Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 106 (2007) 194–204
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Pedestrian detection by means of far-infrared stereo vision

M. Bertozzi [a,*], A. Broggi [a], C. Caraffi [a], M. Del Rose [b], M. Felisa [a], G. Vezzoni [a]

[a] *Dip. Ing. Informazione, Parco Area Delle Scienze 181A, I-43100 Parma, Italy*
[b] *US Army TARDEC, Warren, MI, USA*

Received 22 November 2005; accepted 24 July 2006
Available online 31 December 2006
Communicated by James Davis and Riad Hammoud

Abstract

This article presents a stereo system for the detection of pedestrians using far-infrared cameras. Since pedestrian detection in far-infrared images can be difficult in some environmental conditions, the system exploits three different detection approaches: warm area detection, edge-based detection, and disparity computation. A final validation process is performed using head morphological and thermal characteristics. Currently, neither temporal correlation, nor motion cues are used in this processing.

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

Keywords: Infrared; Stereo; Pedestrian detection

1. Introduction

Pedestrian detection is an important field of research for both commercial and government organizations. Autonomous and semi-autonomous vehicles need to identify people while traversing through the terrain in order to take appropriate actions to avoid them. Driver awareness systems need the ability to alert drivers of potential paths of pedestrians when driving through urban areas. Additionally, the US Army requires pedestrian detection for path following, mule operations, and intent based anti-tamper surveillance systems to protect its robotic vehicles [1–3].

Unfortunately pedestrian detection is a challenging task. Thus, there have been many different approaches to solve this problem. Some use LADAR or laser scanners to retrieve a 3D map of the terrain and detect pedestrians [4–7], another uses ultrasonic sensors to determine the reflection of pedestrians [8]. Radar is also popular for detecting pedestrians similar to ultrasonic sensors; by measuring the reflection of possible targets and determining if they are pedestrians or not [9,10]. A natural choice for a pedestrian detection sensor is vision because it is based on how people perceive humans (through visual cues). Within this problem set there are monocular vision systems [11–13], stereo vision systems [14–20], and infrared vision [21–25] systems. Each vision system has its own advantages and disadvantages.

For the US Army, specifically the Tank Automotive Research, Development and Engineering Center (TARDEC), vision based detection is important for non-evasive pedestrian detection systems in the areas of path following, mule operations, surveillance, and driver awareness. The problem increases in difficulty when considering the movement of the sensors, uncontrolled outdoor environments, and variations in pedestrian's appearance and pose.

The following presents a pedestrian detection system based on stereo far-infrared images. The system exploits three different approaches for detecting objects: warm areas detection, vertical edge detection, and disparity approach. The algorithm combines this information to

* Corresponding author. Fax: +390521905723.
E-mail addresses: bertozzi@ce.unipr.it (M. Bertozzi), broggi@ce.unipr.it (A. Broggi), caraffi@ce.unipr.it (C. Caraffi), mike.delrose@us.army.mil (M. Del Rose), vezzoni@ce.unipr.it (G. Vezzoni).
URL: http://vislab.unipr.it (M. Bertozzi).

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2006.07.016

create a list of relevant pedestrian candidates. The list is then filtered by size/shape ratios of humans and head detection. An active contours approach is currently under testing for shape refinement.

This paper is organized as follows: Section 2 introduces all parts of the algorithm and Section 3 presents the results of this approach and the performance of the system. Section 4 summarizes and concludes the paper.

2. Algorithm description

In the far-infrared domain the image of an object depends on the amount of heat the object emits namely on its temperature. Generally, the temperature of a pedestrian is higher than the environment and a person's heat radiation is sufficiently high compared to the background. Such human shapes appear brighter than the background in infrared images easing the detection process. In fact, a previously developed approach [25] exploits this feature to detect pedestrians.

Unfortunately, pedestrians are not always brighter than the background. During the summer or in a sunny scenario pedestrians are often darker than the surrounding environment. In addition, clothing may mask heat radiation therefore leading systems that detect heat only to partially or completely misdetect pedestrians.

In order to cope with this problem, three different underlying approaches have been developed for pedestrian detection: warm areas detection, vertical edges detection, and a disparity space image-based approach. The first approach is devoted to detect warm areas, while the other two are in charged of detecting also cold objects that potentially can be pedestrians. All these approaches build a list of rectangular bounding boxes framing interesting areas.

A following processing is in charge of localizing the homologous bounding boxes in the other image, thus allowing an estimation of bounding boxes distance and position.

Bounding boxes featuring the same distance and a similar position are then grouped together in order to build larger bounding boxes. The result is filtered using constraints about minimum and maximum allowed pedestrian size and aspect ratio.

A further match is used to search for the most evident human shape characteristic in the FIR domain, namely the head. The overall system flow is depicted in Fig. 1.

2.1. Underlying approaches

In the following, three approaches for the detection are described: areas of attention, stereo match, and filterings.

2.1.1. Warm areas detection

The first phase of the algorithm focuses the attention on areas of the input image (Fig. 2a) 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 (Fig. 2b).

In order to select vertical stripes containing hot regions, a column-wise histogram is computed on the resulting image (see Fig. 3a). 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 (Fig. 3b). This procedure yields rectangular bounding boxes framing areas where pedestrians may be located. In order to refine these bounding boxes, the column-wise and row-wise histogram procedure is iteratively applied to each rectangular box until its size is no longer reduced. Results are shown in Fig. 3c.

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

Unfortunately, this approach fails in detecting the correct areas of interest when pedestrians are not warmer than the background or when too many warm objects are present in the scene (like in the urban environment). Therefore, two additional approaches not based on thermal features, are used to produce other areas of interest.

2.1.2. Edges detection

This approach is based on the assumption that human shape features more vertical edges than the background

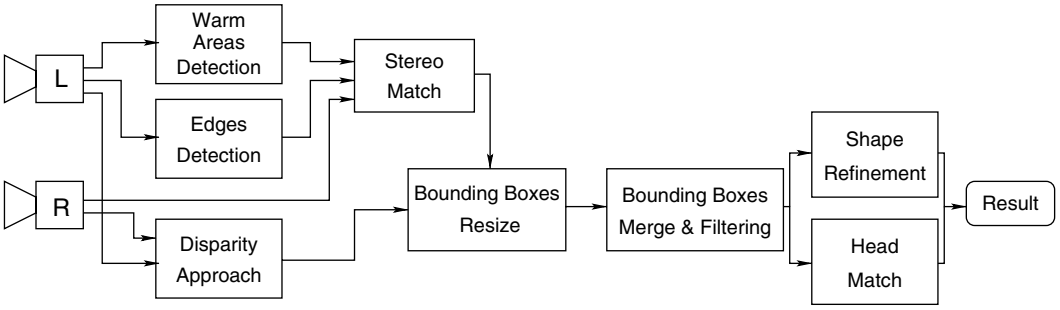

Fig. 1. Overall algorithm flow.

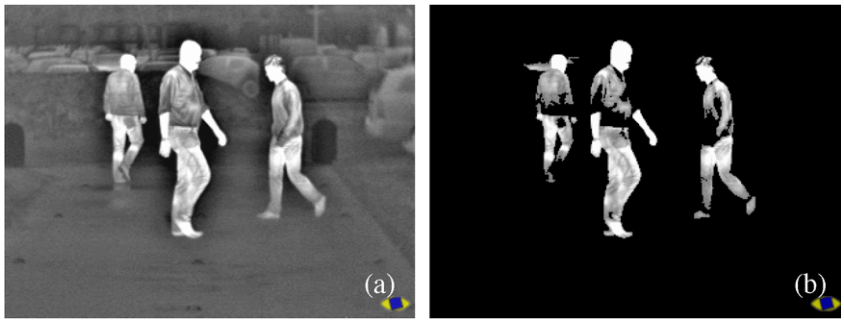

Fig. 2. Preprocessing phase: (a) original input image, (b) focus of attention.

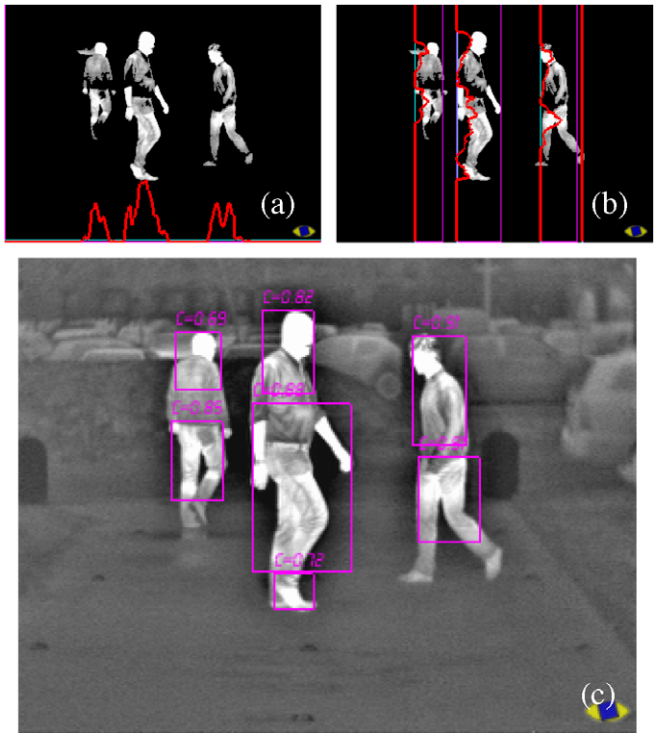

Fig. 3. Warm elements detection: (a) a column-wise histogram detects columns with warm object, (b) following row-histograms for each column detect initial areas of attention, (c) final results are shown on the original image.

or than other objects. Therefore, it is based on the detection of areas that contains a high amount of vertical edges.

Initially, acquired images are filtered using a Sobel operator and an adaptive threshold to detect nearly-vertical edges (Fig. 4b). Unfortunately, objects other than pedestrians contain vertical edges: buildings, cars, poles, etc. Vertical edges of these objects are generally longer and more uniform than the ones that belong to human shapes. Therefore, a filtering phase devoted to the removal of regular vertical edges that are longer than a given threshold is performed. Also isolated pixels are considered as noise and removed (Fig. 4c).

In order to further enhance vertical edges and to group edges of the same pedestrians in the same cluster, a morphological expansion is performed using a 3×7 operator. In the final result (Fig. 4d), only a few clusters of edges are present.

A labelling approach is used to compute connected clusters of pixels; the resulting image is analyzed and a list of bounding boxes containing connected clusters of pixels is built (Fig. 4e). Generally, in complex scenarios, a large number of small clusters is detected affecting both the effectiveness and the efficiency of the subsequent processing steps. Thus, a filtering is used to remove bounding boxes that are too small and contain cold areas only; only sufficiently large boxes or boxes that contain bright pixels survive this phase (Fig. 4f).

2.1.3. Disparity space image

A disparity space image [26] approach is used to strengthen pedestrian detection.

The right image (considered as a reference image) is subdivided in to 3×3 pixels regions and corresponding regions into the left image are searched for. The search for a homologous region is limited to the same row of the left image, since the optical axis of the two FIR cameras are parallel and thus two corresponding rows in the two images are epipolar lines. Moreover, calibration information permits to further bound the search area, reducing both the required computational burden and the risk of wrong matches. For each region in the reference image, the best matching region in the other image is considered and the disparity between their coordinates is computed.

The match is performed using the following correlation formula:

$$C = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3}(L_{i,j} \times R_{i,j})}{\max(\sum_{i=1}^{3}\sum_{j=1}^{3}(L_{i,j})^2, \sum_{i=1}^{3}\sum_{j=1}^{3}(R_{i,j})^2)} \quad (1)$$

where (i,j) are the coordinates of each pixel into the 3×3 regions. The result is a new image, the disparity space image (DSI), in which each pixel encodes a disparity value amongst the homologous pixels of the right and left images.

An aggregation step is used to perform obstacle detection: sufficiently wide areas featuring a similar disparity

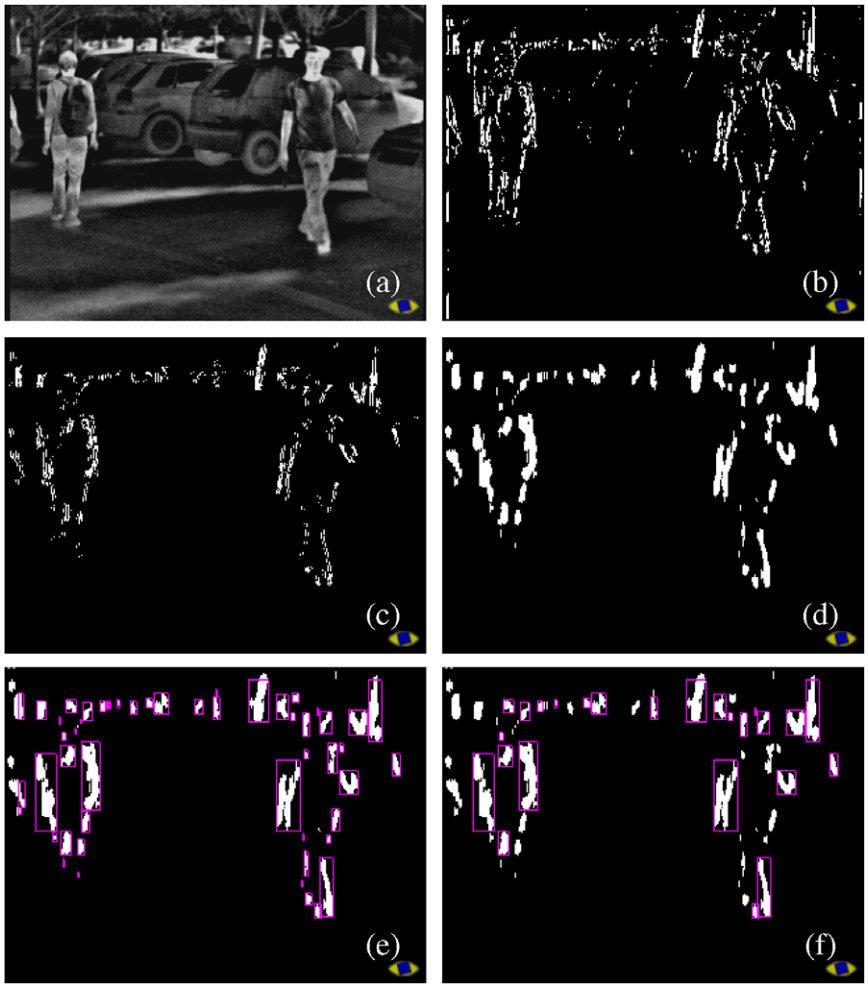

Fig. 4. Edges detection: (a) original image, (b) vertical edges detected using the Sobel operator and a threshold, (c) edges after the removal of strong and regular edges, (d) expanded edges, (e) bounding boxes containing connected clusters of pixels, and (f) final resulting list of bounding boxes.

are considered as obstacles, Fig. 5 shows the computed disparities for the obstacles in the scene (see Fig. 5).

An iterative approach based on the use of histograms similar to the one described in Section 2.1.1 is used to build a list of bounding boxes that contain areas of attention.

This process is fairly different from the other two previously described; in fact, in this case, 3D information about areas of attention are already available since the detection is based on stereo vision. Conversely, an additional step is needed for bounding boxes produced during the processings described in Sections 2.1.1 and 2.1.2 in order to compute the distance and size for attention areas.

2.2. Stereo match

This phase is used to match resulting areas of attention against the other image. Once a correspondence is found, it is possible to compute size and distance of the framed

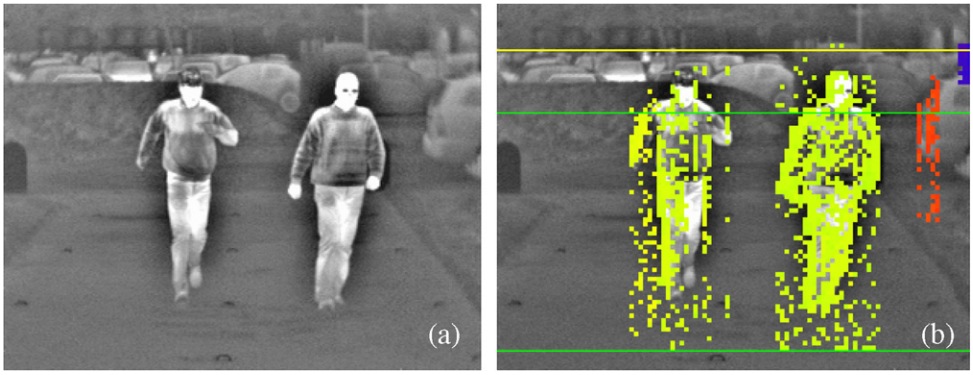

Fig. 5. Disparity space image approach: (a) original image and (b) disparity of obstacles, the darker the pixels the further the obstacle.

object. Only areas of attention computed using the approaches described in Sections 2.1.1 and 2.1.2 are processed during this step since distance and size are already available in the DSI approach.

By supposing that the optical axis 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 the calibration parameters.

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

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

where N is the number of pixels in the considered bounding box, a_{xy} and b_{xy} are grey-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. The algorithm discards the matches whose correlation values are lower than a given threshold (Fig. 6a).

A triangulation technique is used to estimate the distance between each object and the vision system.

2.3. Bounding boxes resize

The knowledge of the distances of bounding boxes allows a refinement of the boxes' size. In fact, assuming the terrain in front of the vision system is flat, it is possible to compute the point of contact between each object framed by a bounding box and the ground. Thus, the bounding boxes bottoms are stretched to the ground [25].

2.4. Bounding boxes merge and filtering

The knowledge of 3D information allows both to merge and preliminarily filter the areas of attention.

Bounding boxes located near the same position in the 3D world are, in fact, all grouped into a single and bigger bounding box (see Fig. 6b). In addition, a number of filters have been devised to get rid of false positives. Too small or too large bounding boxes that can not contain human shapes are removed; pedestrians are expected to stand, thus bounding boxes much wider than taller are discarded (Fig. 6c).

Distance and height of each potential pedestrian are evaluated as well: close bounding boxes that can not contain a whole human shape or far areas that do not allow a sufficiently reliable analysis are discarded as well.

2.5. Search for pedestrians head

For each potential pedestrian the presence of the most evident feature of a human shape in the infrared domain –the head– is searched for. The position of the head is not affected by pedestrian's pose, being always in the upper part of the bounding box and it is, often, warmer than the body. Two different models of a head are used to perform pattern matching operations.

The first model encodes thermal characteristics of a head warmer than the background, namely a binary mask showing a white head on a black background (Fig. 7) . To ease the match, the areas of attention are binarized as well using an adaptive threshold. The model is scaled according to the bounding boxes size and assuming that a head measures nearly 1/6 of the human shape height. The match is performed against an area centered around the top of the bounding box using Eq. (2). The highest correlation value obtained (P_w) is considered as the match quality.

Unfortunately, the head is not always warmer than the background. Environmental conditions, hats, helmets, or hair may mask heat radiation. In order to cope with this problem, an additional head model is used. This model encodes the head shape (Fig. 7) and is used to perform another match in the top area of each bounding box. In this case, the areas of attention are not binarized; for each position of the model, the two average values of pixels that correspond to the internal (white) or external (black) part of the model are computed and the quality of the match computed as the absolute value of the difference between these two averages. A higher difference is obtained in correspondence to object that feature a shape similar to the model. The shape matching quality (P_s) is computed as the highest of such differences.

The final match parameter (P_m) is computed as

$$P_m = 1 - ((1 - P_w) \times (1 - P_s)).$$

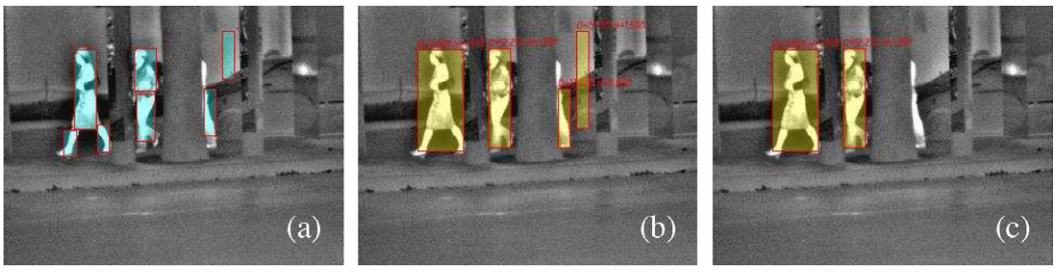

Fig. 6. (a) Homologous bounding boxes in the left image, (b) bounding box bases are resized to reach the ground and elements with similar 3D coordinates are grouped together, (c) bounding boxes that can contain a human shape.

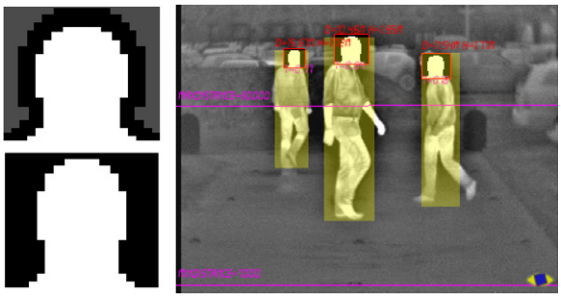

Fig. 7. Head search: the two models used for head matching and the detected match into the areas of attention.

The portion of the bounding box that produces the best match value is recognized as the head of a potential pedestrian while boxes featuring a bad match are discarded as not containing a human shape.

2.6. Pedestrian shape detection by means of snakes

In order to enhance the robustness and reliability of bounding boxes validation, an object contour detection by means of snakes is performed.

A snake is an elastic curve placed on an image that begins to deformate itself from an initial shape in order to adjust to the image features. The deformation is a result of the action of external forces, that attract the snake toward image features, and internal forces, that keep smooth the shape of the curve. The final position corresponds to a minimum of the snake total energy

Fig. 8. Pedestrian detection results: detected pedestrians are shown using a superimposed yellow box. Each box also shows the distance and height of detected object. The two magenta lines show the minimum and maximum distances used for the detection.

[27]. Snakes, or active contour models, have been widely used in image segmentation and understanding to refine the shape of detected object.

Formally, a snake is a spline $\mathbf{v}(s) = [x(s), y(s)]$, $s \in [0,1]$, which moves through the spatial domain of an image to minimize the following energy function:

$$E_{\text{snake}} = \int_0^1 (E_{\text{int}}(\mathbf{v}(s)) + E_{\text{ext}}(\mathbf{v}(s)))\mathrm{d}s \tag{3}$$

E_{int} represents the internal energy. It defines the snake physical properties and depends only on its internal status. It is a function of both bending and stretching forces applied to the snake. E_{ext} is the external energy of the snake and represents the forces pushing the snake toward the desired object. The forces, used for shape detection, are computed using image edges (E_{edge}) and hot areas (E_{hot}). E_{ext} is expressed as a weighted combination of two factors:

$$E_{\text{ext}} = w_1 E_{\text{edge}} + w_2 E_{\text{hot}} \tag{4}$$

The Edge energy functional (5) is computed using a Sobel gradient operator. The result is smoothed by an average filter. This processing helps the snake to converge toward energy minimum, since the snake can be attracted to this minimal from farther away. Fig. 9 shows an example of E_{edge} computation.

$$E_{\text{edge}} = -G_\sigma * \nabla I(x,y) \tag{5}$$

In the far-infrared domain the image of an object depends on its temperature. The pedestrian temperature often is higher than the environment and human shapes appear brighter than the background. Therefore the image intensity is used to attract the snake toward the image hot areas: $E_{\text{hot}} = -I(x,y)$. Fig. 10 shows an example of E_{hot} computation.

A typical snake problem is the initialization stage. In fact, when the initial position is far away from the desired result, the snake often gets stuck in an local minimum of the energy functional. To avoid such a behavior, the snake must be initialized close to the target object. In the discussed system, the snake starts from the bounding box border but the use of the disparity space image is currently under evaluation to initialize the snake closer to the object.

Many solutions [28,27,29] to the snake minimization problem have been proposed in literature. The greedy snake algorithm [28], applied on 5×5 neighborhood, has been used in this work. Fig. 11 shows a few results of shape detection. The detected shape will then be validated by means of neural networks [30] or pedestrian models [31].

3. Results

The developed system has been tested in different situations using an experimental vehicle equipped with two Raytheon 300D 7–14 μ infrared cameras.

Fig. 8 shows a few results: detected human shapes are evidenced using a superimposed yellow box, distance (D) and height (H) are indicated on the top of each box. Thanks to the triangulation information, the system has proven to be able to detect pedestrians even if they partly overlap each other (Fig. 12a). In addition, the use of three different approaches for the detection allows to detect pedestrians in complex scenarios or even when they are not warmer than the background.

Moreover, even though the model used for head detection is quite simple, the head of each pedestrian is properly localized. The model represents a frontal head shape but it is useful also for side shots (Fig. 12b). In Fig. 12c the pedestrian in the foreground is not detected, even if visible, because its height is too small for an human shape and the corresponding bounding box has no valid aspect-ratio, so the filtering phase discarded it.

The main problems are related to the bounding boxes aspect ratio. Sometimes aspect ratio is not a good evaluation criterion for filtering results. Fig. 13a shows that a pedestrian can produce a bounding box that is compatible with other possible objects in the scene, like cars (Fig. 13b). This problem can appear in different situations causing the system to get false positive results despite the head pattern matching filtering.

Another problem concerns groups of pedestrians; if pedestrians are very close to each others and at the same distance from the vision system, they are often detected as a single pedestrian and head detection produces a match anyway (Fig. 13c). Another problem is due to the precision of calibration: any deviation affects distance and size computation for detected objects and subsequent steps, like the refinement of bounding boxes to the ground (Fig. 13d).

Other detection failures are due to occlusions, but this problem is only observed in a few frames, a tracking based system has been already proven to be able to cope with these particular cases [32].

Fig. 14 shows the detection of objects that feature a similar heat/edge distribution of a pedestrian. Unfortunately, objects with a similar size and shape of a pedestrian represent the most frequent cause of misdetection. Therefore, an accurate validation phase must be developed. Unfortunately, the head match filter has proven to be not sufficient for an effective validation. In fact, even if it has been improved with respect to the one discussed in [25], it still needs to be optimized. Moreover, an effective validation process can not be based on head detection only; in fact, several artifacts feature a head like shape and also the sun illumination can heat them causing the detection of a false positive. Currently, snakes are not used in the validation process, but the refined shape that can be obtained is suitable for such a validation. Two different approaches are currently under development: the use of a match with the shape of pre-computed models or a neural network based computation.

The system has been tested using a tool for performance evaluation [33]. A prerecorded sequence of more than 5000 images was manually annotated drawing a bounding box

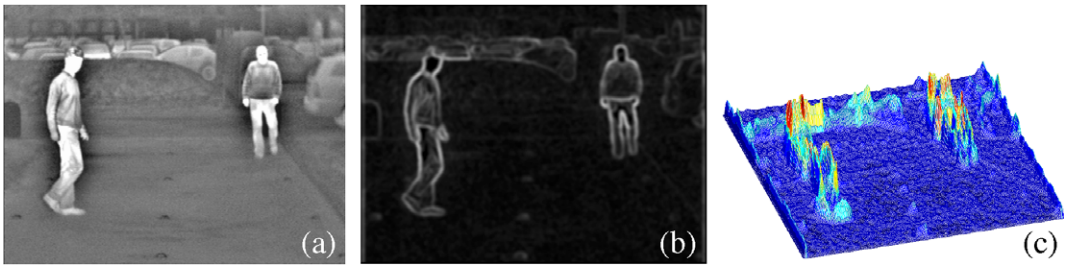

Fig. 9. E_{edge} processing: (a) original image, (b) edge image obtained using Sobel and average operator, (c) edge energy functional.

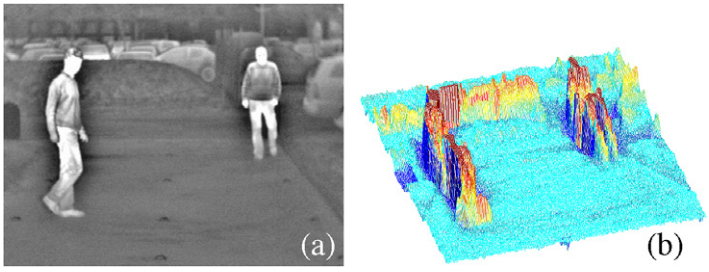

Fig. 10. E_{hot} processing: (a) original image, (b) intensity energy functional.

Fig. 11. Shape detection results: the white rectangles are the bounding boxes where the snakes are initialized and the yellow curves represent the final positions of the snakes.

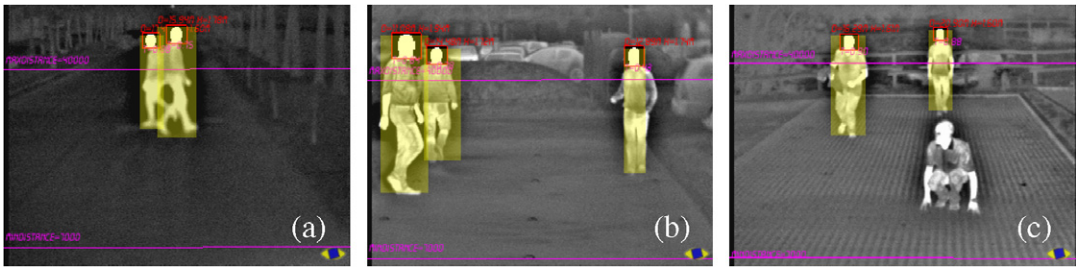

Fig. 12. Results: (a) the system is able to detect pedestrian even if partially occluded, (b) the head model is appropriate whether for frontal or side shots, and (c) pedestrian in foreground is not properly detected due to height and aspect ratio constraints.

over each pedestrian on the scene. The system has been tested on that sequence and the results were compared against the ground truth of the annotation.

The ROC curve shown in Fig. 15 was obtained varying the correlation threshold used for the detection. The first curve (Fig. 15a) has been obtained by running the system enabling warm areas detection, edge detection, and DSI approaches (therefore the system discussed in this article and initially presented in [34]). It can be noticed that in this condition the system is able to correctly detect more than 80% of pedestrians still maintaining a low value for false detections. The second ROC curve (Fig. 15b) has been obtained using warm areas detection only (the system presented in [25]), and thus shows how edge and DSI approaches improve the overall result while barely affecting false detections.

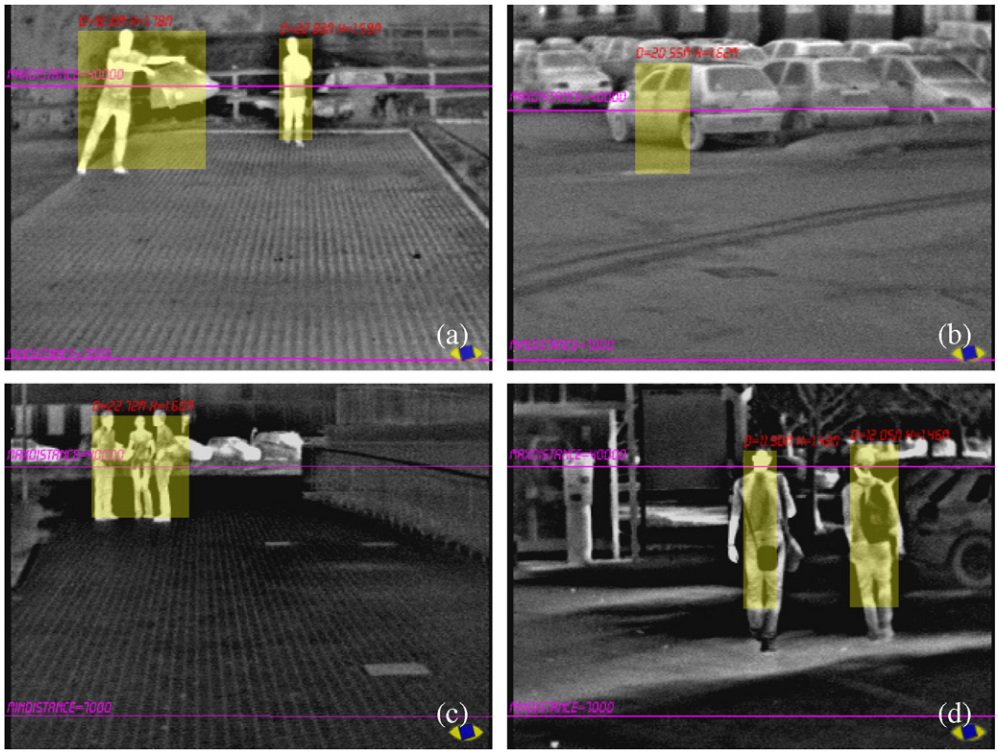

Fig. 13. Detection problems: (a) shows a pedestrian in a unusual positions; its aspect ratio is similar to other possible objects, like in (b); (c) shows a group of pedestrians detected as a single one, and (d) presents a case where an error in the calibration affects distance computation and also box refinement.

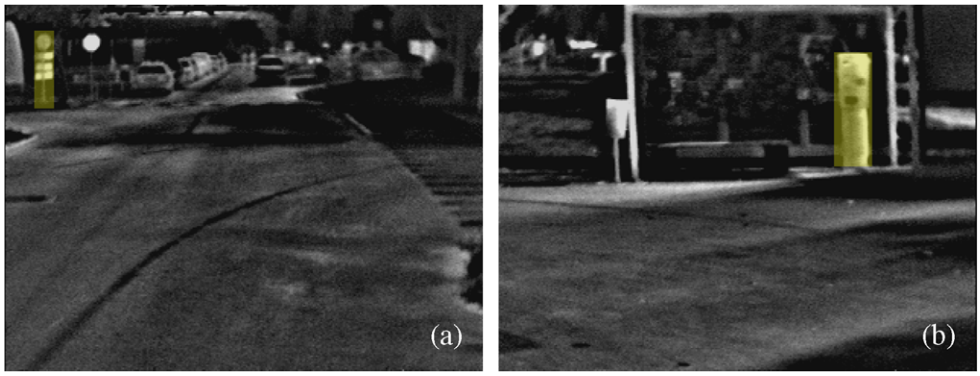

Fig. 14. Misdetections of objects with pedestrian like size or shape: (a) a sign heated by the sun and (b) warm ticket vending machine.

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. Since processing time strongly depends on the complexity of the scene, both urban and rural sequences have been used to evaluate temporal performance; the average execution time on the reference architecture is 84 ms; that means that the system is able to process nearly 12 frames per second.

4. Conclusions

In this paper a stereo vision-based algorithm aimed at the detection of pedestrians in infrared images has been discussed. It has been tested in urban and rural environments using an experimental vehicle equipped with two infrared cameras that work in the 7–14 μ spectrum.

The algorithm is based on three different approaches: the detection of warm areas, the detection of vertical edges, and a DSI computation. Distance estimation, size, aspect ratio, and a head presence are used to select pedestrians. Neither temporal correlation, nor motion cues are currently used for the processing.

Experimental results demonstrated that this approach is promising. The presence of two additional approaches not based on the detection of thermal characteristics permits an increase to the detection ratio with respect to the previous approach [25] and to detect pedestrians

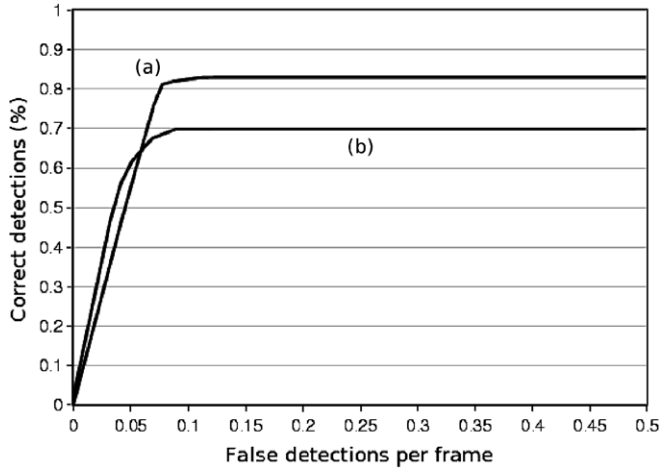

Fig. 15. ROC curves of the results with (a) all underlying approaches activated or (b) only warm areas detection.

even if they are not warmer than the background. Correct detection percentage is high with a very low number of false detections per frame, and the system has proven to work reliably even when pedestrians are partly occluded.

In urban situations, noise produced by the presence of buildings, cars, signals and other objects could increase false detections. In order to enhance the robustness and reliability of the discussed system and to filter out false positives, a refinement system for the shape of detected objects has been developed. A validation phase using the refined shape is currently under development.

A tracking system has been tested as well [32]. It was demonstrated to be useful for removing false negatives due to occlusions, but it is not suited for removing false positives.

References

[1] R. Kania, M. Del Rose, P. Frederick, Autonomous robotic following using vision based techniques, in: R. Raja (Ed.), Proc. Ground Vehicle Survivability Symposium, Monterey, USA, 2005.
[2] M. Del Rose, P. Frederick, Pedestrian detection, in: Proc. Intelligent Vehicle Systems Symposium, Traverse City, USA, 2005.
[3] M. Del Rose, P. Frederick, J. Reed, Pedestrian detection for anti-tamper vehicle protection, in: Proc. Ground Vehicle Survivability Symposium, Monterey, USA, 2005.
[4] K.C. Frerstenberg, J. Dietmayer, V. Willhoeft, Pedestrian recognition in urban traffic using vehicle based multilayer laserscanner, in: Proc. Automobile Engineers Cooperation International Conf., Paris, France, 2001.
[5] K.C. Frerstenberg, U. Lages, Pedestrian detection and classification by laserscanners, in: Proc. IEEE Intelligent Vehicles Symposium 2002, Paris, France, 2002.
[6] A. Fod, A. Howard, M.J. Mataric, Laser-based people tracking, in: Proc. IEEE Intl. Conf. on Robotics and Automation, Washington, DC, USA, 2002.
[7] A. Brooks, S. Williams, Tracking People with Networks of Heterogeneous Sensors, in: Proc. Australian Conf. on Robotics and Automation, Brisbane, Australia, 1999.
[8] NIT Phase II: Evaluation of Non-Intrusive Technologies for Traffic Detection, Tech. Rep. SRF No. 3683, Mn DOT Research Report (2002).
[9] S. Milch, M. Behrens, Pedestrian detection with radar and computer vision, in: Proc. Conf. on Progress in Automobile Lighting, Darmstadt, Germany, 2001.
[10] M. Koltz, H. Rohling, 24 GHz radar sensors for automotive applications, in: Proc. Intl. Conf. on Microwaves and Radar, Warsaw, Poland, 2000.
[11] A. Shashua, Y. Gdalyahu, G. Hayun, Pedestrian detection for driving assistance systems: single-frame classification and system level performance, in: Proc. IEEE Intelligent Vehicles Symposium 2004, Parma, Italy, 2004.
[12] G.P. Stein, O. Mano, A. Shashua, Vision based ACC with a single camera: bounds on range and range rate accuracy, in: Proc. IEEE Intelligent Vehicles Symposium 2003, Columbus, USA, 2003.
[13] L. Zhao, Dressed Human Modeling, Detection, and Parts Localization, Ph.D. dissertation, Carnegie Mellon University (2001).
[14] C. Papageorgiou, T. Evgeniou, T. Poggio, A Trainable Pedestrian Detection System 38 (2000) 15–33.
[15] D. Beymer, K. Konolige, Real-time tracking of multiple people using continuous detection, in: Proc. Intl. Conf. on Computer Vision, Kerkyra, 1999.
[16] H. Shimizu, T. Poggie, Direction estimation of pedestrian from multiple still images, in: Proc. IEEE Intelligent Vehicles Symposium 2004, Parma, Italy, 2004.
[17] K. Grauman, T. Darrell, Fast contour matching using approximate earth mover's distance, Tech. Rep. AI Memo, AIM-2003-026, MIT (2003).
[18] R. Cutler, L.S. Davis, Robust real-time periodic motion detection, analysis and applications, IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (8) (2000) 781–796.
[19] D.M. Gavrila, The visual analysis of human movement: a survey, Computer Vision and Image Understanding 73 (1) (1999) 82–98.
[20] S. Tate, Y. Takefuji, Video-based human shape detection deformable templates and neural network, in: Proc. of Knowledge Engineering System Conf., Crema, Italy, 2002.
[21] J.W. Davis, V. Sharma, Robust background-subtraction for person detection in thermal imagery, in: Proc. Intl. IEEE Wks. on Object Tracking and Classification Beyond the Visible Spectrum, Washington DC, USA, 2004.
[22] H. Nanda, L. Davis, Probabilistic template based pedestrian detection in infrared videos, in: Proc. IEEE Intelligent Vehicles Symposium 2002, Paris, France, 2002.
[23] B. Bhanu, J. Han, Kinematics-based human motion analysis in infrared sequences, in: Proc. IEEE Intl. Workshop on Applications of Computer Vision, Orlando, USA, 2002.
[24] F. Xu, K. Fujimura, Pedestrian detection and tracking with night vision, in: Proc. IEEE Intelligent Vehicles Symposium 2002, Paris, France, 2002.
[25] M. Bertozzi, A. Broggi, M. Del Rose, A. Lasagni, Infrared stereo vision-based human shape detection, in: Proc. IEEE Intelligent Vehicles Symposium 2005, Las Vegas, USA, 2005, pp. 23–28.
[26] A. Broggi, C. Caraffi, R.I. Fedriga, P. Grisleri, Obstacle detection with stereo vision for off-road vehicle navigation, in: Proc. Intl. IEEE Wks. on Machine Vision for Intelligent Vehicles, San Diego, USA, 2005.
[27] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vision 1 (4) (1988) 321–331.
[28] D.J. Williams, M. Shah, A fast algorithm for active contours and curvature estimation, CVGIP: Image Understanding 55 (1) (1992) 14–26.
[29] C. Xu, J. Prince, Snakes, shapes, and gradient vector flow, IEEE Trans. on Image Processing 7 (3) (1998) 359–369.
[30] K. Tabb, S. George, R. Adams, N. Davey, Human shape recognition from snakes using neural networks, in: Proc. Intl. Conf. on Computer Intelligence and Multimedia applications, 1999.

[31] A. Broggi, A. Fascioli, P. Grisleri, T. Graf, M.-M. Meinecke, Model-based validation approachesand matching techniques for automotive vision based pedestrian detection, in: Proc. Intl. IEEE Wks. on Object Tracking and Classification in and Beyond the Visible Spectrum, San Diego, USA, 2005.

[32] E. Binelli, A. Broggi, A. Fascioli, S. Ghidoni, P. Grisleri, T. Graf, M.-M. Meinecke, A modular tracking system for far infrared pedestrian recognition, in: Proc. IEEE Intelligent Vehicles Symposium 2005, Las Vegas, USA, 2005, pp. 758–763.

[33] M. Bertozzi, A. Broggi, P. Grisleri, A. Tibaldi, M. Del Rose, A tool for vision based pedestrian detection performance evaluation, in: Proc. IEEE Intelligent Vehicles Symposium 2004, Parma, Italy, 2004, pp. 784–789.

[34] M. Bertozzi, E. Binelli, A. Broggi, M. Del Rose, Stereo vision-based approaches for pedestrian detection, in: Proc. Intl. IEEE Wks. on Object Tracking and Classification in and Beyond the Visible Spectrum, San Diego, USA, 2005.