Infrared Stereo Vision-based Pedestrian Detection

M. Bertozzi, A. Broggi, and A. Lasagni
Dipartimento di Ingegneria dell'Informazione
Università di Parma
Parma, I-43100, Italy
{bertozzi,broggi,lasagni}@ce.unipr.it

M. Del Rose
Vetronics Research Center
U.S.Army TARDEC
Warren, MI, U.S.A.
DelRoseM@tacom.army.mil

***Abstract*— This paper describes an approach for pedestrian detection in stereo infrared images. The developed system has been implemented on an experimental vehicle equipped with two infrared camera and preliminarily tested in different situations. It is based on the localization and distance estimation of warm areas in the scene; the algorithm groups areas with similar position and considers only results with specific size and aspect ratio. A final validation process, based on the head shape's morphological and thermal characteristics, is used to build the list of potential pedestrian appearing in the scene. Neither temporal correlation, nor motion cues are used in this processing.**

I. Introduction

The human shape detection from a moving vehicle is the target of a cooperation between the Vetronics Technology Area, a division of the Army's TACOM Research, Development, and Engineering Center (TARDEC), and the University of Parma.

U. S. Army is working on programs (Robotic Follow program [1], Crew integration and Automation Test bed program [2]) that are in need of autonomous and semi-autonomous navigation. Human shapes detection is a vital piece to make these programs succeed.

However, the need for automatic human detection goes beyond the Vetronics Technology Area. Within the Army, and throughout the commercial community, the need for detecting people is great. Driver warning systems, security systems, traffic/pedestrian control systems, and automatic switching systems are just a few areas that could greatly benefit from this technology.

Vision-based pedestrian detection is a difficult task: pedestrians usually wear different and differently colored clothes and often are barely distinguishable from the background. Moreover, pedestrians can wear or carry items like hats, bags, umbrellas, and many others, which give a broad variability to their shape. Additional problems that must be considered are: noise produced by the presence of buildings and human artifacts, moving or parked cars, cycles, road signs, signals, different illumination conditions, obstacles and so on.

There are many approaches to pedestrian detection. Some use learning machines like neural networks [3] or support vector machines [4], some are based on the detection of specific patterns, texture or motion clues [5], [6], others are stereo vision based [7].

Only recently, thanks to the decreasing cost of infrared (*IR*) cameras, different systems based on the processing of far-infrared images have been presented [8]–[12].

This work presents an approach based on stereo infrared images to define areas that could contain pedestrians. Warm parts of the scene are detected and, using stereo informations, their position and distance to the cameras are estimated. The algorithm groups detected objects with similar coordinates, creating a list of hot areas. These results are then filtered and only areas with specific size and aspect-ratio are considered and analyzed to find head morphological characteristics. Although the proposed method works on single frames and does not perform tracking, preliminary results have proven to be promising.

This paper is organized as follows: section II introduces all parts of the algorithm and section III presents the results of this approach and the performances of the system. Section IV summarizes and concludes the paper.

II. Algorithm description

The algorithm is composed by seven steps: (1) preprocessing of the right input image; (2) warm areas detection in the right image, yielding rectangular bounding boxes framing interesting areas; (3) localization of homologous bounding boxes in the left image; (4) objects distance and position estimation; (5) grouping of bounding boxes in similar position; (6) filtering of the resulting areas; (7) analysis of head's characteristics.

The first six phases of the algorithm are oriented to detect in which areas of the scene pedestrians could be located. The last phase tries to find, inside the results, the most evident characteristic of a human shape: the head.

A. Preprocessing phase

The first phase of the algorithm focuses the attention on areas of the input image (figure 1.a) with a high intensity value, that represent warm objects. This is obtained using two different threshold values: initially, a high threshold is applied on the pixel values in order to get rid of cold or barely warm areas, selecting only pixels corresponding to very warm objects. Then, pixels featuring a grey level higher than a lower threshold are selected if they are contiguous to other already selected pixels in a region-growing fashion. The resulting image contains only warm contiguous areas that present hot spots (figure 1.b).

B. Warm elements detection

In order to select vertical stripes containing hot regions, a column-wise histogram is computed on the resulting image.

Fig. 1. Preprocessing phase: (*a*) original input image, (*b*) focus of attention.

The histogram is filtered with an adaptive threshold whose value is a fraction of the average value of the whole histogram. Obviously, multiple hot objects may be vertically aligned in the image, so that their contributions add up in the histogram. Nevertheless, warm areas belonging to the same horizontal stripe can be distinguished by computing a new row-wise histogram of the grey-levels for each stripe. This procedure yields rectangular bounding boxes framing areas where pedestrians may be located (see figure 2.a). In order to refine these bounding boxes, the column-wise and row-wise histogram procedure is iteratively applied to each rectangular box (figure 2.b) until its size is no longer reduced. Results are shown in figure 2.c.

In this phase, small bounding boxes are thrown out, as they represent nuisance elements.

C. Candidate areas localization

The content of the resulting bounding boxes is matched with the other stereo image with the purpose of finding the corresponding areas of the image. By supposing that the optical axes of the cameras are parallel, homologous areas can be localized on the same row in the two images and limited search space can be estimated thanks to the knowledge of calibration parameters.

The following Pearson's correlation function is used to evaluate the result of the match:

$$r = \frac{\sum a_{xy} b_{xy} - \frac{\sum a_{xy} \sum b_{xy}}{N}}{\sqrt{(\sum a_{xy}{}^2 - \frac{(\sum a_{xy})^2}{N})(\sum b_{xy}{}^2 - \frac{(\sum b_{xy})^2}{N})}} \quad (1)$$

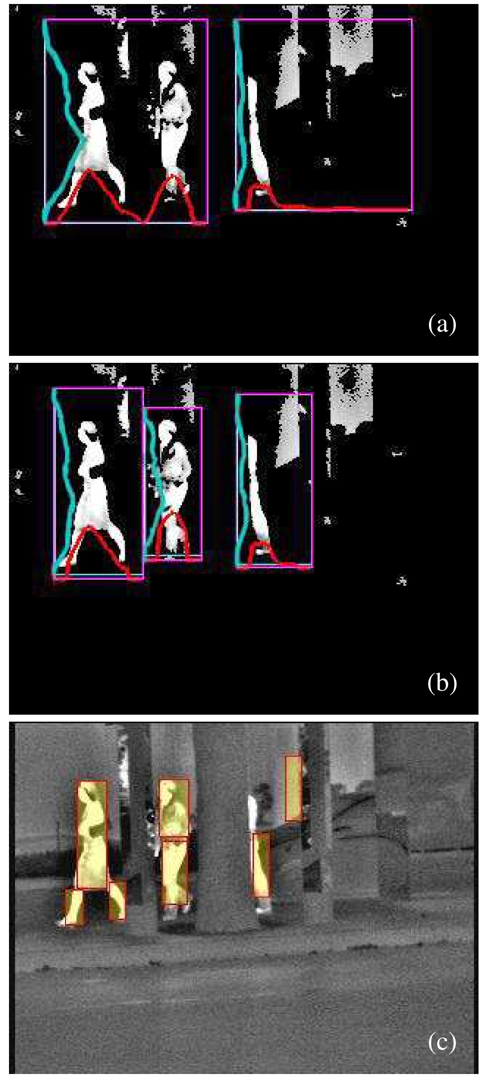


Fig. 2. Warm elements detection: (*a*) intersection of column-wise and row-wise histograms produces wide rectangular bounding boxes, (*b*) a following iteration reduces their size, (*c*) final results are shown on the original image.

where N is the number of pixels in the considered bounding box, a_{xy} and b_{xy} are gre-level values of corresponding pixels of the two images. The bounding box in the other image featuring the maximum value for the correlation is selected as the best match, and the algorithm considers only the best matches whose correlation values are higher than a given threshold (figure 3.a).

A triangulation technique is then used to estimate the distance between each object and the vision system. Therefore, a refinement of the bounding boxes base can take place, based on calibration and perspective constraints. More precisely, the knowledge of position and orientation of the cameras and assumption of a flat road-plane can provide information about human shape's point of contact with the ground. This knowledge is used to stretch the bottom of the bounding box till it reaches the ground and frames the entire shape of the pedestrian.

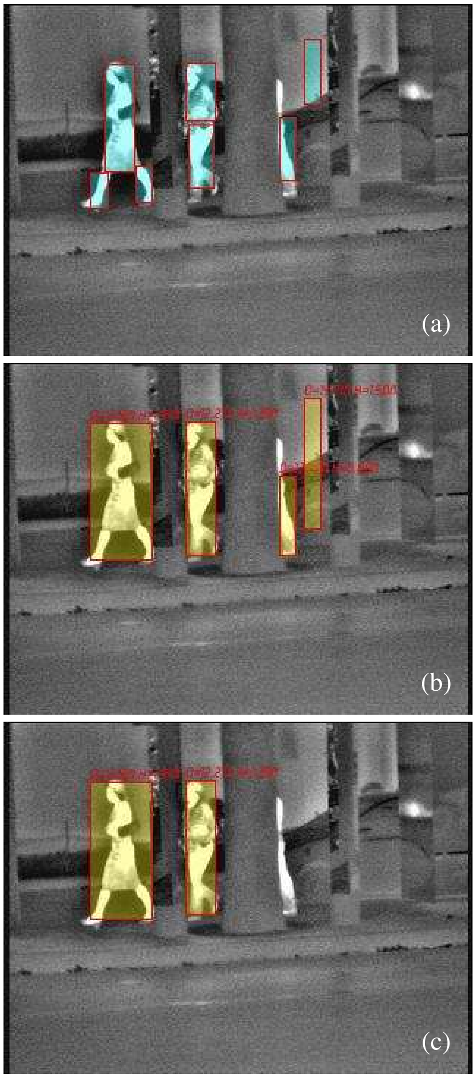


Fig. 3. (*a*) Homologous bounding boxes in the left image, (*b*) bounding box bases are resized to reach the ground and elements with similar 3*D* coordinates are grouped together, (*c*) areas compatible with the presence of pedestrians.

Then, bounding boxes located near to each other are grouped in a single bounding box, that represents an area of interest. A list of warm areas of the scene containing potential pedestrians is then built (see figure 3.b).

Human shape is not always warmer than the background. Thus, the system should manage to also detect incomplete human shapes. More specifically, experimental results showed that the bounding box is sometimes smaller than the human shape when hair or a hat mask the head heat. To cope with this problem the following approach was tested: the top of the boxes is extended until the correlation between the additional areas in the left and right correspondent portions of the stereo images is higher than a given threshold. The results seem promising but an additional processing that takes into account the image texture should be further investigated.

In order to get rid of false positives a number of filters relying on size and aspect ratio have been devised. Small or huge bounding boxes are discarded, because they can not contains human shapes; moreover, it is expected to find standing pedestrian, so also bounding boxes with a significant horizontal component are eliminated. Distance and height of each object are evaluated as well: too close bounding boxes that can not contain whole human shape, too far areas that do not allow to analyze their content with sufficient reliability, or too short or tall objects are discarded.

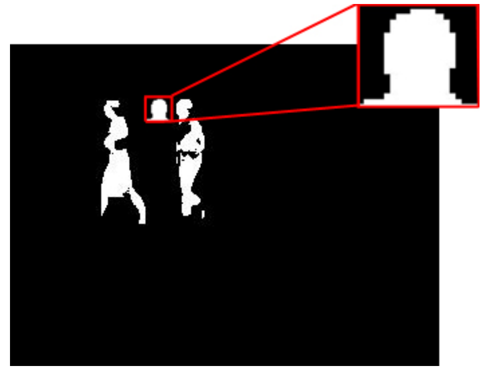


Fig. 4. Candidates areas are binarized and the model is scaled in order to obtain best pattern matching results.

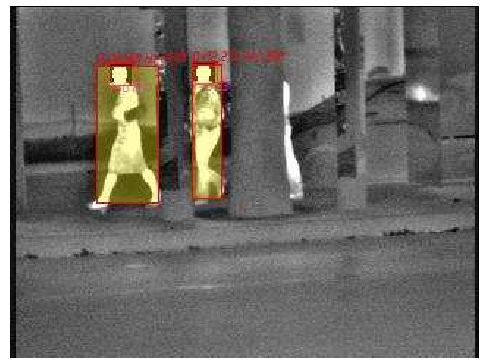


Fig. 5. Pedestrians heads detection.

After the filtering phase, a list of areas that can potentially contain pedestrians is obtained (see figure 3.c).

D. Search for pedestrians head

The last phase of the algorithm is in charge to find, in each area of interest, characteristics which could be related to the presence of a pedestrian. Generally, the most evident feature of a human shape in infrared images is represented by the head. This is due to the fact that head is often warmer than the body and its position is not strongly affected by pedestrian's pose, being always in the upper part of the bounding box. A simple binary model of a head is used to perform pattern matching operations and, in order to produce better results, the considered areas are also binarized and the model is scaled to adapt it to the bounding boxes size, assuming that a head measures nearly $1/6$ of human shape height (figure 4). A match for the head is performed against an area centered around the top of the bounding box.

Different rows of each bounding box are considered to avoid possible mistakes in their dimensions, as the head could not be

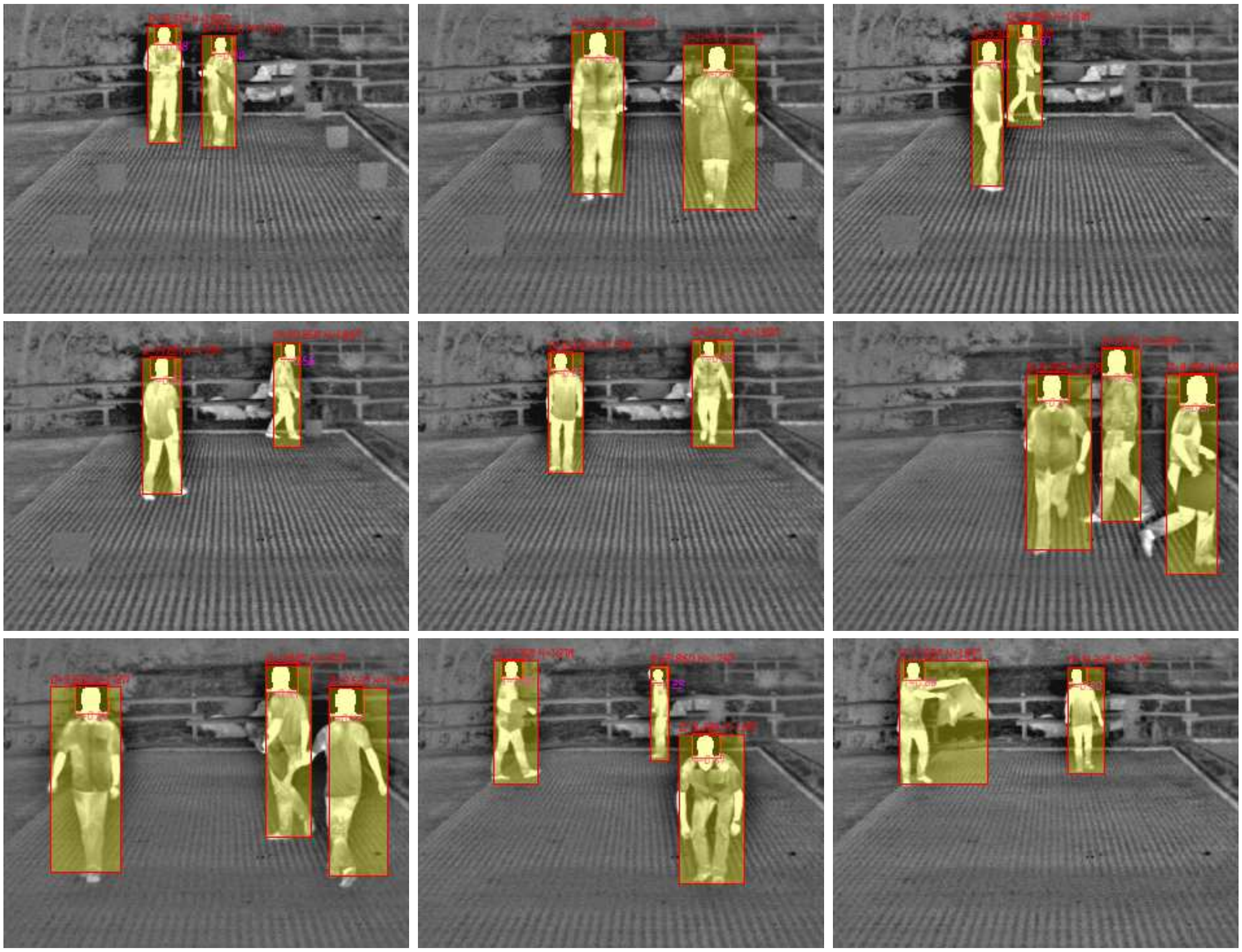


Fig. 6. Pedestrian detection results.

exactly contained in them. To evaluate the result of the match, correlation (1) is used.

The portion of the bounding box that produces the best correlation value is recognized as the head of a potential pedestrian, see figure 5. Bounding boxes featuring a too bad match are discarded as not containing a human shape.

III. Results

The developed system has been preliminarily tested in different situations using an experimental vehicle equipped with two infrared cameras.

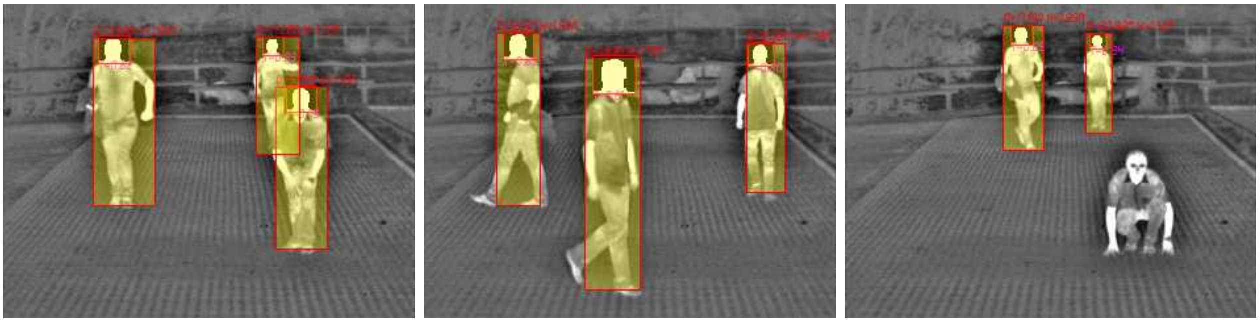
Figure 6 shows some results of this infrared stereo vision-based pedestrian detection approach. Each human shape is remarked by a superimposed yellow box; distance and height are reported as well. It is possible to see that the system is able to detect one or more pedestrians even if they are close to each other or in presence of a complex background. It can be noticed in figure 7 that, thanks to the triangulation information, different human shapes are not mixed even if partly occluded.

Moreover, even though the model used for head detection is quite simple, the head of each pedestrian is properly located. The model represents a frontal head shape but it give good results also for side shots (see figure 7). In figure 7.c pedestrian in foreground is not detected, even if visible, because his height is too small for a human shape and the corresponding bounding box has no valid aspect-ratio, so the filtering phase discards it.

Best results are obtained at night or low temperature conditions: in this cases the images present a high contrast between cold objects, whose pixels are dark, and warm objects, whose pixels are bright, and it is easy to discriminate between them. Otherwise, in high temperature conditions, the system seems not to be suitable to give satisfying results, as there is not enough difference between cold and warm elements.

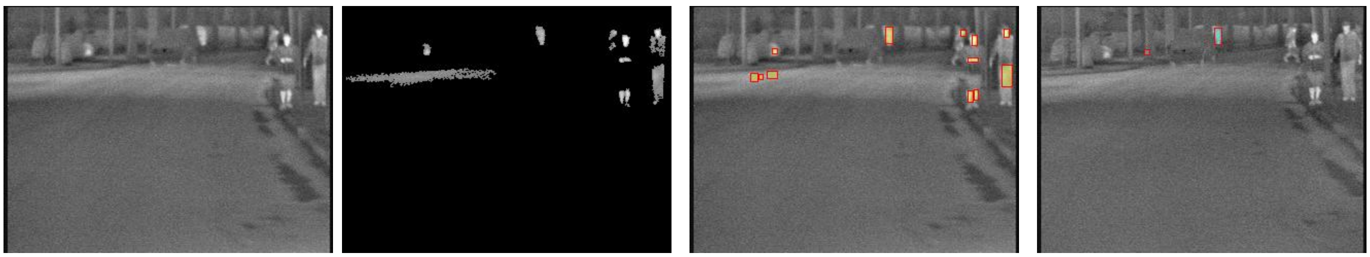
Main problems to check concern aspect ratio and small bounding boxes. Sometimes aspect ratio is not a good evaluation criterion for filtering results. Figure 8 shows that a pedestrian with open arms produces a bounding box that is compatible with other possible objects in the scene, like cars. This problem can appear in different situations, causing the system to get false positive results despite of the head pattern matching filtering.

Another problem concerns small warm elements: the algorithm discards warm objects framed by small bounding boxes because they usually contain noise. This can sometimes cause



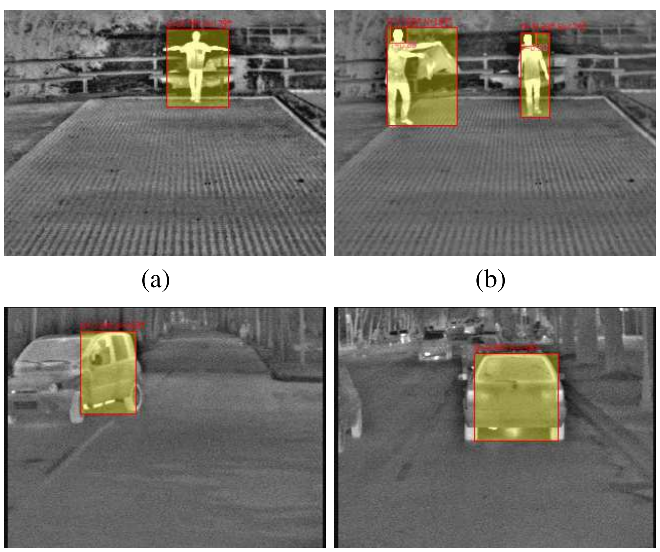
(a) (b) (c)

Fig. 7. Results: (*a*) the system is able to detect pedestrians even if partially occluded, (*b*) the head model is appropriate whether for frontal or side shots, and (*c*) foreground pedestrian foreground is not properly detected due to height and aspect ratio constraints.



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

Fig. 9. Small warm elements problem: (*a*) Right input image, (*b*) preprocessed image, (*c*) warm elements detected, (*d*) homologous bounding boxes localized in the left image. Detected hot objects in right image are too small to produce good results during the homologous localization phase.



(a) (b)

(c) (d)

Fig. 8. Aspect ratio problems: (*a*) and (*b*) show pedestrians detected assuming unusual positions; their bounding boxes have an aspect ratio which is compatible with other possible objects in different situations, like in (*c*) and (*d*).

important informations loss and false negatives. However, removing this filter is not a good solution, as shown in figure 9: under some circumstances, small warm elements in the preprocessed image are considered, but these boxes don't produce appreciable results in homologous localization, so that no pedestrians can be detected.

Other detection failures are due to occlusions, but this problem is observed only in few frames, thus tracking could be used to cope with these particular cases.

A tool for vision based pedestrian detection performance evaluation [13] has been used to obtain ROC curves shown in figure 10. The first curve has been obtained running the system with the head detection disabled. It can be noticed that in this condition the system is able to correctly detect more than 80% of pedestrians in the scene. Lowering stereo correlation threshold do not improve this percentage. The second curve has been obtained enabling the head detection and varying the correlation threshold in the pattern matching. Correct detection percentage rapidly reaches the maximum value with a very low number of false detection per frame.

Tests have been performed using a Pentium IV processor at 2.80 GHz equipped PC with 512 kBytes of cache memory and 1 GByte of RAM.

Several sequences of images have been acquired from a moving vehicle and have been processed in order to evaluate the algorithm execution time. For each frame of a sequence, elaboration time is strongly related with the number of pedestrians it contains. Therefore, the sequences have been subdivided into four categories:

Results are shown in table I. It is possible to see that each single frame is processed in an average time of 35.25 ms. This means that the system is able to elaborate about 29 frames per second and this is a good result, considering that cameras works at a slower frame rate (25 fps). Therefore, algorithm

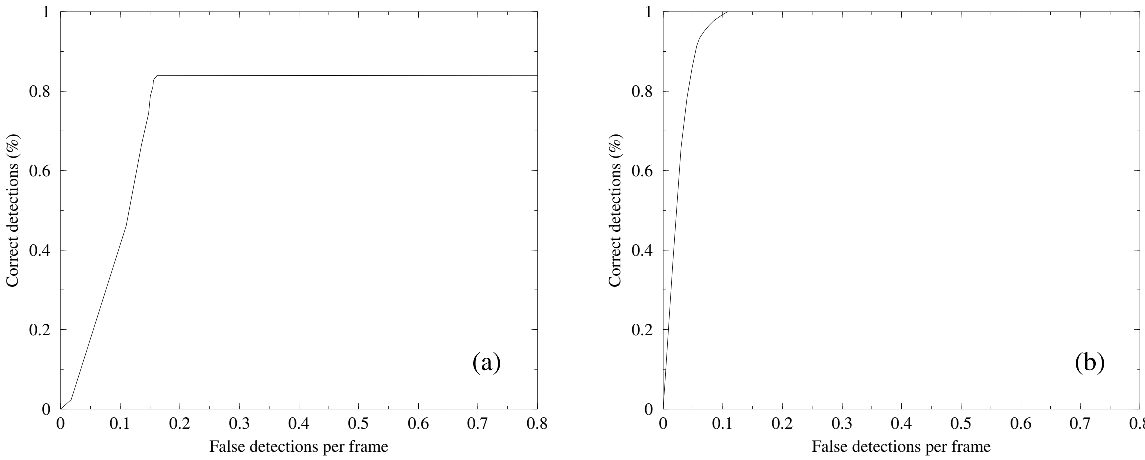


Fig. 10. ROC curves: (*a*) stereo matching and (*b*) head matching.

speed allow to add some more processing step in order to obtain a better pedestrian detection.

Sequence	t. min (*ms*)	t. max (*ms*)	**t. avg** (*ms*)	***fps***
Extraurban	16	129	**32**	**31.25**
Urban few pedestrians	14	87	**38**	**26.31**
Urban, more pedestrians	12	128	**41**	**24.39**
Urban, crowded	12	100	**30**	**33.33**
Average	**13.5**	**102.75**	**35.25**	**28.82**

TABLE I

PERFORMANCE OF THE SYSTEM IN DIFFERENT CONDITIONS.

IV. CONCLUSIONS

In this paper a stereo vision-based algorithm aimed to the detection of pedestrians in infrared images has been presented. It has been tested in urban and extraurban environments using an experimental vehicle equipped with two infrared cameras.

The algorithm is based on the detection of warm areas in the scene that are compatibles with human presence, based on distance estimation, size, aspect ratio, and head shape localization. The detection of pedestrians is computed through the processing of single frames (temporal correlation is not considered).

Experimental results are good and demonstrate that this approach is promising, especially in extraurban and low temperature conditions. Correct detection percentage is high with a very low number of false detections per frame, and the system has proven to work also when pedestrians are partly occluded.

In urban situations, noise produced by the presence of buildings, cars, signals and other objects could increase false detections. Therefore, in order to enhance the robustness and reliability of the discussed system, alternative warm objects search methods, filters and a tracking algorithm are currently under development.

ACKNOWLEDGMENT

This work was supported by the European Research Office of the U. S. Army under contract number N62558-03-M-0041.

REFERENCES

[1] B. Brendle and J. Jaczkowski, “Robotic Follower: Near-Term Autonomy for Future Combat Systems,” in *Procs. SPIE - AeroSense Conference 2002*, vol. 4715, Orlando, FL, Apr. 2002.

[2] B. Brendle, “Cockpit Development in the Crew Integration and Automation Testbed advanced Technology Development Program,” in *Procs. SPIE - AeroSense Conference 2003*, vol. 5080, Orlando, FL, Apr. 2003.

[3] 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.

[4] S. Kang, H. Byun, and S.-W. Lee, “Real-Time Pedestrian Detection Using Support Vector Machines,” *Lecture Notes in Computer Science*, vol. 2388, p. 268, Feb. 2002.

[5] V. Philomin, R. Duraiswami, and L. Davis, “Pedestrian Tracking from a Moving Vehicle,” in *Procs. IEEE Intelligent Vehicles Symposium 2000*, Detroit, USA, Oct. 2000, pp. 350–355.

[6] Y. Song, X. Feng, and P. Perona, “Towards detection of humans,” in *Procs. Conf. on Computer Vision and Pattern Recognition*, vol. 1, South Carolina, USA, June 2000, pp. 810–817.

[7] K. Fujimoto, H. Muro, N. Shimomura, T. Oki, Y. K. K. Maeda, and M. Hagino, “A Study on Pedestrian Detection Technology using Stereo Images,” *JSAE Review*, vol. 23, no. 3, pp. 383–385, Aug. 2002.

[8] 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.

[9] 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.

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

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

[12] A. Broggi, M. Bertozzi, , R. Chapuis, F. C. A. Fascioli, and A. Tibaldi, “Pedestrian Localization and Tracking System with Kalman Filtering,” in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 584–589.

[13] M. Bertozzi, A. Broggi, P. Grisleri, A. Tibaldi, and M. D. Rose, “A Tool for Vision based Pedestrian Detection Performance Evaluation,” in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 784–789.