Robotica (2011) volume 29, pp. 495–505. © Cambridge University Press 2010
doi:10.1017/S0263574710000299

Far infrared pedestrian detection and tracking for night driving

Daniel Olmeda, Arturo de la Escalera and José María Armingol

Intelligent Systems Laboratory, Department of Systems Engineering and Automation, Universidad Carlos III de Madrid, C./ Butarque 15, 28911 Leganes, Spain

(Received in Final Form: June 14, 2010. First published online: July 29, 2010)

SUMMARY

This paper presents a module for pedestrian detection from a moving vehicle in low-light conditions. The algorithm make use of a single far infrared camera based on a microbolometer. Images of the area ahead of the vehicle are analyzed to determine if any pedestrian might be in its trajectory. Detection is achieved by searching for distributions of temperatures in the scene similar to that of the human body. Those areas with an appropriate temperature, size, and position in the image are classified, by means of a correlation between them and some probabilistic models, which represents the average temperature of the different parts of the human body. Finally, those pedestrians found are tracked in a subsequent step, using an unscented Kalman filter. This final stage of the algorithm enables the algorithm to predict the trajectory of the pedestrian, in a way that does not depend on the movement of the camera. The aim of this system is to warn the vehicle's driver and reduce the reaction time in case an emergency break is necessary.

KEYWORDS: Computer vision; IR pedestrian detection; Driver Assistance System; Unscented Kalman filter; Intelligent Transportation Systems.

1. Introduction

The understanding of traffic is an important concern nowadays. There is wide agreement that traffic should evolve into something that can be sustained in the future. Two are the main concerns with the current traffic systems: economicwise and safetywise. Traffic accidents are one of the main causes of deaths and permanent physical disabilities in every country with an important presence of vehicles.[11] A reliable traffic infrastructure is also an important factor in economies. The constant growth of the number of vehicles is pushing the current roads their flow limit. From both points of view, the traffic architecture has to be improved. The scientific community is also participating from this interest, with very interesting ideas as where the future of traffic will be. A relatively new knowledge area, and one of the most actively developed, is the study of intelligent transportation systems (ITS). These systems focus both in traffic reliability and safety. The solution proposed for both is to take over time responsibilities from the human driver and relocate them to an automatic system. These systems can integrate the information of every vehicle on the road and synchronize their movements, obtaining a much more fluid traffic. And because these system can have a much wider sensorial information than a single person can, the risk of an accident can be decreased.

Advanced Driver Assistance Systems (ADAS) broaden the senses of a human driver with information that would not normally be available due to the position inside of the vehicle, as well as circumstances the driver fails to see. These systems are thought of as an intermediate solution in time between current infrastructure and a future fully automated traffic systems. In this article the authors present an ADAS module for automatic detection of pedestrian in urban environments and in low-light conditions. The system make use of a far infrared thermal camera to search for the heat that the pedestrians emit in conditions that, otherwise, would be unfit for exploiting a system based on visible light cameras, such as (and specially) night driving. Information about the position of pedestrians crossing the road is submitted to the driver for a decision about the vehicle's control. These kind of ADAS systems can also assist fully automated vehicles.

1.1. Research summary

We present a brief summary of the algorithm, which will be further explained in Sections 3 and 4.

Section 3 describes the detection and classification step of the algorithm. In the first place, warm regions with head like shapes are looked for in the image. Around positive matches a region of interest (ROI) is created. Inside those ROIs, objects with temperatures in a certain range are extracted and fed to the classifier. The presence of a pedestrian is confirmed matching the surviving ROIs with a set of probabilistic models. Section 4 describes the next step of the system: the tracking of those pedestrians by means of an unscented Kalman filter (UKF). The prediction of the filter is also fed to the classifier, in case the ROI extractor fails to detect a pedestrian that was already being followed.

An outline of the process is summarized in Fig. 1.

2. State of the Art

The control of a road vehicle is a complex task. The environment is not fully controlled, so there is always an unknown probability of encountering an obstacle. In this work the authors focus on the detection of a special kind of obstacles: pedestrians. In urban scenarios vehicles and pedestrians share the same ground, so there is higher

* Corresponding author. E-mail: dolmeda@ing.uc3m.es

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

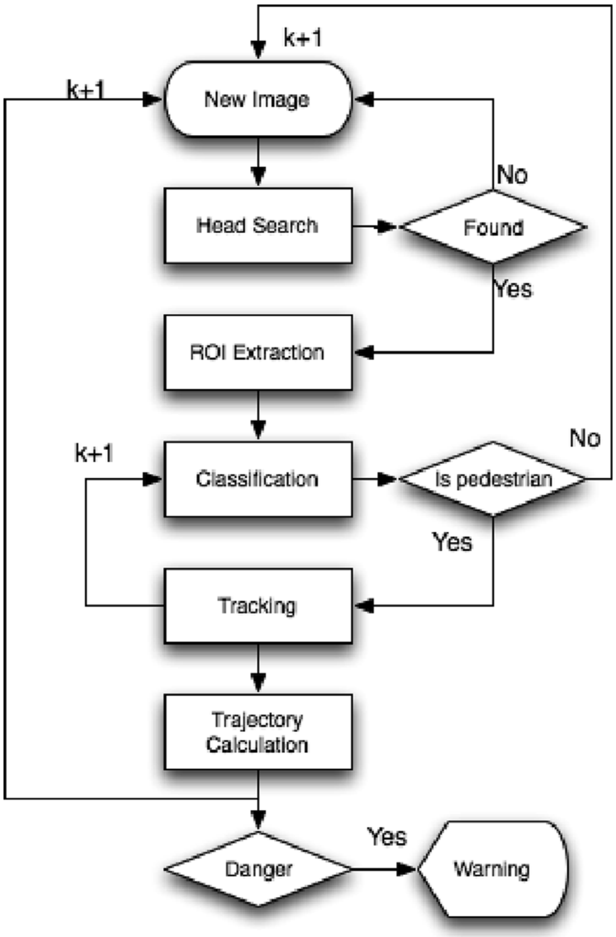


Fig. 1. Flow chart of the algorithm.

probability of a collision than in highway traffic. It is a particularly dangerous situation because pedestrians are much more likely to be hurt than the occupants of the vehicle, even at low speeds.

However, detection of pedestrians from a moving vehicle is not trivial, as they can appear with fairly different shapes, on a large part of the image, and in a random fashion. The use of computer vision to solve this situations is justified as other approaches, such as lidar scanners, although delivering very precise measurements of distance, does not provide enough information to discriminate between different types of obstacles. On top of that, vision is a non intrusive method. On the downside, the performance of a computer vision application is very dependent on the illumination conditions. There is a rich bibliography about pedestrian detection using cameras in the visible range light. As for night driving, there are two possibilities: to illuminate the scene with infrared leds and capture it with near infrared cameras,[6] or the use of thermal cameras that captures the emission of objects in the far infrared spectrum.

Far infrared images have a very valuable advantage over the visible light ones. They do not depend on the illumination of the scene. The output of those cameras is a projection on the sensor plane of the emissions of heat of the objects, that is proportional to the temperature. Most systems take advantage of this characteristic and select the regions of interest based on the distribution of warm areas on the image.[1,2,4,9] Another important feature is the intensity of the borders between pedestrians and their background. Those are used in systems that select regions of interest by the proximity of local shape features such as edgelets or histograms of oriented gradients.[3,8,14]

On systems that search for the temperature distribution, the discriminating feature of pedestrians would be the shape of the object. Regions of interest are correlated with some predefined probabilistic models.[2,9] In infrared images this approach is simple, yet robust.

Tracking can greatly simplify the task of pedestrian detection and cope with temporal occlusions or misdetections. It can also be used to predict trajectory and time left for collision between pedestrian and vehicle. Yet, this step is usually neglected in papers describing far infrared pedestrian detector. The most common solution is the use of kalman filtering to determine the pedestrian position.[13]

The use of far infrared cameras, besides all its advantages, is usually unable to cope with every scenario. Infrared cameras are unable to replace visible light cameras, as they present some disadvantages. As the outside temperature raises to high levels, the sensor's noise render the images useless for extracting distinctive features for pedestrian detection. Besides, direct sunlight, no matter what temperature, affects infrared images, as reflection on some surfaces make them appear hotter than they really are. The tendency is to integrate infrared vision cameras with other sensors (e.g. radar, visible light images) in a system that decides which information would be more useful in different circumstances.

3. Pedestrian Detection in Far Infrared Images

Pedestrian detection is achieved in this system by means of a far infrared camera. The sensor of these cameras represent in an image the radiance that the objects in the scene emits or reflects. The infrared radiance emitted from an object that hits the sensor depends on the external temperature of that object and its distance to the camera.

In this kind of application, pedestrians can be at any distance. Objects with a temperature close to the expected of a human body are searched for at different radiance scales.

Pedestrians in far infrared images presents some very distinctive characteristics, such as very pronounced external edges and a particular distribution of the body temperature. Usually the pedestrians head and legs are the parts of the body that emits more heat, being their apparent temperature barely lower than their real one. Warm areas of the image are extracted based on their apparent temperature, neglecting objects with temperatures that does not match those of the human body. Chest and arms are more often covered by thicker clothes, specially in winter, therefore their apparent temperature is usually only a little higher than that of the background. The border's definition is higher if the difference between the pedestrians and the backgrounds temperature is significant.

The application of a far infrared camera for pedestrian detection is intended to complement a visible light stereo system in situations in which its results degrade, such as under low illumination conditions. Driving at night,

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

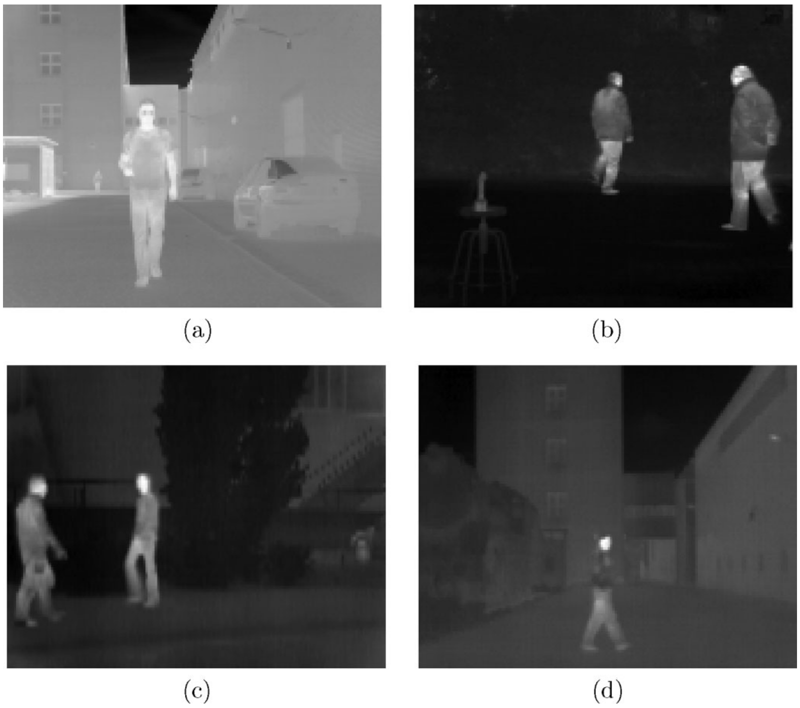


Fig. 2. Infrared images under different illumination and temperature conditions. (a) Reflection of direct sunlight against a brick wall (28°C); (b) cold winter scene (5°C); (c) scene under heavy clouds (12°C); (d) summer night scene (15°C).

pedestrians in images of a far infrared camera present very pronounced edges, and the distribution of their pixels intensities can easily be separated from that of the background.

Mobile vision applications, such as the ones part of a driver assistance system, have some intrinsic difficulties: the environment changes in almost each frame and in a way that can only be partially anticipated. The image can contain objects with a temperature higher than the human body, such a heated parts of a vehicle. Therefore, warm areas of the image are extracted based on their apparent temperature, neglecting objects with temperatures that does not match those of the human body.

3.1. Sensor calibration

Since the system only looks for pedestrians, the sensor have been calibrated focusing on a good detection of the lower and upper temperatures of the human body, and also for the average temperature of the head. These calibrated curves are obtained for each resolution used.

The gray level of each pixel of infrared images represents the amount of heat that the sensor captures. The camera used is based on the non-refrigerated microbolometer and, as such, its sensibility to external radiancies changes in a way that is function of the flux of radiance coming from inside the camera, as a result of its temperature. That sensibility is function of the sensor's and the object's temperatures. As such, the practical range of operation is limited in these kind of cameras. A significant raise in temperature would result in large errors in the calibrated curves. In that case, a recalibration would be necessary. The algorithm has been tested successfully driving at night and within an environment temperature between 5°C and 20°C. This range could be extended to lower temperatures, as Fig. 3 shows that the camera sensibilities is almost linear for such temperatures, but testing under these conditions was not possible. At higher temperatures, the sensitivity of the sensor degrades very quickly, as shown in Fig. 3. Under sunny conditions, results have been disappointing as the reflection of sunlight on certain flat surfaces, such as brick walls, makes the algorithm unable to detect pedestrian most times. Figure 2 shows infrared images under different illumination and temperature conditions.

3.2. Camera model

The proposed system makes use of raw images of the microbolometer, obtained through a 14 bit A/D. The gray level of each pixel of these images represents the amount of heat that the sensor captures. The camera used is based on the non-refrigerated microbolometer and, as such, its sensibility changes in a way that is function of its temperature. It is necessary, then, to calibrate that sensibility.

Figure 3 represents the overall sensibility curve obtained. However, the computation of the threshold for this curve in each frame would be unreasonable high. That is why it is simplified locally and only applies to the usual work temperatures of the camera. Three sensibility curves are precomputed for: the higher and lower temperatures of the head and the lower temperature of the body.

Within the work temperature of the camera the sensibility can be approximated to a cubic function of the sensors temperature.

The camera is modeled as a pin-hole. The intrinsic characteristics are know and so is the position and orientation of the camera. The world system of coordinates is placed on

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

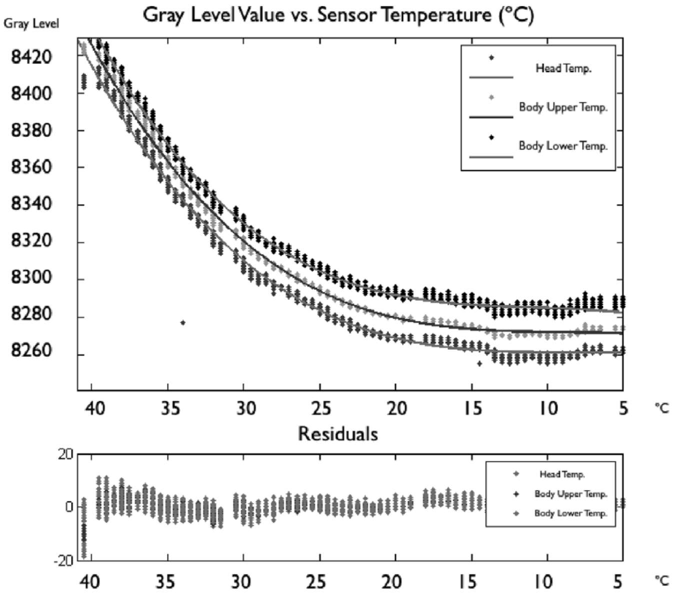


Fig. 3. Gray level of three constant temperatures of the human body against the sensor temperature.

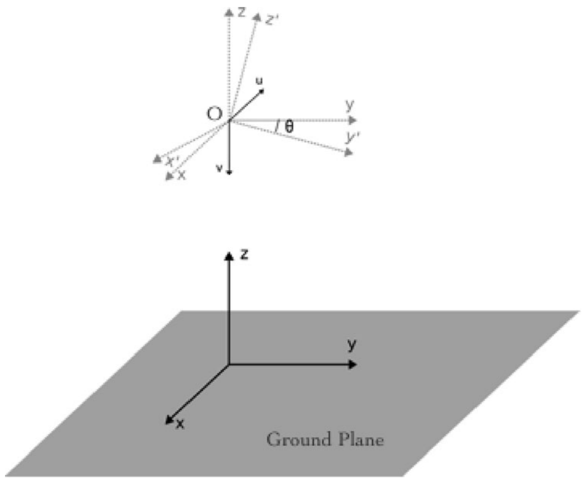


Fig. 4. Reference system of world and camera coordinates.

the ground plane, moving along with the vehicle and so does the camera position (Fig. 4).

The position of the pedestrian is modeled as a gaussian distribution in the xy plane of the ground. To determine accurately its distance to the camera, the homography of the ground plane onto the sensor is calculated for each frame (Eq. (1)). The projection of a 3D point in the image plane can be done if it is known its relative position to a certain plain. In this case, the camera position relative to the ground plane is known and can be assumed that it is constant. A more detailed explanation of the system setup can be found in Section 5.1. The rotation of the camera is known via a three degrees gyroscope:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = P \cdot W \cdot \begin{bmatrix} X \\ Y \\ Z \\ S \end{bmatrix}, \qquad (1)$$

where P is the intrinsics matrix. The camera movement between frames is modelled as W, which is comprised of the rotation matrix R and the translation vector T from the camera coordinate system to the ground. U and V are the image homogenous coordinates. The true pixel coordinates, $u = \frac{U}{S}$ and $v = \frac{V}{S}$, have their center in point O of Fig. 4, which is the optical center of the camera.

$$P = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}, \qquad (2)$$

$$W = \begin{bmatrix} R & T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\left(\frac{\pi}{2}+\theta\right) & -\sin\left(\frac{\pi}{2}+\theta\right) & t_y \\ 0 & \sin\left(\frac{\pi}{2}+\theta\right) & \cos\left(\frac{\pi}{2}+\theta\right) & 0 \end{bmatrix}. \qquad (3)$$

The results of the projection of a world point in the image plane are specially sensitive to variations of the skew angle θ (See Fig. 4). In the presented system this angle is know for each capture frame with the help of an gyroscope with three degrees of freedom attached to the same base as the camera. A point on the ground plane is projected on the image as:

$$u = c_u - \frac{X \cdot f_u}{Y \cdot \sin\left(\frac{\pi}{2}+\theta\right)} \qquad (4)$$

$$v = c_v - \frac{f_u \cdot t_y}{Y \cdot \sin\left(\frac{\pi}{2}+\theta\right)} - \frac{f_v}{\tan\left(\frac{\pi}{2}+\theta\right)} \qquad (5)$$

where f_u and f_v are the focal lengths on the u and v directions of the image; c_u and c_v are the coordinates of the center of the image. These four parameters are measured in pixels.

The intrinsics are obtained in a calibration process that involves the use of a special chessboard pattern, a matrix of incandescent lamps.

Finally, knowing that the image coordinates are $u = \frac{U}{S}$ and $v = \frac{V}{S}$, the relation between those and the world coordinates of the ground plane can be calculated with Eqs (4) and (5).

3.3. *Extraction of warm areas*

Pedestrians on the image are searched for at different resolutions, as the radiance that the sensor receives depends on the distance of the object to the camera. Each resolution consider an horizontal section of the image, with its lower points being the ground plane at different distances.

Extraction of the warm areas is done by thresholding the image in two phases: the first one tries to extract the heads of the pedestrians in the images; the second, the whole pedestrian silhouette. Objects within the normal temperature of the human body are thresholded. The result is a binarized image, containing blobs that can represent parts of the human body, specially heads and hands (Fig. 5). Since this first phase searches for the pedestrian head, those blobs that are not in the upper half of the image are ignored. Those blobs that are not within a reasonable size are also excluded.

Once the head candidates have been selected, a first set of regions of interest are generated. The highest point of the

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

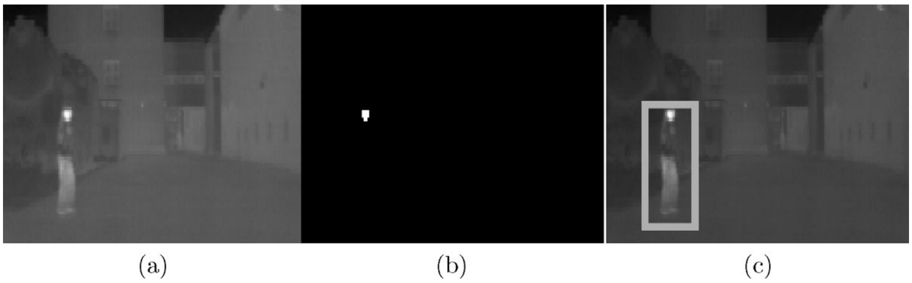


Fig. 5. Selection of regions of interest: (a) original image; (b) hotspots; (c) regions of interest.

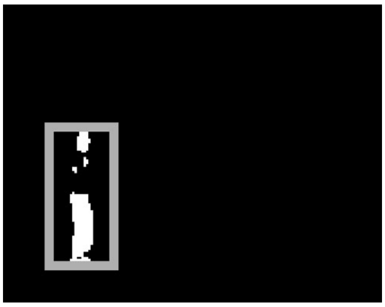


Fig. 6. Search of the pedestrian inside the previously calculated ROI.

head is also the top of the box, while the lowest point is at the closest point of the ground at that resolution. This way, the whole body of the pedestrian is included in the box, if there is any (Fig. 5c). The width of the bounding boxes is set to be 3/7 of the height, as it is a usual proportion of the human body. This width is big enogh to accommodate inside pedestrians of all sizes. At this point only the position of the head in the image is know, thus these bounding boxes have to be big enough to contain any pedestrian, no matter what height. A first approach is to suppose that the head is 200 cm from the ground plane on which the pedestrian is walking. The distance of the pedestrian to the camera (w_y) is given by Eq. (6), where w_z is the pedestrian's height, v is the position of the top of the region of interest, in image coordinates, and h is the camera's distance to the ground plane. The base of the region of interest is calculated with Eq. (5), for this new distance. The width of the bounding boxes is set to be 3/7 of the height, as it is a usual proportion of the human body.

$$w_y = \frac{(w_z - h) \cdot f_v}{v - c_v} \tag{6}$$

The regions of interest generated from the original image are now binarized with a threshold of t_1, that is the lower temperature established for the human body. Since most pedestrians height is less than 190 cm the distribution of temperatures inside the ROI will only seize a fraction of it. It is then resized, keeping the same proportions, assuming that the lowest part of the pedestrian are the feet, and that those are resting on the flat ground ahead the vehicle (Fig. 6).

3.3.1. Vertical symmetry. Pedestrians edges on far infrared images are usually very well defined against a cool background. Besides, human beings presents a very high vertical symmetry, so it can be a good descriptor to separate them from the non-pedestrian class. Edges are extracted from the original image using a vertical Sobel filter, both for positive and negative borders. Only the ROIs obtained in previous steps of the algorithm are considered, so vertical edges of each pedestrian have to be symmetrical around an axis located approximately at $w_i/2$, where w_i is the width of each ROI i. Equation (7) defines this symmetry, normalized by the size of the box.

$$S = \sum_{i=0}^{h} \left[\sum_{j=0}^{w} \left(I_P \left(i, \frac{w}{2} + j \right) \cdot I_N \left(i, \frac{w}{2} - j \right) \right) \right] \cdot \frac{1}{w \cdot h} \tag{7}$$

where I_P and I_N are the images of the positive and negative Sobel filter, respectively; w and h are the width and height of the ROI. If the symmetry S is not over a certain threshold, the ROI is deleted and will not be taken into account in the future.

3.3.2. Correlation with non-deformable models. Final verification of the extracted regions is done by means of gray scale correlation with some precomputed models. In far infrared images the most recognizable feature of pedestrians is the silhouette of the body temperature against the background. So, the correlation takes place between the final ROIs extracted thresholded with temperature t_1 (see Fig. 6) and the models, whose creation process is explained as follows.

From several sequences processed as explained in Section 3.3, extracted ROIs containing potential pedestrians are manually classified. The models are created computing the mean of the value of each pixel for the training group. Equation (8) returns the value of each pixel $P(x, y)$, being $R_i(x, y)$ the selected ROIs.

$$P(x, y) = \sum_{i=0}^{N} \frac{R_i(x, y)}{N} \tag{8}$$

where N is the number of ROIs selected for the creation of each model. In this case, every model has been generated out of $N = 50$ candidates.

Pedestrians have very different appearance depending on the their gait cycle. The main difference is due to the position of the legs. Thats why those ROIs that eventually contain a pedestrian are grouped in four different categories: open, almost open, almost closed and closed legs. An example

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

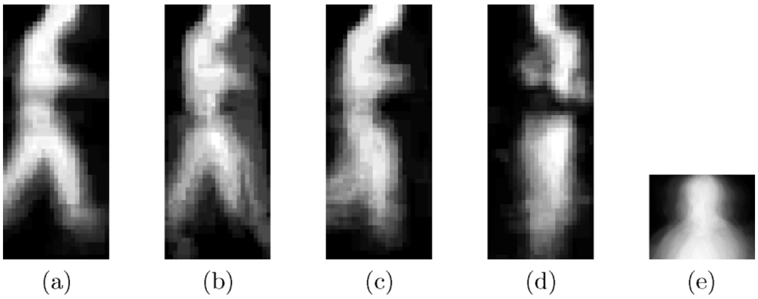


Fig. 7. Examples of pedestrian models. (a) Opened Legs; (b) Semi-opened legs (1); (c) Semi-opened legs (2); (d) Closed legs; (e) Head-only model.

of these models can be seen on figures 7(a), 7(b), 7(c) and 7(d). This approach enables the algorithm to correctly identify a wider diversity of shapes but it takes longer to process four correlations for each candidate. To reduce the number of calculations a fifth model is created for a common characteristic of pedestrians: the head. An example of this model can be seen on Fig. 7(e). For each candidate, the top third of the bounding box is correlated with this head model before carrying on with the other four. If there is not a satisfactory score, the region is deleted. Besides the improvement in computational time required, the use of this head model can prevent misdetections of pedestrians that are partially occluded, for example, wait to cross the road and standing behind a parked car. In those cases only the upper half of the person is visible and would return a negative from a full-body model correlation. Pedestrians located close to the camera present very rich details on the image, as far as long wave infrared vision goes. However, as this distance increases, the contour gets softer and the overall appearance of the candidates changes. To take account of all these possible shapes, four sets of the previous model have been created. Manually selected ROIs containing pedestrians are also sorted based on the distance to the camera. During the detection phase of the algorithm, correlation will only take place between the candidate and the set of models created for that particular range of distances, in which the candidate is. Examples of models for different distances can be seen on figure 8. Models have been created for four distance ranges: 5 to 15 m, 15 to 25 m, 25 to 40 m and over 40 m. Pedestrians closer than 5 m to the camera are sometimes incomplete in the images, and a good classification is not possible.

Correlation is done by Eq. (9), introduced in ref. [10].

$$c = \frac{\sum_{i=0}^{N}[(p(x,y)_i - 0.5)(M(x,y)_i - 0.5)]}{\sum_{i=0}^{N}[p(x,y)_i - 0.5]} \tag{9}$$

where $p(x, y)$ is each pixel of candidate ROI, $M(x, y)$ is each pixel of the model and N is the number of pixel inside the ROI.

4. Tracking

The Unscented Kalman Filter (UKF)[5,7] extends the general Kalman filter to non-linear transformations of a random variable without the need of linearization, as the Extended Kalman Filter (EKF) does.[12] This is particularly useful in the acquisition step of data through a visual system. As noted in Eqs (4) and (5), this transformation is highly non-linear. Thus, the use of the UKF is justified over the EKF, achieving better results with approximately the same computational demands. For self containment, the equations of the UKF are included. In Section 4.3 both the Kalman Filter and the UKF, are compared to each other in terms of performance and accuracy.

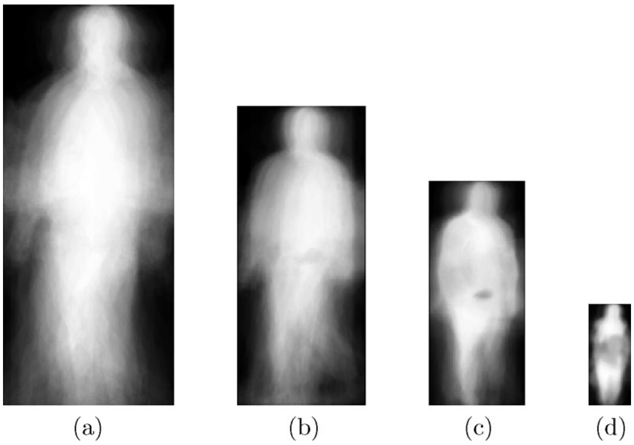


Fig. 8. Examples of pedestrian models for different ranges of distances to the camera. (a) 5 to 15 m; (b) 15 to 25 m; (c) 25–40 m; (d) 40 to ∞.

The tracking module follows the movement of each pedestrian that have been detected, therefore the state vector includes four variables: position and velocity in the x and y directions (Eq. (10)).

$$\hat{x} = \begin{bmatrix} p_x & p_y & v_x & v_y \end{bmatrix} \tag{10}$$

where p_x and p_y are the position and v_x and v_y are the velocity of the pedestrian being tracked.

The Unscented Kalman Filter propagates the random variable across the non-linear system using a minimal set of deterministically chosen weighted sigma points. The mean and variance of the transformed variable are accurate up to the second order of Taylor series expansion.

For a random variable of dimension n with mean $\bar{x}$ and covariance P the sigma points χ are:

$$\chi_0 = \bar{x} \tag{11}$$

$$\chi_i = \bar{x} + \sqrt{(n+\lambda)P} \quad i = 1, \ldots, n \tag{12}$$

$$\chi_i = \bar{x} - \sqrt{(n+\lambda)P} \quad i = n+1, \ldots, 2n \tag{13}$$

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

where $\lambda = \alpha^2(n+\kappa)$ is an scaling factor that determines how much spread are sigma points around the mean $\bar{x}$. In this case the values of α and κ are set to $\alpha = 0.01$ and $\kappa = 200$.

The selected weighted sigma points are propagated though the non-linear function f and the mean and covariance of the state are approximated.

For each sigma point, two weights are calculated, w^c and w^m:

$$w_0^c = \frac{\lambda}{n+\lambda} + 1 - \alpha^2 + \beta \quad \beta = 2 \tag{14}$$

$$w_0^m = \frac{\lambda}{n+\lambda} \tag{15}$$

$$w_i^m = w_i^c = \frac{1}{2(n+\lambda)} \tag{16}$$

Once the selected sigma points are propagated though the non-linear function f (Eq. (17)), weights w^m are used to approximate the mean (Eq. (18)) and w^c to approximate the covariance (Eq. (19)) of the state.

$$\gamma_i = f(\chi_i) \quad i = 0, \ldots .2n \tag{17}$$

$$\bar{y} = \sum_{i=0}^{2n} w_i^m \gamma_i \tag{18}$$

$$P_y = \sum_{i=0}^{2n} w_i^c [\gamma_i - \bar{y}][\gamma_i - \bar{y}]^T \tag{19}$$

4.1. Time update

On this step the movement model can be assembled as a time update of a simple Kalman filter since the observation steps are relatively small. The movement of the pedestrian is modeled as rectilinear between two consecutive frames and with constant velocity. True acceleration and any non-linearity are included in the update inside the error Q of the covariance P. The detected pedestrian position is simplified as it being in the intersection of its vertical symmetry axis with the ground plane. This way, the filter only tracks a single virtual point that moves always on the same plane. This prediction stage takes into account both the movement of the pedestrian and that of the vehicle.

$$\hat{x}_{t+1} = M \cdot R \cdot x_t + t_r \tag{20}$$

$$P_{t+1} = M \cdot R \cdot P_t \cdot (M \cdot R)^t + Q \tag{21}$$

The pedestrian's movement model is expressed with matrix M, in Eq. (22), where for each measurement, the predicted state of the position for the next moment (Eq. (21)) is the current plus the distance walked in the time between calculations.

As previously explained, the pedestrian movement is expected to be rectilinear and with constant velocity. This model is expressed with matrix M, in Eq. (22), where for each measurement, the predicted state of the position for the next moment is the current plus the distance walked in the time between calculations.

$$M = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

The vehicle movement is modeled as a combination of a translation across the ground plane and a rotation around the z axis, perpendicular to that same plane. Matrix R rotates both the relative position of the pedestrian to the vehicle and the direction of the velocity vector. The information about the vehicle's rotation angle α is always know though the merge of GPS and gyroscope measurements. This way, the motion of the vehicle, and consequently that of the camera, is compensated, and the real motion of the pedestrian can be isolated.

$$R = \begin{bmatrix} R_p & 0 \\ 0 & R_v \end{bmatrix} \tag{23}$$

$$R_p = R_v = \begin{bmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{bmatrix} \tag{24}$$

The process noise matrix is given by Eq. (25), where (a_x, a_y) is the acceleration of the vehicle.

$$Q = \begin{bmatrix} \frac{a_x^2 t^3}{3} & \frac{a_x^2 t^2}{2} & 0 & 0 \\ \frac{a_x^2 t^2}{2} & a_x^2 t & 0 & 0 \\ 0 & 0 & \frac{a_y^2 t^3}{3} & \frac{a_y^2 t^2}{2} \\ 0 & 0 & \frac{a_y^2 t^2}{2} & a_y^2 t \end{bmatrix} \tag{25}$$

4.2. Measurement update

The weighted sigma points are propagated through the transformation f, the homography of the ground plane onto the image sensor (see Eq. (1)).

The measurements of the pedestrians positions are obtained through a camera that is modeled as a pin-hole, without any aberrations. The mean and covariance of the predicted state are used to generate the sigma points as explained before. Then those points are propagated though the function f. This function is highly non-linear and a perfect candidate to use with the Unscented Transformation, as results with an Extended Kalman Filter can sometimes be more than disappointing.

$$\gamma_t^i = f\left(\chi_{t-1}^i\right) \tag{26}$$

As explained in Section 3.2 the image coordinates of an object that lies on the ground plane, known the fixed position of the camera over the vehicle, are function of the relative position of the object to the camera. This relation, for this particular case, is expressed in Eqs (4) and (5).

The set of chosen sigma points are propagated through the system, attending at the state that they represent. For each

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

state of the state vector the function is different and expressed as follows.

$$\gamma_i^0 = \frac{\chi_i^0 f_u}{\chi_i^1} + u_o \tag{27}$$

$$\gamma_i^1 = \frac{h f_v}{\chi_i^1} + v_o \tag{28}$$

$$\gamma_i^2 = \chi_i^2 \tag{29}$$

$$\gamma_i^3 = \chi_i^3 \tag{30}$$

The position measurements are derived from the image coordinates, while the velocity elements suffer no transformation. Since they are non observed variables of the system they can be assumed to be independent of the image coordinates.

The new propagated sigma points are used to obtain the predicted mean and covariance (Eqs (18) and (19)).

Finally, the new state is calculated. The last measurement y is included into this last step to update the state. The difference between the actual measurement and the expected one is resized by the Kalman Gain (K).

$$P_{xy} = \sum_{i=0}^{2n}\sum_{j=0}^{2n} w_{i,j}^c[\chi_{i,t|t-1} - \hat{x}_{t|t-1}][\gamma_{i,t|t-1} - \hat{y}_{t|t-1}]^T \tag{31}$$

$$K = P_{xy} P_y^{-1} \tag{32}$$

$$\hat{x}_t = \hat{x}_{t-1} + K(y - \hat{y}_{t-1}) \tag{33}$$

$$P_t = P_{t-1} - K P_y K^T \tag{34}$$

4.3. Kalman filter vs. UKF

The authors presents a brief comparison of the results obtained for the same sequence, tracking a single pedestrian from a static vehicle, with a simple Kalman Filter and an Unscented Kalman Filter. The UKF outperforms the Kalman filter in situations where the state follows a non linear transformation, or when information unknown to the model changes the state. Pedestrian tracking suffers from both. Although the pedestrian movement would be assumed to be linear, the movement of the camera, attached to an accelerating vehicle, is not. Pedestrians trajectories can also change suddenly. As an example, Fig. 9 shows the results of a test tracking the same single pedestrian. Both are initialized incorrectly, but manager to start the tracking. However, the simple Kalman FIlter is unable to follow the pedestrian once the heading changes.

4.4. Matching

For each pedestrian that enter the camera field of view a new UKF is created to follow it. However, since the actual number of pedestrians is *a priori* unknown, it is necessary to apply a matching algorithm between the detected pedestrians in a frame and the ones that are already being followed. Matching is based on the same variables as the state vector of the filter: relative position to the camera and velocity.

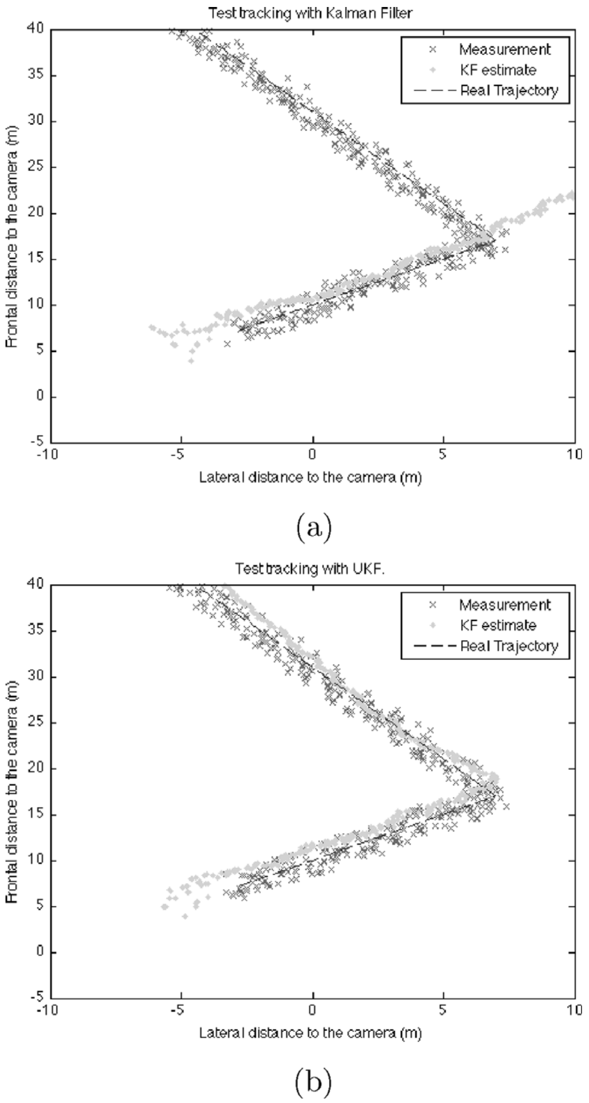


Fig. 9. Test tracking of a pedestrian that changes heading abruptly. (a) Kalman FIlter; (b) Unscented Kalman Filter.

For each frame takes place a comparison between the position and velocity of new pedestrian (B_i^t) in the scene and pedestrians that are already being tracked ($B_{L_i}^{t-1}$). For two candidates to be a correct match two conditions are set. First, position and velocity of the new pedestrian have to be within the limits that sets the covariance of the state. In the infrequent event that two or more pedestrian are found inside a single uncertainty area and that their velocities are also very similar matching is done with the one that minimizes cost Eq. (35). That is, those with a more similar state vector.

$$C_{L_i} = \rho \left\|X_L^{t-1} - X_i^t\right\| + (1-\rho)\left|\left(V_L^{t-1} - V_i^t\right)\right| \tag{35}$$

where X is the position in the ground plane, V is the velocity, L is the labeled pedestrian and ρ is the importance granted to the position over the speed. The cost is calculated for every $1 \leq i \leq N$, being N the number of pedestrians found inside the uncertainty area. Those pedestrians that have found no match are not immediately deleted. Their state is updated, skipping the measurement update, and are put on hold, waiting for it to reappear. This case can happen if

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

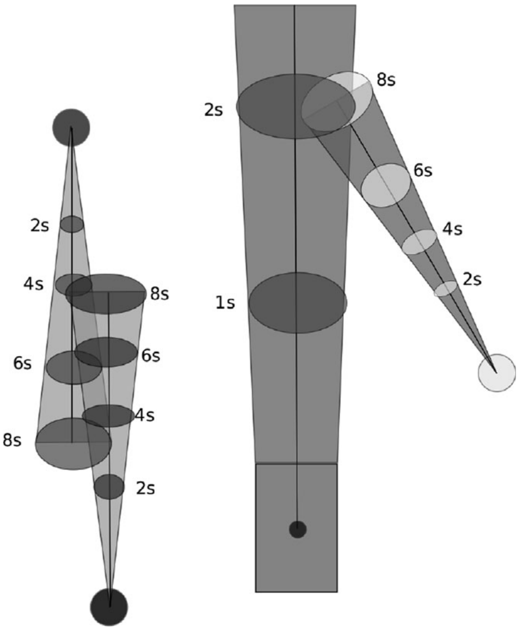


Fig. 10. Uncertainty areas of the vehicle and the pedestrians positions in the near future.

two pedestrians are crossing the street almost at the same speed and one is overlaying the other in the image for a few frames.

The purpose of tracking pedestrians from a moving vehicle is to anticipate a collision. The information about the speed and heading of those can help to determine its future trajectory. Of course, the assumption that pedestrians moves in a rectilinear trajectory and with constant velocity is very accurate between two consecutive frames, but as we try to predict its behavior further in the future the uncertainty of its position expands very quickly. To determine the risk of a collision, every time that a new frame is processed, the algorithm updates a future trajectory for the vehicle and every pedestrian, constantly incrementing the covariance of future positions. A representation can be found on Fig. 10.

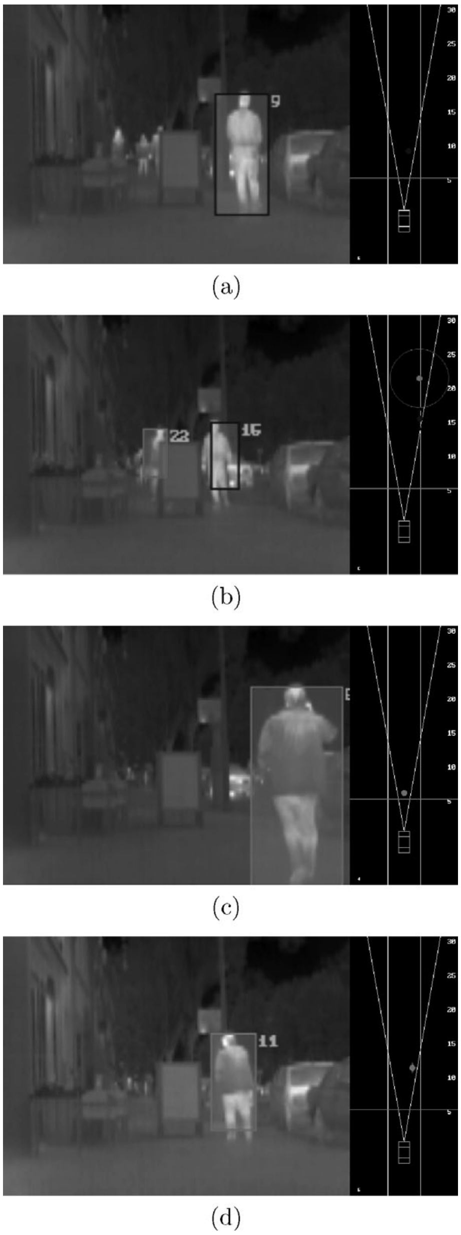
5. Results

5.1. System setup

The present system has been implemented as part of the IVVI experimental vehicle. The different algorithms are executed on one Intel Pentium D 2 Ghz processor. The camera used is a Flir Indigo Omega with sensitivity between the 7.5 µm and 13.5 µm.

The algorithm has been tested on an experimental platform. The results are based on 5 h of sequences, recorded in real urban traffic. It has operated at near real time, with usual speeds of 19 fps. The lowest speed registered have been 11 fps, while the algorithm was trying to track a crowd of pedestrians crossing the street in front of the vehicle.

As of the current development state the results have been satisfactory, classifying correctly almost 96% of bounding boxes closer than 45 m 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. Another option would be using a longer focal length lens.



(a)

(b)

(c)

(d)

Fig. 11. Pedestrians detected and tracked in a static sequence.

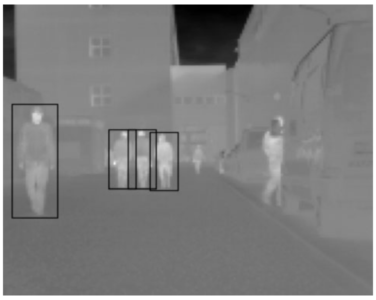
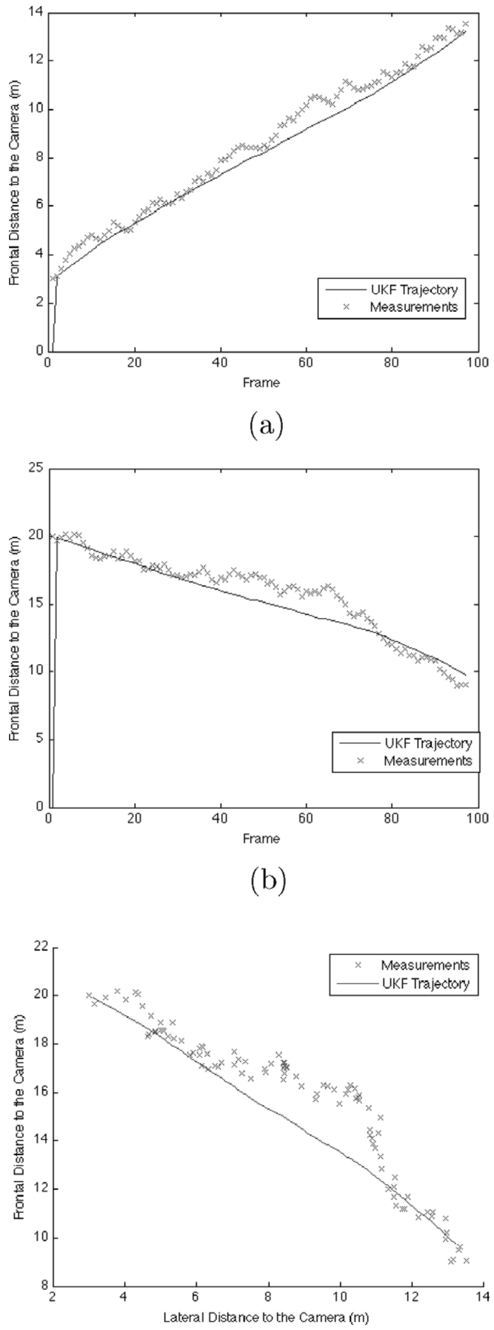


Fig. 12. False negative examples.

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press



(a)

(b)

Fig. 13. Detected position of a single pedestrian and the resulting trajectory of the UKF. (a) Frontal distance to the camera; (b) Lateral distance to the camera; (c) Movement of the pedestrian over the XY plane.

The algorithm have also been proven to be very solid against misdetections; that is, the rate of static objets classified as pedestrians. Under stated conditions of temperature and illumination, the false positive rate is of one in 300 images, which is approximately one every 20 s. However, the driver is not warned, if theres isn't a repetition of detections, so that occasional misclassifications don't affect normal driving.

The system has been tested driving through urban locations, and below 50 km/h. Urban driving is of special interest because there is where most of the accidents involving pedestrians happens.

A processed sequence is shown in Fig. 11.

Figure 12 is an example of two false negatives. One of the pedestrians is too far away, and its shape can hardly be identified as a person. The other one is partially occluded, specially the head. As the algorithm search for the head, the ROI extractor fails in this case.

Figure 13 depicts the trajectory of a single pedestrian being followed. It includes the noisy measurements of its position, as well as the output state of the UKF. It shows that bad measurements, do not affect the state of the position. As such, the false alarms are reduced and the driver is not annoyed unnecessarily.

On each iteration of the pedestrian tracking, the candidate confidence is updated. The driver is only warned if this confidence is above a certain threshold, after at least five frames from first detection. Driving at 50 km/h and processing 19 fps the alert sound is triggered when the pedestrian is 35 m ahead of the vehicle. The driver has then approximately 2.5 s to react.

6. Conclusions and Future Work

In this paper, a pedestrian detection system in FIR images based on template matching has been presented. It detects pedestrian within a range of 1–45 m in front of the vehicle, and predict a short-term trajectory based on the results of a tracking step by means of an unscented Kalman filter.

The results have been promising. However, new objectives have been considered to improve the algorithm. 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. It is also planned the fusion of the visual information of the pedestrian detector with a lidar system, to obtain accurate distance measurements.

Acknowledgments

This research was partially funded by CICYT Project POCIMA (TRA2007-67374-C02-01) and CICYT Project VISVIA (TRA2007-67786-C02-02).

References

1. M. Bertozzi, A. Broggi, P. Grisleri, T. Graf and M. Meinecke, "Pedestrian Detection in Infrared Images," *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE* (May 2003) pp. 662–667.
2. M. Bertozzi, A. Broggi, C. Hilario, R. Fedriga, G. Vezzoni and M. Del Rose, "Pedestrian Detection in Far Infrared Images Based on the Use of Probabilistic Templates. *Intelligent Vehicles Symposium, 2007 IEEE* (May 2007) pp. 327–332.
3. M. Bertozzi, A. Broggi, M. Del Rose, M. Felisa, A. Rakotomamonjy and F. Suard, "A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier," *IEEE Intelligent Transportation Systems Conference* (2007).
4. E. Binelli, A. Broggi, A. Fascioli, S. Ghidoni, P. Grisleri, T. Graf and M.-M. Meinecke, "A modular tracking system for far infrared pedestrian recognition," *Proceedings of IEEE Intelligent Vehicles Symposium* (2005) pp. 758–763.

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press

5. S. Julier and J. Uhlmann, “A new extension of the Kalman filter to nonlinear systems,” *Int. Symp. Aerosp./Def. Sens.* **3**, 26–38 (Jan. 1997).
6. B. Ling, M. I. Zeifman and D. R. P. Gibson, “Multiple pedestrian detection using IR led stereo camera,” *Intell. Robot Comput. Vision XXV: Algorithms, Tech. Active Vision* **6764**(1), 67640A (2007).
7. M. Meuter, U. Iurgel, S.-B. Park and A. Kummert, “The unscented Kalman filter for pedestrian tracking from a moving host,” *Intelligent Vehicles Symposium, 2008 IEEE* (2008) pp. 37–42.
8. R. Miezianko and D. Pokrajac, “People Detection in Low Resolution Infrared Videos,” *Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008. IEEE Computer Society Conference on*, (May 2008) pp. 1–6.
9. H. Nanda and L. Davis, “Probabilistic template based pedestrian detection in infrared videos,” *Intell. Vehicle Symp., 2002. IEEE* **1**, 15–20 (May 2002).
10. D. Olmeda, C. Hilario, A. Escalera and J. M. Armingol, “Pedestrian Detection and Tracking Based on Far Infrared Visual Information,” *Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems* (2008) pp. 958–969.
11. Z. Sun, G. Bebis and R. Miller, “On-road vehicle detection: A review,” *Pattern Anal. Mach. Intell., IEEE Trans.* **28**(5), 694–711 (2006).
12. M. Wan and J. Herve, “Adaptive target detection and matching for a pedestrian tracking system,” *Syst. Man Cybern. 2006. SMC ’06. IEEE Int. Conf.* **6**, 5173–5178 (Sep. 2006).
13. F. Xu, X. Liu and K. Fujimura, “Pedestrian detection and tracking with night vision,” *IEEE Trans. Intell. Transp. Syst.* **6**(1), 63–71 (Jan. 2005).
14. L. Zhang, B. Wu and R. Nevatia, “Pedestrian detection in infrared images based on local shape features,” *Comput. Vision Pattern Recognit., 2007. CVPR ’07. IEEE Conf.* (May 2007) pp. 1–8.

https://doi.org/10.1017/S0263574710000299 Published online by Cambridge University Press