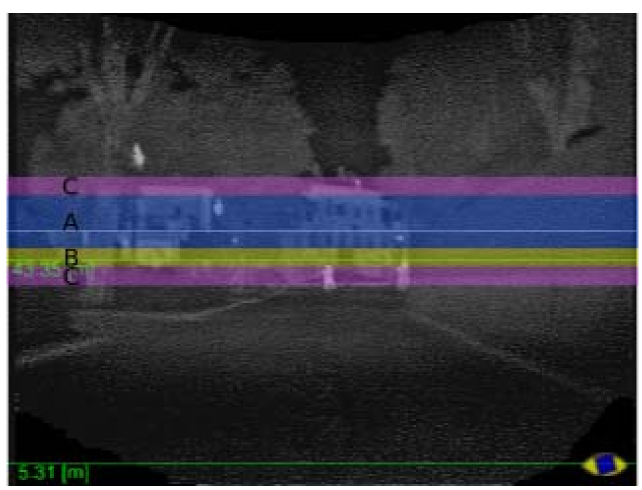
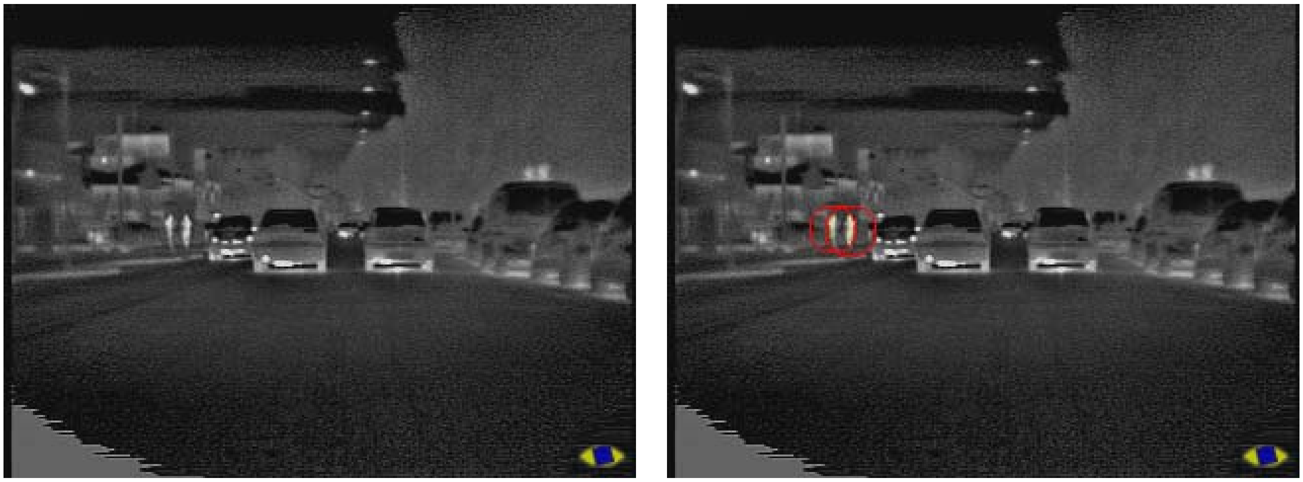
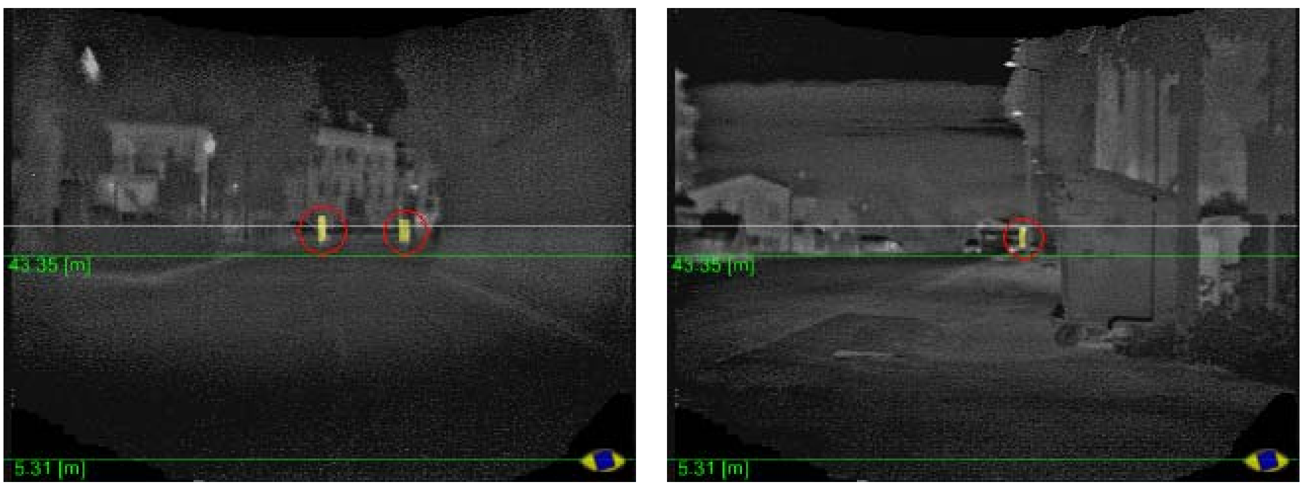


Fig. 5. Image acquired when driving on a speed bumper. The heavy calibration error causes the pedestrian to appear inside region C, so that it cannot be recognized.



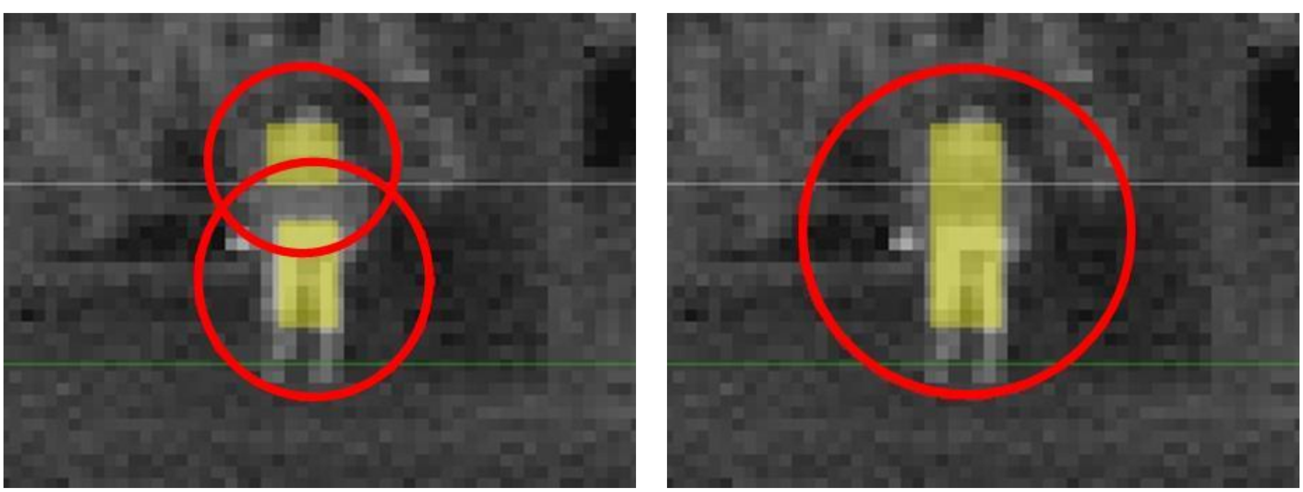
(a) (b)

Fig. 6. An example of algorithm output: in the original image (a) two distant pedestrians are present; they are both recognized by the algorithm (b).



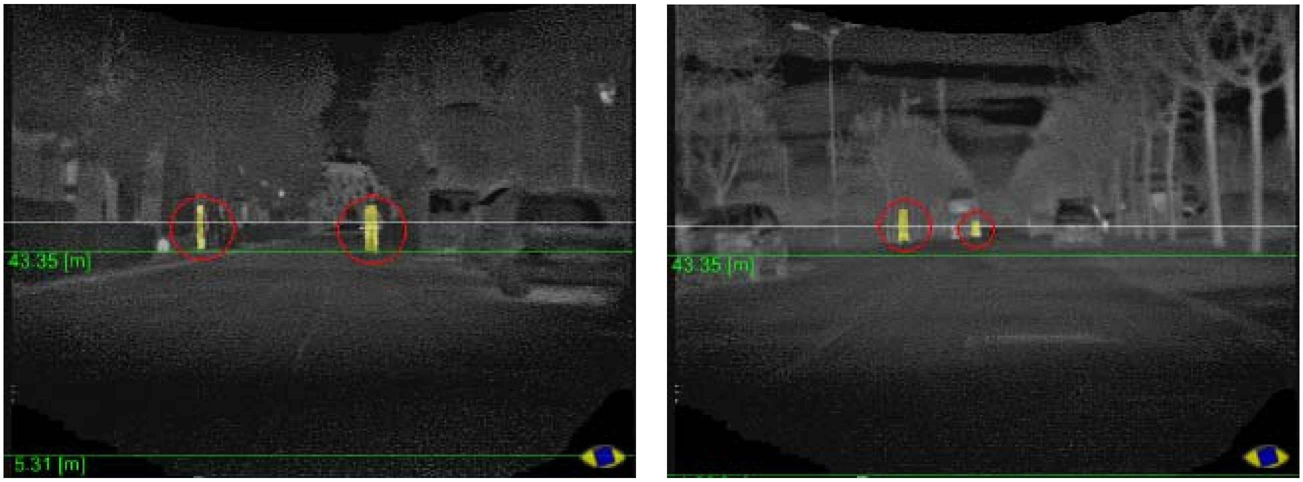
(a) (b)

Fig. 7. Recognition in suburban (a) and rural (b) environments.



(a) (b)

Fig. 8. In (a), a pedestrian is recognized as two separate hot spots, due to the cold trunk; the merging phase can solve this problem by creating a single, larger hot spot (b) that can be recognized as a pedestrian in the following validation step.



(a) (b)

Fig. 9. In (a) a cyclist is recognized as a pedestrian, together with a standing person. In (b) an example of misdetection: the hot spot on the right is a vehicle light.

Even if the results obtained so far are very promising, some system improvements can be considered. For example, by adopting a stereo camera pair and adding a stereo match processing, the distance of detected pedestrians could be precisely measured. These new data could also help in developing new filtering techniques capable of eliminating false positives, especially downtown.

Fusion with other sensors turns out to be quite difficult, since wide-angle radars commonly installed on cars can hardly find a pedestrian (that is a weak reflector) beyond 40-50 m. A laserscanner may offer a better performance, but at such distances its resolution is limited; furthermore, the commonly used single-plane laserscanners, when used for long distance applications, suffer from the same problem of inaccuracy produced by vehicle pitch. For this task, vision – although not 100 % reliable – seems to be the best technology, and probably stereovision would provide the best results.

REFERENCES

[1] T. Gandhi and M. M. Trivedi, "Pedestrian Collision Avoidance Systems: a Survey of Computer Vision based Recent Studies," in *Procs. IEEE Intelligent Transportation Systems 2006*, Sept. 2006, pp. 976–981.

[2] T. Tsuji, H. Hattori, M. Watanabe, and N. Nagaoka, "Development of Night-vision System ," *IEEE Trans. on Intelligent Transportation Systems*, vol. 3, no. 3, pp. 203–209, Sept. 2002.

[3] X. Liu and K. Fujimura, "Pedestrian Detection using Stereo Night Vision," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 6, pp. 1657–1665, Nov. 2004, ISSN 0018-9545.

[4] F. Xu and K. Fujimura, "Pedestrian Detection and Tracking with Night Vision," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

[5] A. Broggi, A. Fascioli, M. Carletti, T. Graf, and M.-M. Meinecke, "A Multi-resolution Approach for Infrared Vision-based Pedestrian Detection," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 7–12.

[6] U. Meis, M. Oberländer, and W. Ritter, "Reinforcing the Reliability of Pedestrian Detection in Far-Infrared Sensing," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 779–783.

[7] M. Bertozzi, A. Broggi, M. Del Rose, and A. Lasagni, "Infrared Stereo Vision-based Human Shape Detection," in *Procs. IEEE Intelligent Vehicles Symposium 2005*, Las Vegas, USA, June 2005, pp. 23–28.

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

[9] M. Hu, "Multiple Probabilistic Templates Based Pedestrian Detection in Night Driving with a Normal Camera," in *Procs. IEEE Intl. Conf. on Innovative Computing, Information and Control 2006*, vol. 2, Beijing, China, Aug. 2006, pp. 574–577.