Pedestrian Detection and Tracking Based on Far Infrared Visual Information

Daniel Olmeda, Cristina Hilario, Arturo de la Escalera, and Jose M. Armingol

Universidad Carlos III de Madrid
Intelligent Systems Laboratory, Department of Systems Engineering and Automation
C./ Butarque 15, 28911 Leganes, Spain
armingol@ing.uc3m.es
http://www.uc3m.es/islab

Abstract. This article discusses a pedestrian detector for an experimental vehicle, based on visual information from a single far infrared camera. The system considers three consecutive processes for each image. Once the candidates' heads have been extracted, regions of interest are resized based on the distance to the camera and then filtered by its vertical edges symmetry. Once the bounding boxes of possible pedestrians are picked a spatial correlation with some template models takes place. Finally, detected pedestrians are tracked, contrasting their position on successive frames. The results are satisfactory, classifying correctly almost 96% of pedestrians closer than 45m to the vehicle.

1 Introduction

Given today's rate of growth of the global automovilistic sector it is expected that, in a few decades time, the traffic engineering, as we understand it, will become obsolete. Certainly, road infrastructures will not be able to keep up with the needs of the increasing traffic. The most likely solution will be an optimization of the road flows in which, progressively, more and more task will be reassigned from the driver to the vehicles. However, until the time when fully automatic driving becomes a reality, it is necessary the development and implementation of driving assistance technologies. On those, the driver is the one who assumes most of the tasks and responsibilities. These systems provide information on real-time to the driver so that most of the cognitive limitations or eventual distractions can be diminished. In this context, a pedestrian detector system for low visibility conditions based on far infrared vision has been developed.

On the driving scene, pedestrian are comprehensibly the most fragile element. On more than an 80% of the accidents involving a motor vehicle, limited or total liability can be assigned to the driver. The perception and reaction delay time against unusual events on the road is one of the most important considerations in the attempt of reducing casualties. This is even more significant on drivers that are over a certain age. A 50 year old driver needs twice as much light as a 30 year old [3]. The final goal of the system, whose description follows, is therefore,

J. Blanc-Talon et al. (Eds.): ACIVS 2008, LNCS 5259, pp. 958–969, 2008.

to communicate the driver all possible pedestrians whose trajectory is expected to cross imminently the vehicles', so that a collision can be avoided.

The proposed system is based on monocular far infrared (FIR) vision. In night driving conditions, in which pedestrian corporal temperature fairly exceed that of his vicinity, far infrared vision allows the segmentation of the acquired image, based on its gray level value, discriminating between pedestrian and background. Detected warm objects are scanned, searching for a certain pattern, and finally the expected trajectory is calculated. In the event of an intersection between this trajectory and the path of the vehicle, an acoustic warning will inform the driver about the direction on which the detected pedestrian is approaching.

1.1 The IVVI Project

The presented algorithm has been implemented on the IVVI experimental vehicle (Intelligent Vehicle based on Visual Information) [4]. IVVI is an investigation platform whose principal aim is the development of an Advanced Driver Assistance System (see figure 1). It comprises four modules:

1. *Anti-collission*: Informs the driver about pedestrian or other vehicles on the road.
2. *Speed supervision*: A warning signal is triggered when the driving speed excess the limit established for that kind of road. To achieve this, a traffic signals recognition phase is included as part of the module.
3. *Overtaking assistant*: The blind spot is scanned for possible cars in the adjacent lane, and the left line is checked to be discontinuous.
4. *Somnolence awareness*: Driver's behaviors are watched searching for symptons of lessened attention.

Fig. 1. (top) IVVI: Intelligent Vehicle based on Visual Information. ; (bottom-left) Vehicle's screen; (bottom-right) Infrared camera.

This paper is organized as follows. Section 2 describes related work on computer vision, focussing on pedestrian detection and its latest progress. The presented system is introduced in section 3 and detailed in section 4. The obtained results are presented in section 5. Finally, in section 6 conclusions and future work are discussed.

2 State of the Art

Pedestrian detection on experimental vehicles has seen some significant development in the last few years [1]. Computer vision is widely used for pedestrian detection given that other approaches (e.g. radar) can be unreliable discriminating between similar solid object by their own means. However it is quite common to integrate radar systems on vision pedestrian detection frameworks as their distance measures are highly accurate. Most of the researches on real time vision detection focus on visible light range images, both monocular and stereo configurations [4], [5], [6]. As for night detection, the most used method is far infrared (FIR) vision [2], [3], [7], [8].

Many developed systems involve a pedestrian extraction phase, distinguishing it from its background, followed by a validation step. The extraction is based, among other methods, on movement detection by means of background substraction, on distances in stereo vision systems, or on object features such as brightness or textures. On vehicle based frameworks, in which movement detection is unreliable, the most commonly used algorithms are based on search for warm areas or borders.

The most important objective of onboard pedestrian detection is to avoid possible collisions. Thereby, on top of detection, many systems perform also a tracking of the regions of interest (ROI). From this tracking is derived a prediction of the future behavior of the pedestrian. This phase usually includes a Kalman filter. This filter provides great stability to the system, a good performance and is relatively easy to implement. In other systems, the tracking is based on the spatial superpose of the regions of interest on successive frames [13]. This procedure is often used on algorithms in which processing speed of the image is a crucial factor, and the error confidence interval doesn't need to be too narrow. Some algorithms performs a pedestrian search only each certain number of frames, reducing significantly the computational cost, and follow the pedestrian with a Kalman filter [8], [9].

The use of far infrared cameras, besides all its advantages, is usually unable to cope with every scenario. The results that can be obtained with a relatively high external temperature makes the use of these cameras impracticable in these situations. However, its use is entirely justified in night driving or in those situations in which the visibility is very limited by fog or rain. The results obtained are very promising and give reasons for keeping on with the investigations.

The tendency is to integrate infrared vision cameras with other sensors (e.g. radar, visible light images) in a system that decides based on the information of them all [7].

3 System Description

The purpose of using infrared images on pedestrian detection is that of coping with circumstances on which daylight is not enough to use normal vision cameras. Infrared images' most distinctive characteristic is to display the amount of heat that an object emits. On the histogram, higher values correlate with pixels of the image belonging to a warm object, and cold objects will appear on the image as darker pixels.

Unlike many other solutions proposed, the presented system acquires data by means of a firewire camera with an image depth of 14 bits. CCIR or RS170 cameras apply a predetermined gain curve to the images. The result is that, though those images seem more intuitive to the human eye, usually misrepresent the temperature of the scene objects. The use of the sensors raw data allows the algorithm to select a more accurate threshold to discriminate between warm objects and background.

The used camera operates in the far infrared interval, sensible to wavelengths between $7.5\mu m$ and $13.5\mu m$. For this region of the frequency spectrum, the obtained images depict the thermal radiations of the objects in scene. A patent disadvantage of infrared images, when compared with visible light ones, is its narrow divergence of intensities. These images are blurry. Therefore, the search of pedestrian can't be based on distinguishable feature points or textures, as these appear in an aleatoric manner. Instead, the presence of a pedestrian is validated by a shape contour detection. A person can assume very different appearances, depending on the action taking place. For example, someone sitting or laying on the ground has an outline quite different as that of another standing still. To cope with the wider possible sort of body shapes, four different models has been developed. Each one takes in account a distinguishable position of arms and legs of a walking pedestrian, such being the most common way in which those are presented on the acquired sequences.

Those models are correlated with the infrared images. However, considering that the size of the possible pedestrian is *a priori* unknown, a multi-resolution search would be prohibitive for a real-time algorithm. Consequently it is necessary to define regions of interest that have high prospects of containing pedestrians before the correlation step. The ROIs are determined depending on the gray-level intensities of the image, assuming that high intensities values match hot objects.

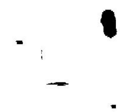
An additional search step scan the image looking for symmetry of vertical edges.

4 Algorithm

4.1 Search of Warm Areas

As noted previously, warm areas are segmented thresholding the image. Since the acquired 14 bit image represents the scene's temperature with an almost linear function proportional to the sensor temperature, a threshold function obtained

(a) Image with improved contrast.



(b) Hotspots in the image.

(c) Head-only model.

Fig. 2. Search of pedestrians heads

by direct experimentation is applied to extract objects with a temperature close to that of the human body. The resulting blobs usually represent the head or arms of the present pedestrians. The blobs' bounding boxes are resized and compared with a head-only model by means of gray-scale correlation (see fig. 2) as detailed in section 4.2.

Simetry of vertical edges. Warm objects edges on infrared images are well defined as they are highly contrasted with their background. Besides, it is known that positive and negative edges of pedestrians will be symmetrical around an axis. (See figure 3).

Once the pedestrians' heads are found, the algorithm sets the width of the image column that is more likely to contain the pedestrian. This decision is based on the vertical symmetry of increasingly wider columns whose axis is to be located around the center of the pedestrians head. This search is restrained to columns just wide enough to contain the pedestrian based on the height of the head above the ground. This width e shall maximize equation 1, where a is the vertical axis of the blob identified as a head.

$$S(e) = \sum_{w=0}^{N} (P_{a-w} \cdot N_{a+w}) \tag{1}$$

Edges are calculated with the finite impulse response filter proposed by Prewitt, for both positive (P) and negative (N) ones.

Selection of candidate areas. An adaptative linear gain curve is applied to areas of the image with a high vertical symmetry. Then those regions are

(a) Positive edges. (b) Negative edges. (c) Symmetry axis.

Fig. 3. Vertical edges symmetry

thresholded by means of the Otsu method [11]. The purpose of this is to extract the silhouette of the warm objects detected.

Huge or tiny blobs are excluded out of the thresholded image as they cannot contain a pedestrian. Warmer body parts are head, hands and legs. Chest and ankles are usually cooler, appearing in the image a darker pixels, and as a result can be misclassified as background. So it is possible that different blobs belong to the same pedestrian. Considering the shape of a pedestrian, every possible blob in which it can be divided must be in a limited width column. The horizontal projection of the thresholded areas as shown in figure 4).

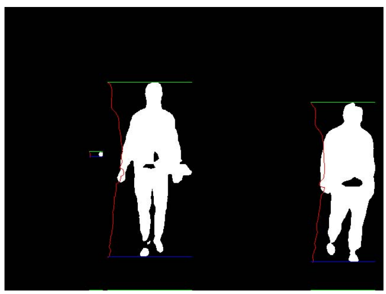


Fig. 4. Horizontal projection of warm symmetric objects

Obviously, some blobs may not belong to a certain pedestrian, although they are located on the same column. That's why a further filtering is necessary. Those blobs that, having an appropriate size, are further than a certain distance to the group's centroid are also deleted. Adequate blobs not affected by the filter are confined in a rectangular box, whose size is proportional to the distance of the candidate to the camera. Since the system is based on a monocular solution, whose intrinsic parameters are known distance is determined upon the ground contact point of the pedestrian, using the pin-hole simplification.

4.2 Correlation

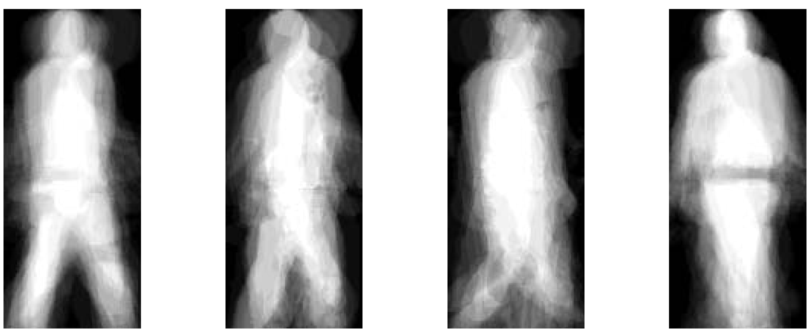
Model's creation. Correlation as a matching indicator is widely used because of its simplicity and good results on certain kinds of images. However, its main drawback is its high computational demand. Once bounding boxes are determined, the computational cost of a correlation is perceptibly reduced. Thus justifying its use.

The models are created starting with the same process of warm areas extraction described in 4.1. Out the bounding boxes obtained from several sequences 240 have been selected as they include pedestrians in fairly normal attitudes. Those selected boxes are divided in four groups attending to arms and legs position. Only walking or running pedestrians are considered, given that other atypical poses (such as a sit-down position) are very difficult to tell apart from other inanimate objects using infrared vision.

The model's template image is obtained calculating the average of each image of the group of ROIs selected. Each pixel's intensity value of the resulting model would then be proportional to it's probability of belonging to a pedestrian. That probability is determined with equation 2.

$$p(x,y) = \frac{\sum_{i=0}^{N} s(x,y)}{255 \cdot N} \tag{2}$$

In equation 2, $s(x,y)$ is the gray level of the i pixel of the selected ROIs, and $p(x,y)$ the probability of the model's pixels of belonging to a pedestrian. The gray-level value of them would be, then $M(x,y) = p(x,y) * 255$. The resulting images are depicted in figure 5.



(a) Opened Legs. (b) Semi-opened legs (1). (c) Semi-opened legs (2). (d) Closed legs.

Fig. 5. Examples of pedestrian models

However, this approach have been proved to be unreliable to correctly classify pedestrian that are too far away from the camera, as their silhouettes are quite different from those close to it. To cope with this four sets of models have been developed at multiple resolutions, as indicated in table 1.

Correlation methods. Normalized correlation of gray scale images as a pattern matching method can be considered as a convolution, in which the model

Table 1. Models resolution for each distance range

Distance(m)	Resolution
$[5-15]$	210×90
$[15-25]$	126×54
$[25-40]$	79×34
$[40-\infty]$	40×17

is the kernel. Thus, as said before, computational cost is its main disadvantage. Moreover, it has to be repeated in multiple scales, as pedestrians can be both far and close to the camera. A multi-resolution search means a quite slow overall algorithm, unacceptable on real time applications. However, after the warm area search the number of calculations are reduced to one per ROI and model.

Several correlations formulae have been tested. The normalized gray-scale correlation equation normalizes its result between -1 and 1, being 1 a total match, and scores below 0 a lower likeliness than that expected by chance.

The obtained results with normalized correlation have been positive, however, the computational cost is still too high. Equation 3 takes significantly less time to process, with similar results [2], [7].

$$c = \frac{\sum_{i=0}^{N} \left[(p_i - 0.5)\,(M_i - 0.5) \right]}{\sum_{i=1}^{N} |p_i - 0.5|} \tag{3}$$

In equation 3, p is the value of intensity of each pixel of the thresholded ROI, being 1 if it belongs to the subject and 0 if it belongs to the background. M is the value of the intensities of the model pixels normalized between 0 and 1. The result is the correlation value of the positive pixels of the image. Because the background doesn't make any contribution to equation 3 it is necessary to calculate the deviation of dark pixels of the model with the black from the background of the ROI, using the same equation. The average of these two results is the value of the correlation. Each region of interest is matched against the four models, being the confidence of containing a pedestrian the higher of the four correlations. Only ROIs with a certainty of being pedestrians above a certain threshold are considered from now on. The result is a list of pedestrians that feed the tracking algorithm.

4.3 Labeling and Tracking

The pedestrian detection modules of driving assistance systems are intended to anticipate and avoid a potential collision between them and the vehicle. Then, besides the location of the pedestrian, it is necessary to know its trajectory and speed. Any pedestrian trajectory expected to cross that of the vehicle triggers the warning alarm system. The alarm notifies the driver the direction where the pedestrian is coming from. There are two important considerations to be contemplated in order to optimize the algorithm: how much precision is needed calculating the trajectory, and the algorithm speed. The relation between them is inverse. Besides, there is an additional trouble following objects in infrared images. That is, the scarce of unique feature points.

The validation module feeds the tracker with a list of detected pedestrians. On this step, pedestrians found on an image (B_i^t) are compared with those of the previous frame ($B_{L_i}^{t-1}$), where L is the label assigned to that tracked pedestrian. In case two bounding boxes are superposed on two consecutive frames, the new ROI is assigned the pedestrian label if it also adhere to certain features, such as size and spatial position, that minimizes equation 4.

$$C_{L_i} = \rho \left\| \, (X_L^{t-1} + V_L^{t-1}) - X_i^t \right\| + (1-\rho) \left| (\nabla S_L^{t-1} + S_L^{t-1} - S_i^t) \right| \tag{4}$$

In equation 4, ρ is the importance granted to the speed over the size. The cost is calculated for $1 \leq i \leq N$, being N the number of pedestrians found. The lower cost over a given threshold indicates the pedestrian label.

With each frame acquisition the labeled pedestrian information is updated (size, centroid coordinates, and edges of the bounding box). With this information the short-term future behavior of tracked pedestrian is extrapolated, that is, the variation on size (∇S_L), speed (∇V_L)and the lineal equation of the centroid trajectory (eq. 5). Taking into account a number n of consecutive frames, the centroid coordinates of the pedestrian can be calculated for an fr number of frames in the future.

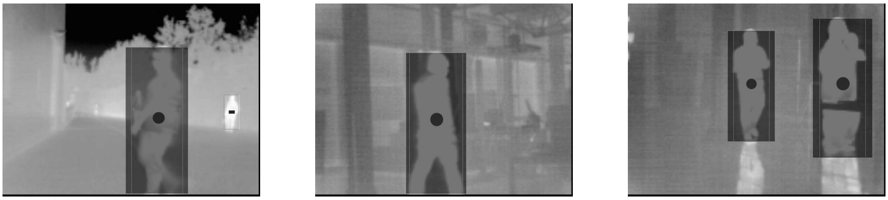
$$P_f = (Vx{\cdot}fr, Vy{\cdot}fr, \nabla S{\cdot}fr) \ni \begin{cases} V_x^t = (V_x^{t-1} + (X_c^t - X_c^{t-n}))/2 \\ V_y^t = (V_y^{t-1} + (Y_c^t - Y_c^{t-n}))/2 \\ \nabla S^t = \left(\nabla S^{t-1} + f \cdot L\left(\frac{1}{d_t} - \frac{1}{d_{t-1}} \right) \right)/2 \end{cases} \tag{5}$$

This tracking method is able to cope with detection failures or occlusions. If a pedestrian is not found where the algorithm expects it to be, it is labeled as lost and updated each frame with the accumulated cinematic information for a limited number of frames. When the detection module finds it again, it is assigned the same label that it had before the occlusion.

5 Results

The presented pedestrian detector system is part of the IVVI experimental vehicle. The ilumination and temperature circumstances on which those sequences were recorded range from night (5°C) to sunny midday (17°C).

The IVVI is equipped with two PC platform computers, each with a Pentium D 2 Ghz processor. The Indigo far infrared camera has been configured to output data via a firewire connection. The average processing speed with images that have, at least, one pedestrian has been $26fps$. If no warm object is present, processing times have always been below $5ms$. As of the current development state

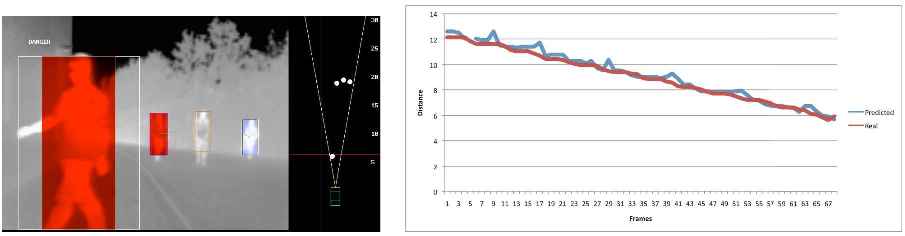


(a) Closest pedestrian inside vehicle's trajectory. The other one is out of it.

(b) A single person crossing in front of the vehicle.

(c) Pedestrian and group of pedestrians standing still ahead the vehicle.

Fig. 6. Boxes with a circle contains pedestrians in a dangerous spot. Boxes with a rectangle do not represent danger.



(a) Detected pedestrians and bird-eye view.

(b) Real pedestrian distance to the camera and its prediction.

Fig. 7. Distance to the camera prediction

the results have been satisfactory, classifying correctly almost 96% of bounding boxes closer than $45m$ to the vehicle. Further objects have a very low resolution, thus having a failure rate much higher. It is, therefore, necessary to develop a new parallel algorithm that can handle such long distances instead of adapting this one. However, driving through urban locations, and below $50km/h$ such a distance is more than enough to avoid the most dangerous situations. The algorithm have also been proven to be very solid against misdetections.

Figure 6 depicts different processed images. Those pedestrians found to be in a risky place or those whose trajectory intersect that of the vehicle are tagged with a circle. Boxes marked with a rectangle are not in jeopardy (for example, those walking along the sidewalk).

The tracker module is able to follow and predict the position of multiple pedestrians in the near future. Figure 7(b) represents the predicted distance of the left-most pedestrian in figure 7(a) against its real position. This prediction is made 10 frames in advance.

Since detection is first based upon the search of the pedestrian head, it is able to discern between pedestrian grouped even if they partially overlap or touch each other.

Finally, the driver is warned of the presence of such obstacles through an audio alarm. For each pedestrian found a sound buffer is set so that it represents the position in 3D space, from the perspective of the driver.

6 Conclusion

In this paper, a pedestrian detection system in FIR images based on template matching has been presented. It detects pedestrian within a range of $1m$ to $45m$ in front of the vehicle, and predict a short-term lineal trajectory.

The results have been promising. However, new objectives have been considered to improve the algorithm. First, and because it is not immune to occlusions, partially covered pedestrians are not correctly classified. It is intended to develop a new set of models to overcome this limitation. A single model will be used to match the upper part of the body. If the correlation returns a positive result the lower part of the ROI would be matched against several leg models.

Secondly, driving at a relative high speed, sometimes two ROIs on consecutive frames containing the same pedestrian don't suffice equation 4. This is specially true when the vehicle is turning direction. A GPS and inertial module will be included to acquire yaw angle information, so that velocity coordinates can be correctly updated.

Finally, it is intended to merge this module with a pedestrian detector that exploits visible light images, so that each cancel the disadvantages of the other.

Acknowledgment

This research was partially funded by the Cicyt Project VISVIA TRA2007-67768-C02-02 and by the Comunidad de Madrid through the project CCG06-UC3M/DPI-0418.

References

1. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), 694–711 (2006)
2. Nanda, H., Davis, L.: Probabilistic Template Based Pedestrian Detection in Infrared Videos. In: Procs. IEEE Intelligent Vehicles Symposium 2002, Paris, France (June 2002)
3. Fang, Y., Yamada, K., Ninomiya, Y., Horn, B.K.P., Masaki, I.: A Shape-Independent-Method for Pedestrian Detection with Far-Infrared-Images. IEEE Transactions on Vehicular Technology 53(6) (November 2004)
4. Hilario, C., Collado, J.M., Armingol, J.M., de la Escalera, A.: Pedestrian Detection for Intelligent Vehicles based on Active Contour Models and Stereo Vision. In: 10th International Workshop on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain (February 2005)
5. Bu, F., Chan, C.-Y.: Pedestrian Detection in Transit Bus Application: Sensing Technologies and Safety Solutions. In: Intelligent Vehicles Symposium, June 13-15. IEEE, Los Alamitos (2007)

6. Broggi, A., Bertozzi, M., Chapuis, R., Fascioli, F.C.A., Tibaldi, A.: Pedestrian Localization and Tracking System with Kalman Filtering. In: Procs. IEEE Intelligent Vehicles Symposium 2004, Parma, Italy, June 2004, pp. 584–589 (2004)
7. Bertozzi, M., Hilario Gómez, C., Fedriga, R.I., Vezzoni, G., Del Rose, M.: Pedestrian Detection in Far Infrared Images based on the use of Probabilistic Templates. In: Intelligent Vehicles Symposium, June 13-15. IEEE, Los Alamitos (2007)
8. Binelli, E., Broggi, A., Fascioli, A., Ghidoni, S., Grisleri, P.: A modular tracking system for far infrared pedestrian recognition. In: Intelligent Vehicles Symposium, June 13-15. IEEE, Los Alamitos (2007)
9. Xu, F., Fujimura, K.: Pedestrian Detection and Tracking with Night Vision. In: Procs. IEEE Intelligent Vehicles Symposium 2002, Paris, France (June 2002)
10. Kittler, J., Illingoworth, J., Foglein, J.: Threshold Selection Based on a Simple Image Statistic. Computer Vision, Graphics, and Image Processing 30, 125–147 (1985)
11. Otsu, N.: A Threshold Selection Method from Gray-level Histograms. IEEE Transaction on Systems, Man and Cybernetics SMC-9,, 62–66 (1979)
12. Chang, C.-I., Du, Y., Wang, J., Guo, S.-M., Thouin, P.D.: Survey and comparative analysis of entropy and relative entropy thresholding techniques. IEE Proceedings Vision, Image and Signal Processing 153(6), 837–850 (2006)
13. Fuentes, L.M., Velastin, S.A.: People tracking in surveillance applications. In: Proceedings 2nd IEEE Int. Workshop on PETS, Kauai, Hawaii, USA, December 9 (2001)