Robust People Segmentation by Static Infrared Surveillance Camera

José Carlos Castillo, Juan Serrano-Cuerda, and Antonio Fernández-Caballero

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. In this paper, a new approach to real-time people segmentation through processing images captured by an infrared camera is introduced. The approach starts detecting human candidate blobs processed through traditional image thresholding techniques. Afterwards, the blobs are refined with the objective of validating the content of each blob. The question to be solved is if each blob contains one single human candidate or more than one. If the blob contains more than one possible human, the blob is divided to fit each new candidate in height and width.

1 Introduction

In the surveillance field [11], [8], [9] the use of infrared cameras are being intensively studied in the last decades. 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 [13]. 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 [6], [2], a thresholded thermal image forms the first stage of processing after which methods for pose estimation and gait analysis are explored. In [10], a simple intensity threshold is employed and followed by a probabilistic template. A similar approach using Support Vector Machines is reported in [12]. Recently, a new background-subtraction technique to robustly extract foreground objects in thermal video under different environmental conditions has been presented [3]. A recent paper [7] presents a real-time egomotion estimation scheme that is specifically designed for measuring vehicle motion from a monocular infrared 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 [5].

In this paper, we introduce our approach to real-time robust people segmentation through processing video images captured by an infrared camera.

N. García-Pedrajas et al. (Eds.): IEA/AIE 2010, Part I, LNAI 6096, pp. 348–357, 2010.

2 Robust People Segmentation Algorithm

The proposed human detection algorithm is explained in detail in the following sections related to the different phases, namely, people candidate blobs detection, people candidate blobs refinement and people confirmation.

2.1 People Candidate Blobs Detection

The algorithm starts with the analysis of input image, $I(x, y)$, captured at time t by an infrared camera, as shown in Fig. 1a. 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 $I(x, y)$ to $I^1(x, y)$ (see Fig. 1b). The normalization assumes a factor γ, calculated as the mean gray level value of the las n input image, and uses the mean gray level value of the current image, $\overline{I}$.

$$I^1(x, y) = \frac{I(x, y) \times \gamma}{\overline{I}} \tag{1}$$

where $I^1(x, y)$ is the normalized image. Notice that $I^1(x, y) = I(x, y)$ when $\overline{I} = \gamma$.

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

$$\theta_i = 3 \times \frac{5}{4}\gamma \tag{2}$$

$\delta = \frac{5}{4}\gamma$ 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.

$$I^1(x, y) = \begin{cases} I^1(x, y), & \text{if } I^1(x, y) \leq \theta_i \\ \overline{I^1}, & \text{otherwise} \end{cases} \tag{3}$$

The algorithm uses a threshold to perform a binarization for the aim of isolating the human candidates spots. The threshold θ_c, 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 threshold is calculated as:

$$\theta_c = \frac{5}{4}(\gamma + \sigma_{I^1}) \tag{4}$$

where σ_{I^1} is the standard deviation of image $I^1(x, y)$. 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.

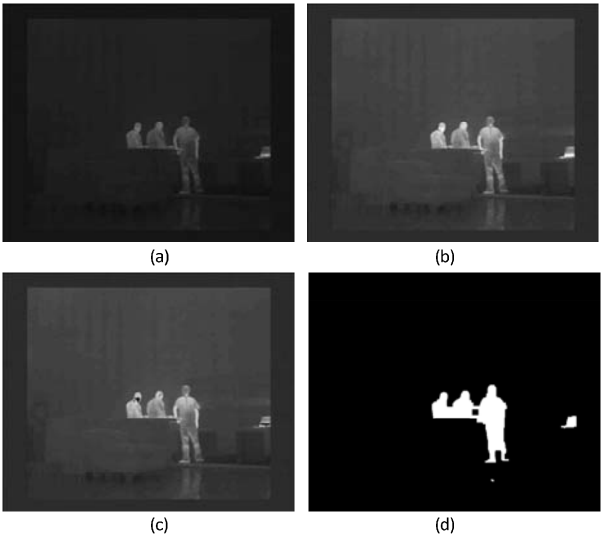

Fig. 1. (a) Input infrared image. (b) Scaled frame. (c) Incandescence elimination. (d) Thresholded frame.

Now, image $I^1(x,y)$ is binarized using the obtained threshold θ_c. Pixels above the threshold are set as maximum value $max = 255$ and pixels below are set as minimum value $min = 0$.

$$I_b^1(x,y) = \begin{cases} min, \text{ if } I^1(x,y) \leq \theta_c \\ max, \text{ otherwise} \end{cases} \tag{5}$$

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. 1.

$$I_o^1(x,y) = I_b^1(x,y) \circ \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{6}$$

$$I_c^1(x,y) = I_o^1(x,y) \bullet \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{7}$$

Afterwards, the blobs contained in the image are obtained. A minimum area, A_{min}, - function of the image size - is established for a blob to be considered to contain humans.

$$A_{min} = 0.0025 \times (r \times c) \tag{8}$$

where r and c are the number of rows and columns, respectively of input image $I(x, y)$. As a result, the list of blobs, L_B, containing people candidates in form of blobs $b_\lambda[(x_{start}, y_{start}), (x_{end}, y_{end})]$, is generated. λ stands for the number of the people candidate blob in image $I^1(x, y)$, whereas (x_{start}, y_{start}) and (x_{end}, y_{end}) are the upper left and lower right coordinates, respectively, of the minimum rectangle containing the blob. As an example, consider the resulting list of blobs related to Fig. 1 and offered in Table 1.

Table 1. People candidates blobs list

λ	x_{start}	y_{start}	x_{end}	y_{end}	area
1	297	270	482	458	35154
2	608	344	645	376	1254

2.2 People Candidate Blobs Refinement

In this part, the algorithm works with the list of blobs L_B, present in image $I^1(x, y)$, obtained at the very beginning of the previous section. At this point, there is a need to validate the content of each blob to find out if it contains one single human candidate or more than one. Therefore, the algorithm processes each detected blob separately.

Let us define a region of interest (ROI) as the minimum rectangle containing one blob of list L_B. A ROI may be defined as $R_\lambda = R_\lambda(i, j)$, when associated to blob $b_\lambda[(x_{start}, y_{start}), (x_{end}, y_{end})]$. Notice that $i \in [1..max_i = x_{end} - x_{start} + 1]$ and $j \in [1..max_j = y_{end} - y_{start} + 1]$.

People vertical delimiting. The first step consists in scanning R_λ by columns, adding the gray level value corresponding to each column pixel, as shown in equation (9).

$$H_\lambda[i] = \sum_{j=1}^{max_j} R_\lambda(i, j), \forall i \in [1..max_i] \tag{9}$$

This way, a histogram $H_\lambda[i]$ showing which zones of the ROI own greater heat concentrations is obtained. A double purpose is pursued when computing the histogram. In first place, we want to increase the certainty of the presence and situation of human heads. Secondly, as a ROI may contain several persons that are close enough to each other, the histogram helps separating human groups (if any) into single humans. This method, when looking for maximums and minimums within the histogram allows differentiating among the humans present in the particular ROI.

Now the histogram, $H_\lambda[i]$, is scanned to separate grouped humans, if any. For this purpose, local maxima and local minima are searched in the histogram to establish the different heat sources (see Fig. 2a). To assess whether a histogram column contains a local maximum or minimum, a couple of thresholds

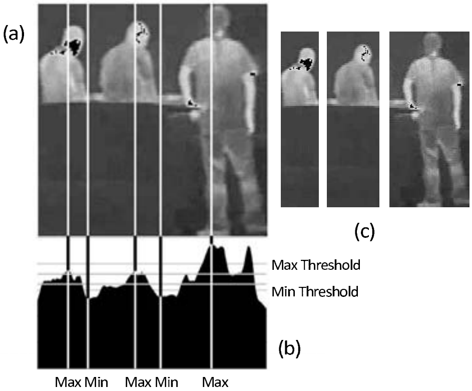

Fig. 2. (a) Input ROI. (b) Histogram. (c) Columns adjustment to obtain three human candidates.

are fixed, $\theta_{v_{max}}$ and $\theta_{v_{min}}$. Experimentally, we went to the conclusion that the best thresholds should be calculated as:

$$\theta_{v_{max}} = 2 \times \overline{R_\lambda} + \sigma_{R_\lambda} \tag{10}$$

$$\theta_{v_{min}} = 0.9 \times \overline{R_\lambda} \times max_j \tag{11}$$

Each different region that surpasses $\theta_{v_{max}}$ is supposed to contain one single human head, as heads are normally warmer than the rest of the people body covered by clothes. That is why $\theta_{v_{max}}$ has been set to the double of the sum of the average gray level plus the standard deviation of the ROI. On the other hand, $\theta_{v_{min}}$ indicates those regions of the ROI where the sum of the heat sources are really low. These regions are supposed to belong to gaps between two humans. We are looking for regions where the column summed gray level is below a 90% of the mean ROI gray level value. Fig. 2b shows the histogram for input ROI of Fig. 2a. You may observe the values for $\theta_{v_{max}}$ and $\theta_{v_{min}}$, corresponding to three peaks (three heads) and two valleys (two separation zones). Fig. 2c shows the three humans as separated by the algorithm into sub-ROIs, $sR_{\lambda,\alpha}$.

People horizontal delimiting. All humans contained in a sub-ROI, $sR_{\lambda,\alpha}$, obtained in the previous section possess the same height, namely the height of the original ROI. Now, we want to fit the height of each sub-ROI to the real height of the human contained. Rows adjustment is performed for each new sub-ROI, $sR_{\lambda,\alpha}$, generated by the previous columns adjustment, by applying a new threshold, θ_h.

The calculation is applied separately on each sub-ROI to avoid the influence of the rest of the image on the result. This threshold takes the value of the sub-ROI average gray level, $\theta_h = \overline{sR_{\lambda,\alpha}}$. Thus, sub-ROI $sR_{b,\lambda,\alpha}$ is binarized in order to delimit its upper and lower limits, obtaining $sR_{\lambda,\alpha}$, as shown in equation (12) similar to equation (5).

$$sR_{b,\lambda,\alpha}(i,j) = \begin{cases} min, \text{ if } sR_{\lambda,\alpha}(i,j) \leq \theta_h \\ max, \text{ otherwise} \end{cases} \tag{12}$$

After this, a closing is performed to unite spots isolated in the binarization, getting $sR_{c,\lambda,\alpha}$ (see Fig. 3b).

$$sR_{c,\lambda,\alpha}(i,j) = sR_{b,\lambda,\alpha}(i,j) \bullet \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{13}$$

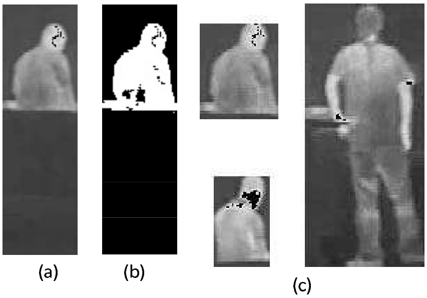

Fig. 3. (a) Input sub-ROI. (b) Binarized sub-ROI. (b) Rows adjustment to delimit three human candidates.

Next, $sR_{c,\lambda,\alpha}$ is scanned, searching pixels with values superior to *min*. The upper and lower rows of the human are equal to the first and last rows, respectively, containing pixels with a value set to *max*. The final result, assigned to new ROIs, $\Re_\kappa$, may be observed in Fig. 3c. The blobs associated to the split ROIs are enlisted into the original blobs list, L_B (see Table 2).

Table 2. Refined people candidates blobs list

κ	x_{start}	y_{start}	x_{end}	y_{end}	area
1	339	286	395	354	3808
2	396	270	481	458	15980
3	298	289	338	354	2600

3 Data and Results

We have tested our proposal with a well-known indoor infrared video benchmark, namely the “Indoor Hallway Motion” sequence included in dataset 5 “Terravic Motion IR Database” provided within the OTCBVS Benchmark Dataset Collection [4].

Firstly, in Fig. 4 the results of applying the algorithm to a sequence where a person is crossing the scene form right to left. Fig. 4a shows the scenario. Notice the presence of a hot spot in the image that could lead to confusion, but the algorithm does not make any mistake. Fig. 4b shows the first frame where the person is detected. The detection works fine - a hit of a 100% - through all the frames (see Fig. 4c) until the person reaches the hot spot region. Here (Fig. 4d), the human region also covers the hot spot. From this moment on, and until the person exits the scene, the performance of the person segmentation hit is a 98% (e.g. Fig. 4e and f).

The second example shows another person crossing from right to left, while, at the same time, a second human is slowly approaching the camera from the rear. As you may observe (Fig. 5a to c), the first human is perfectly segmented. The person approaching is wearing clothes that do not emit any heat. Thus, only the head is visible, and it is very difficult to detect such a spot as belonging to a human. Nonetheless, in a 63% of the frames where the partially occluded human is present, he is detected (see Fig. 5d and e). Unfortunately, when this person is too close to the camera, it is not detected as a human (see Fig. 5f).

The third example shows a human entering the scene from the front right and walking to the rear left (see Fig. 6. Now, this person shows an abnormal and uniform high heat level. This is a challenge for our proposal, as we try to locate the human head in relation with the body (in terms of heat and size). The performance of the algorithm in this sequence is about a 60% of hits when the person is close to the camera (see Fig. 6a to c) and grows to a 90% when the human goes far, as shown in Fig. 6e and f. Fig. 6d shows one of the three frames where a false positive is gotten in a total of 140 frames of the sequence.

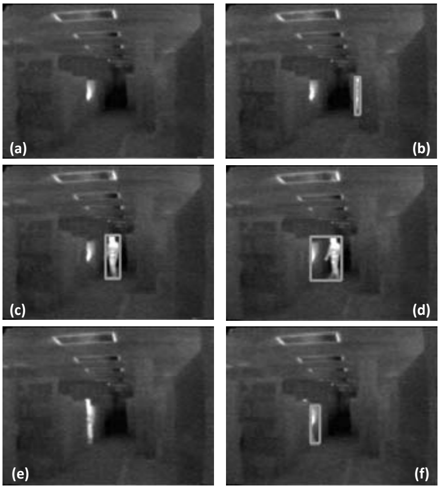

Fig. 4. Crossing from right to left. (a) Frame #220. (b) Frame #250. (c) Frame #290. (d) Frame #292. (e) Frame #324. (f) Frame #333.

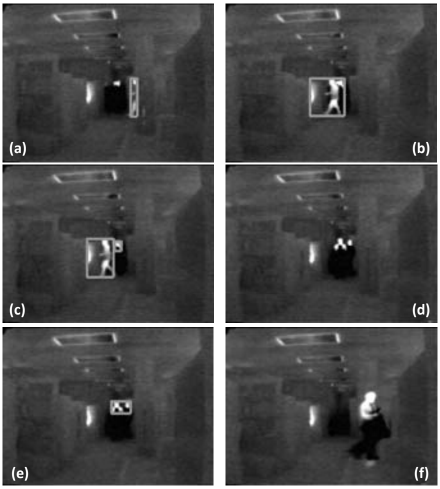

Fig. 5. Crossing from right to left, and approaching from rear to front. (a) Frame #2326. (b) Frame #2357. (c) Frame #2362. (d) Frame #2420. (e) Frame #2411. (f) Frame #2530.

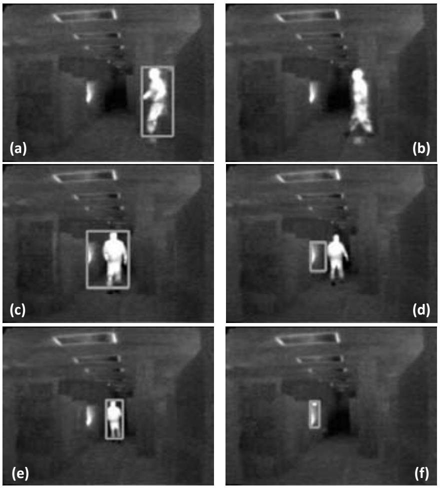

Fig. 6. Walking from front right to rear left. (a) Frame #5135. (b) Frame #5150. (c) Frame #5175. (d) Frame #5209. (e) Frame #5229. (f) Frame #5264.

Finally, we show in Fig. 7 a sequence of two people entering the scene from different sides, and walking together to the rear. As shown in Fig. 7a to c, the segmentation process is efficient. From frame #8344 on, corresponding to Fig. 7d, the humans are not correctly divided, as they are too close to the rear. As shown in Fig. 7e and f, until the humans disappear from the scene, the segmentation of the group is right for a 47%.

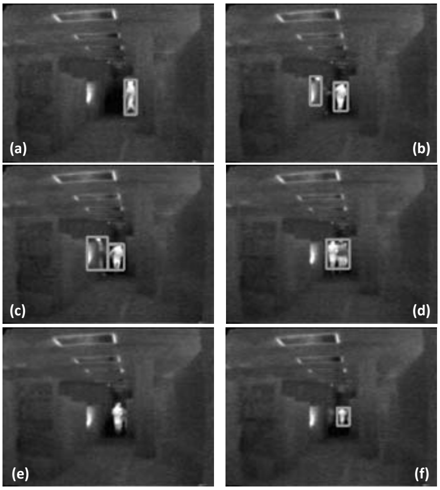

Fig. 7. Two persons walking to the rear. (a) Frame #8268. (b) Frame #8293. (c) Frame #8297. (d) Frame #8344. (e) Frame #8377. (f) Frame #8478.

4 Conclusions

A new approach to real-time people segmentation through processing images captured by an infrared camera has been extensively described. The proposed algorithm starts detecting human candidate blobs processed through traditional image thresholding techniques. Afterwards, the blobs are refined with the objective of solving the question if each blob contains one single human candidate or has to be divided into smaller blobs. If the blob contains more than one possible human, the blob is divided to fit each new candidate in height and width. The results obtained so far in indoor scenarios are promising. We have been able of testing our person segmentation algorithms on a well-known test bed.

Acknowledgements

This work was partially supported by the Spanish Ministerio de Ciencia e Innovación under project TIN2007-67586-C02-02, and by the Spanish Junta de Comunidades de Castilla-La Mancha under projects PII2I09-0069-0994 and PEII09-0054-9581.

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. Davis, J.W., Sharma, V.: Background-subtraction in thermal imagery using contour saliency. International Journal of Computer Vision 71(2), 161–181 (2007)
4. Davis, J.W., Keck, M.A.: A two-stage template approach to person detection in thermal imagery. In: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision, vol. 1, pp. 364–369 (2005)
5. 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)
6. 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)
7. 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)
8. 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)
9. 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)
10. 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)
11. 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)
12. Xu, F., Liu, X., Fujimura, K.: Pedestrian detection and tracking with night vision. IEEE Transactions on Intelligent Transportation Systems 6(1), 63–71 (2005)
13. Yilmaz, A., Shafique, K., Shah, M.: Target tracking in airborne forward looking infrared imagery. Image and Vision Computing 21(7), 623–635 (2003)