Segmenting Humans from Mobile Thermal Infrared Imagery

José Carlos Castillo, Juan Serrano-Cuerda, Antonio Fernández-Caballero, and María T. López

Universidad de Castilla-La Mancha, Departamento de Sistemas Informáticos & Instituto de Investigación en Informática de Albacete Campus Universitario s/n, 02071-Albacete, Spain `caballer@dsi.uclm.es`

Abstract. Perceiving the environment is crucial in any application related to mobile robotics research. In this paper, a new approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the indoor autonomous mobile platform mSecurit^{TM} is introduced. The approach starts with a phase of static analysis for the detection of human candidates through some classical image processing techniques such as image normalization and thresholding. Then, the proposal uses Lukas and Kanade optical flow without pyramids algorithm for filtering moving foreground objects from moving scene background. The results of both phases are compared to enhance the human segmentation by infrared camera. Indeed, optical flow will emphasize the foreground moving areas gotten at the initial human candidates detection.

1 Introduction

Perceiving the environment is crucial in any application related to mobile robotics research [8,3]. The information surrounding the robot can be used to navigate, avoid barriers and execute a given mission [10]. As an outstanding motion detection method, optical flow is being widely used in mobile robot navigation. Optical flow plays a central role in the ability of primates to recognize movement. Image flow divergence has been used to orient a robot within indoor hallways by estimating time to collision [4] and differences in optical flow magnitude have been used to classify objects moving at different speeds to simplify road navigation in traffic [9]. Optic flow has also been used in a dynamical model of visually-guided steering, obstacle avoidance and route selection [6]. An approach uses optical flow information to track features on objects from the time they appear on screen until they interact with the local sensors of the robot [13].

Moreover, the use of autonomous robots or vehicles can provide significant benefits in the surveillance field [19,14]. And, many algorithms focusing specifically on the thermal domain have been explored. The unifying assumption in most of these methods is the belief that the objects of interest are warmer than their surroundings [21]. Indeed, some animals can see in total darkness, or even

see colors beyond the visual spectrum, that humans have never seen. Thermal infrared video cameras detect relative differences in the amount of thermal energy emitted/reflected from objects in the scene. As long as the thermal properties of a foreground object are slightly different (higher or lower) from the background radiation, the corresponding region in a thermal image appears at a contrast from the environment. In [11,2], a thresholded thermal image forms the first stage of processing after which methods for pose estimation and gait analysis are explored. In [18], a simple intensity threshold is employed and followed by a probabilistic template. A similar approach using Support Vector Machines is reported in [20]. Recently, a new background-subtraction technique to robustly extract foreground objects in thermal video under different environmental conditions has been presented [5]. A recent paper [12] presents a real-time ego-motion estimation scheme that is specifically designed for measuring vehicle motion from a monocular infra-red image sequence at night time. In the robotics field, a new type of infrared sensor is described [1]. It is suitable for distance estimation and map building. Another application using low-cost infrared sensors for computing the distance to an unknown planar surface and, at the same time, estimating the material of the surface has been described [7].

In this paper, we introduce our approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the indoor autonomous mobile platform mSecurit^{TM} (see Fig. 1)developed by the Spanish private company MoviRobotics S.L. The approach starts with a phase of static analysis (on the current image frame) for the detection of human candidates. Then, a dynamic analysis [15,17] (taking the previous and the current images) by means of an optical flow algorithm based on Lukas and Kanade [16] approach without pyramids is run. The algorithm aligns two images to achieve the best matches and determines motion between both images. The approach assumes the images to be roughly aligned and uses Newton-Raphson iteration for the gradient of error. Lastly, the results of both phases are compared to efficiently segment the humans.

2   Human Detection Algorithm

The proposed human detection algorithm is explained in detail in the following sections related to the different phases, namely, human candidates blob detection, image motion analysis, and human blobs segmentation.

2.1   Human Candidates Blob Detection

The algorithm starts with the analysis of a single image, $I(r, c, t)$, captured at a time t by the camera. This phase is considered as static because it uses only the information of the current frame to perform the detection. Firstly, a change in scale, as shown in equation (1) is performed. The idea is to normalize all images to always work with a similar scale of values, transforming $I(r, c, t)$ to $I'(r, c, t)$. The normalization assumes a factor $\gamma = 60$, as our empirical experience

Fig. 1. The mSecurit^{TM} mobile surveillance robot

shows that this value corresponds to the mean gray level value of an image, $\overline{I}(t)$, captured at a standard environment temperature (see figure 2b).

$$I'(r,c,t) = \frac{I(r,c,t) \times \gamma}{\overline{I}(t)} \qquad (1)$$

where $I'(r,c,t)$ is the normalized image. Notice that $I'(r,c,t) = I(r,c,t)$ when $\overline{I}(t) = \gamma$.

The next step is the elimination of incandescent points (corresponding to light bulbs, fuses, and so on), which can confuse the algorithm by to showing zones with high temperature. As the image has been scaled, the threshold θ_i calculated to eliminate these points is related to the normalization factor γ. Indeed,

$$\theta_i = 3 \times \frac{5}{4}\gamma \qquad (2)$$

$\delta = \frac{5}{4}\gamma$ introduces a tolerance value of a 25% above the mean image value. And, $3 \times \delta$ provides a value high enough to be considered an incandescent image pixel. Thus, pixels with a higher gray value are discarded and filled up with the mean gray level of the image.

$$I'(r,c,t) = \begin{cases} I'(r,c,t), & \text{if } I'(r,c,t) \leq \theta_i \\ \overline{I'}(t), & \text{otherwise} \end{cases} \qquad (3)$$

The algorithm uses a threshold to perform a binarization for the aim of isolating the human candidates spots. The threshold θ_h, obtains the image areas containing moderate heat blobs, and, therefore, belonging to human candidates. Thus, warmer zones of the image are isolated where humans could be present. The soft threshold is calculated as:

$$\theta_h = \frac{5}{4}(\gamma + \sigma_{I'}) \qquad (4)$$

a)   b)   c)

Fig. 2. (a) Input IR image frame. (b) Scaled frame. (c) Closed soft threshold.

where $\sigma_{I'}$ is the standard deviation of image $I'(r,c,t)$. Notice, again, that a tolerance value of a 25% above the sum of the mean image gray level value and the image gray level value standard deviation is offered.

Now, image $I'(r,c,t)$ is binarized using the threshold. Pixels above the threshold are set as maximum value $max = 255$ and pixels below are set as minimum value $min = 0$.

$$B_s(r,c,t) = \begin{cases} min, \text{ if } I'(r,c,t) \leq \theta_h \\ max, \text{ otherwise} \end{cases} \tag{5}$$

Next, the algorithm performs morphological opening (equation (6)) and closing equation (7)) operations to eliminate isolated pixels and to unite areas split during the binarization. These operations require structuring elements that in both cases are 3×3 square matrixes centered at position $(1,1)$. These operations greatly improve the binarized shapes as shown in Fig. 2.

$$B_o(r,c,t) = B_s(r,c,t) \circ \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{6}$$

$$B_c(r,c,t) = B_o(r,c,t) \bullet \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{7}$$

Afterwards, the blobs contained in the image are obtained. A minimum area, B_{min}, - function of the image size - is established for a blob to be considered to contain humans. As a result image $B_r(r,c,t)$ is obtained by eliminating non-human blobs from image $B_c(r,c,t)$.

2.2   Image Motion Analysis

In this phase, dynamic analysis - or image motion analysis - by optical flow calculation is performed. Optical flow has been selected as it discards the scene movement due to the proper robot motion. A simple subtraction-based approach would indicate that everything is in movement, making impossible to differentiate really moving objets in the completely moving scene. Thus, as the majority
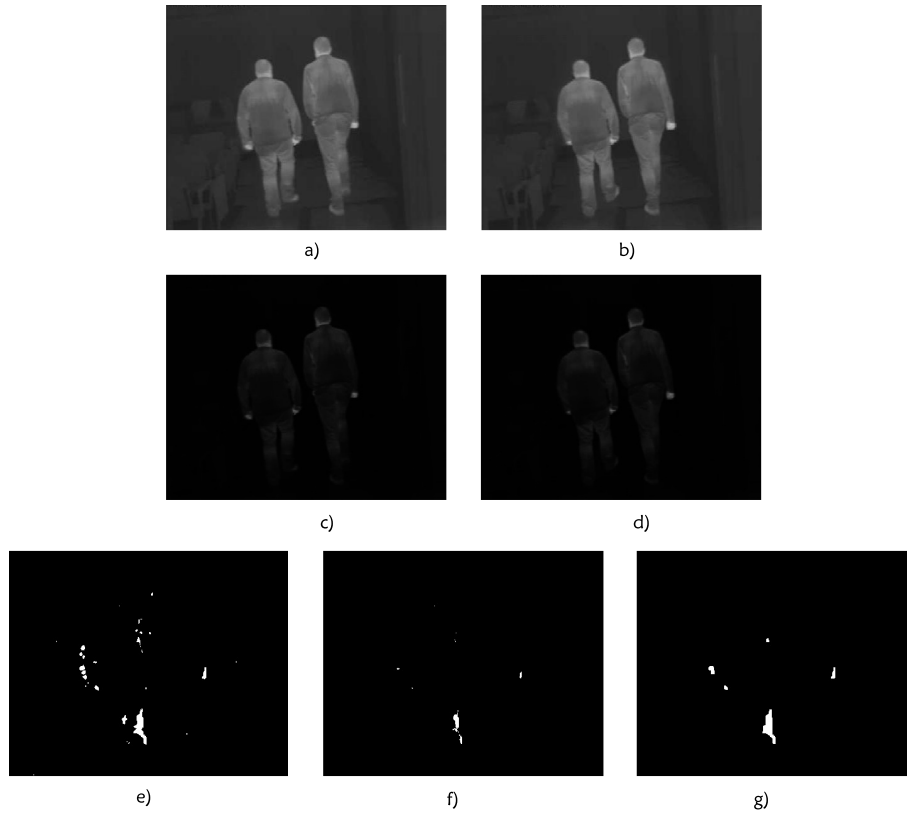
Fig. 3. (a) Previous frame. (b) Current frame. (c) Multiplied previous frame. (d) Multiplied current frame. (e) Soft thresholded moments. (f) Hard thresholded moments. (g) Matched thresholds.

movement is the scene movement, optical flow discards it to only focus in other different direction movements [16].

Evidently, this phase uses two image frames, the previous image, $I(r, c, t-1)$, and the current one, $I(r, c, t)$ (see Fig. 3a and 3b). In first place, the current and the previous frames are multiplied to enhance the contrast, such that the dark values become darker and the bright values become brighter (see Fig. 3c and 3d). This way, the calculation of the optical flow is facilitated.

The dynamic analysis requires the calculation of the moments corresponding to each pixel movement on the normalized input images, applying equation 1 on images $I(r, c, t-1)$ and $I(r, c, t)$. The optical flow calculation results into two gray level images, where each pixel reflects the angular moment detected, storing the movements in X and Y axes. Firstly, the algorithm performs the speed calculation of the optical flow. The selected optical flow approach is the Lucas-Kanade without pyramids algorithm. This algorithm is fast and offers an excellent success vs. speed ratio. The calculated speeds, as a result of the optical
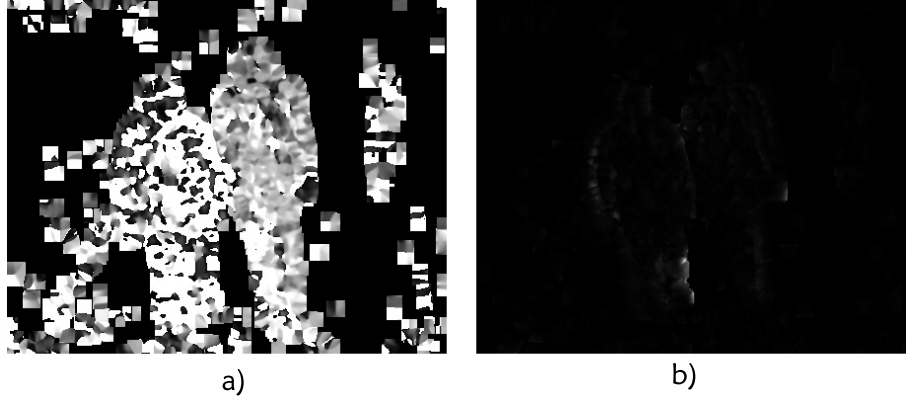
a)   b)

Fig. 4. Optical flow calculation. (a) Moments. (b) Angles.

flow, are turned into angles, $\alpha(r, c, t)$, and magnitudes, $m(r, c, t)$. Fig. 4a shows the magnitudes (moments), that is to say, the amount of movement at each pixel (r, c) between $I'(r, c, t - 1)$ and $I'(x, y, t)$, in form of a moments image, $M(r, c, t)$. Similarly, Fig. 4b shows the direction of the movement (angles). The results clearly indicate that angles are less important than moments. Indeed, on the one hand, non-rigid objects' movements go into very different directions, and, on the other side, angles with low moments may be caused by image noise.

To efficiently use the moments image $M(r, c, t)$, its histogram, as shown in Fig. 5 has been studied for many cases. As you may observe, most values are in the $[0, 64]$ interval, but very close to 0. Indeed, the average value is close to 1 in these moments images. Therefore, two thresholds, a moments soft threshold $\mu_s = 10$ and a moments hard threshold $\mu_h = 25$, are used to delimit the blobs of possible humans. The aim of the soft threshold, μ_s, is to obtain the most representative values, whereas the hard threshold, μ_h, is used to refine a better matching between zones that show an elevated movement and zones with less movement but connected to the previous ones. Thus, the zones where movement has been detected are extended, and the zones with reduced movements are eliminated.

Therefore, firstly, the moments soft threshold μ_s is applied to the moments image $M(r, c, t)$ to obtain image $M_s(r, c, t,)$, as shown in Fig. 3e). The related formula is:

$$M_s(r, c, t) = \begin{cases} min, \text{ if } M(r, c, t) \leq \mu_s \\ max, \text{ otherwise} \end{cases} \tag{8}$$

Afterwards, an opening filter is applied to erase isolated pixels, getting $M_o(r, c, t,)$ (see equation (9)). In this case, disconnected areas can arise, as parts of the image may have gone in different directions.

$$M_o(r, c, t) = M_s(r, c, t) \circ \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{9}$$
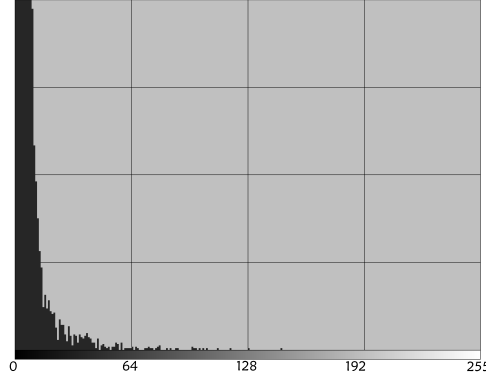
Fig. 5. Moments histogram

After this, the moments hard threshold, $\mu_h = 25$, is applied to $M(r, c, t,)$ in order to obtain image $M_h(r, c, t)$ (see Fig. 3f and equation (10)).

$$M_h(r, c, t) = \begin{cases} min, \text{ if } M(r, c, t) \le \mu_h \\ max, \text{ otherwise} \end{cases} \tag{10}$$

Now, the blobs present in $M_o(r, c, t,)$ are compared to the blobs of $M_h(r, c, t)$. The aim is to verify if each blob detected with the hard threshold is contained in a spot detected with the soft threshold. The spots that do not meet this condition are discarded. Finally, the resulting image, called refined moments image $M_r(r, c, t)$, and shown in Fig. 3g, only contains the blobs that have met the previous condition. This image is used during the next phase to improve the certainty about the human presence.

2.3   Human Blobs Segmentation

This phase enhances the human detection by combining the results of the previous phases, that is, the human candidates blobs image and the refined moments image.

Indeed, this phase performs an *AND* operation between the output images of both previous phases, $B_r(r, c, t)$ and $M_r(r, c, t)$. The aim here is to take advantage of the optical flow information to improve the detection performed in the static analysis. This is, the optical flow emphasizes the moving areas gotten at the initial human candidates detection. The possibilities that these moving shapes are humans are increased, as the resulting image

$$P_r(r, c, t) = B_r(r, c, t) \cap M_r(r, c, t) \tag{11}$$

verifies if there exists a sufficient amount of pixels in movement within the zones delimited in $B_r(r, c, t)$.

3   Results

The algorithm was tested on a motherboard (an Intel Celeron M 430 at 1.73 GHz) and processor installed on the mSecurit^{TM} mobile robot. The RAM unit has a capacity of 512 MB. The performance results in terms of real-time capability of the algorithms described are excellent, as the method deals with the 6 frames per second provided by the FLIR camera installed on the mSecurit^{TM} mobile platform.

Fig. 6 shows the output of applying the proposed human detection algorithm on an IR video sequence. As you may easily observe, in all frames captured, the human is perfectly detected.
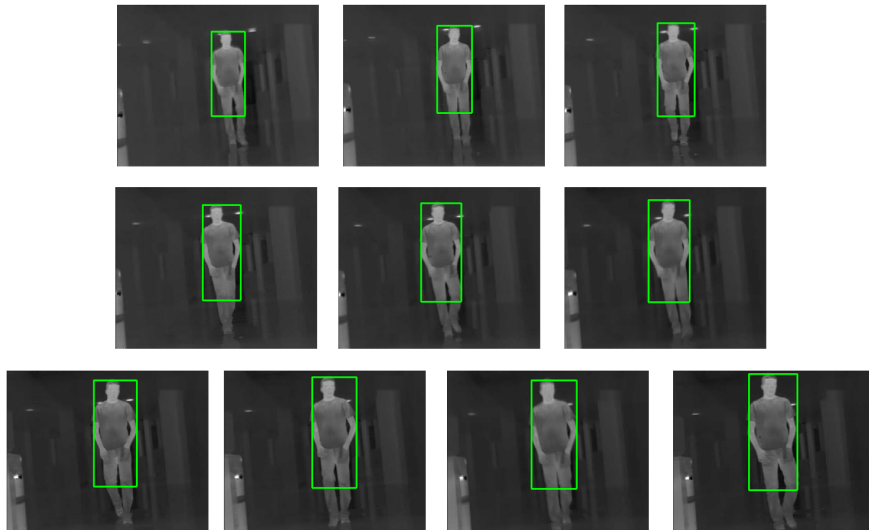
Fig. 6. A complete human detection video IR sequence

4   Conclusions

In this paper, our approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the Spanish private company MoviRobotics S.L. indoor autonomous mobile platform mSecurit^{TM} has been presented. The approach starts with a phase of static analysis for the detection of human candidates. Then, a dynamic analysis by means of and optical flow algorithm based on Lukas and Kanade approach without pyramids is run. The algorithm aligns two images to achieve the best matches and determines motion between both images. The approach assumes the images to be roughly aligned and uses Newton-Raphson iteration for the gradient of error. The initial results are promising and we are now engaged in performing tests in different visual surveillance scenarios.

Acknowledgements

References

1. Benet, G., Blanes, F., Simó, J.E., Pérez, P.: Using infrared sensors for distance measurement in mobile robots. Robotics and Autonomous Systems 40(4), 255–266 (2002)
2. Bhanu, B., Han, J.: Kinematic-based human motion analysis in infrared sequences. In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, pp. 208–212 (2002)
3. Cherubini, A., Oriolo, G., Macrí, F., Aloise, F., Cincotti, F., Mat, D.: A multimode navigation system for an assistive robotics project. Autonomous Robots 25(4), 383–404 (2008)
4. Coombs, D., Herman, M., Hong, T., Nashman, M.: Real-time obstacle avoidance using central flow divergence, and peripheral flow. IEEE Transactions on Robotics and Automation 14(1), 49–59 (1998)
5. Davis, J.W., Sharma, V.: Background-subtraction in thermal imagery using contour saliency. International Journal of Computer Vision 71(2), 161–181 (2007)
6. Fajen, B.R., Warren, W.H., Temizer, S., Kaelbling, L.P.: A dynamical model of visually-guided steering, obstacle avoidance, and route selection. International Journal of Computer Vision 54(1-3), 13–34 (2003)
7. Garcia, M.A., Solanas, A.: Estimation of distance to planar surfaces and type of material with infrared sensors. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 1, pp. 745–748 (2004)
8. Gascueña, J.M., Fernández-Caballero, A.: Agent-based modeling of a mobile robot to detect and follow humans. In: Håkansson, A., et al. (eds.) KES-AMSTA 2009. LNCS (LNAI), vol. 5559, pp. 80–89. Springer, Heidelberg (2009)
9. Giachetti, A., Campani, M., Torre, V.: The use of optical flow for road navigation. IEEE Transactions on Robotics and Automation 14(1), 34–48 (1998)
10. Guo, L., Zhang, M., Wang, Y., Liu, G.: Environmental perception of mobile robot. In: Proceedings of the 2006 IEEE International Conference on Information Acquisition, pp. 348–352 (2006)
11. Iwasawa, S., Ebihara, K., Ohya, J., Morishima, S.: Realtime estimation of human body posture from monocular thermal images. In: Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 15–20 (1997)
12. Jung, S.-H., Eledath, J., Johansson, S., Mathevon, V.: Egomotion estimation in monocular infra-red image sequence for night vision applications. In: IEEE Workshop on Applications of Computer Vision, p. 8 (2007)
13. Lookingbill, A., Rogers, J., Lieb, D., Curry, J., Thrun, S.: Reverse optical flow for self-supervised adaptive autonomous robot navigation. International Journal of Computer Vision 74(3), 287–330 (2007)

14. López, M.T., Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Visual surveillance by dynamic visual attention method. Pattern Recognition 39(11), 2194–2211 (2006)
15. López, M.T., Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Motion features to enhance scene segmentation in active visual attention. Pattern Recognition Letters 27(5), 469–478 (2006)
16. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (1981)
17. Mira, J., Delgado, A.E., Fernández-Caballero, A., Fernández, M.A.: Knowledge modelling for the motion detection task: The algorithmic lateral inhibition method. Expert Systems with Applications 27(2), 169–185 (2004)
18. Nanda, H., Davis, L.: Probabilistic template based pedestrian detection in infrared videos. In: Proceedings of the IEEE Intelligent Vehicle Symposium, vol. 1, pp. 15–20 (2002)
19. Pavón, J., Gómez-Sanz, J., Fernández-Caballero, A., Valencia-Jiménez, J.J.: Development of intelligent multi-sensor surveillance systems with agents. Robotics and Autonomous Systems 55(12), 892–903 (2007)
20. Xu, F., Liu, X., Fujimura, K.: Pedestrian detection and tracking with night vision. IEEE Transactions on Intelligent Transportation Systems 6(1), 63–71 (2005)
21. Yilmaz, A., Shafique, K., Shah, M.: Target tracking in airborne forward looking infrared imagery. Image and Vision Computing 21(7), 623–635 (2003)