# Detection and Tracking of Pedestrians in Infrared Images

Daniel Olmeda, Arturo de la Escalera and Jos M Armingol
Intelligent Systems Laboratory, Department of Systems Engineering and Automation
Universidad Carlos III de Madrid
C./ Butarque 15, 28911 Leganes, Spain,
dolmeda@ing.uc3m.es

***Abstract*—This article presents a pedestrian detector by means of a far infrared thermal camera, as part of an integrated driver assistance system. The final goal of the system is to warn the driver of pedestrians ahead of the vehicle. Detection is achieved by the extraction and evaluation of the main feature of pedestrians in far infrared images: the distribution of heat of the human body. Those objects with an appropriate size and temperature are correlated with probabilistic models that represent the average temperature of the different parts of the human body. Each pedestrian is tracked independently by an Unscented Kalman Filter.**

***Index Terms*—Computer Vision, ITS, ADAS, Pedestrian, Detection, Tracking**

## I. Introduction

Lately our society is witnessing an increasing interest about safety in traffic systems. Traffic Safety is an important concern in developed countries. 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.

The purpose of Advanced Driver Assistance Systems (ADAS) is to evaluate the surroundings and warn the driver of potentially dangerous situations. 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 emmit in conditions that, otherwise, would be unfit for exploiting a system based on visible light cameras, such as (and specially) night driving.

This paper is organized as follows: In Section II, we present a brief introduction of vision-based driver assistance systems. The proposed algorithm is described in sections III and IV. Section III focus on detection and localization of pedestrians ahead of the vehicle. In section IV the tracking algorithm from a mobile platform is explained. Section V includes an explanation about the implementation of the algorithms in the research platform IVVI (Intelligent Vehicle based on Visual Information), as well as some initial results. To overcome some of the noticed problems future research directions are presented in Section VI.

## II. State of the art

One of the more important tasks of ADAS systems is the detection of obstacles on the road. The presence of pedestrians is a particularly dangerous situation because, in case of a collision, they are much more likely to be hurt than the occupants of the vehicle [1]. 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, doesn't 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, 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 [2] [3] [4] [5]. Another important feature is the intensity of the borders between pedestrians and their background. Those are used in systems that select regions of interest based on the proximity of local shape features such as edgelets or histograms of oriented gradients [6] [7] [8] .

On systems relying on local shape features the classification step usually involves the use of AdaBoost. As for 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 [3] [5]. 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

978-1-4244-4398-7/09/$25.00 ©2009 IEEE

collision between pedestrian and vehicle. Yet, this step is usually neglected in papers describing far infrared pedestrian detection. The most common solution is the use of kalman filtering to determine the pedestrian position [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. 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.

## III. Detection

Pedestrian detection is achieved in this system by means of a far infrared camera. The sensor of these cameras represents 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.

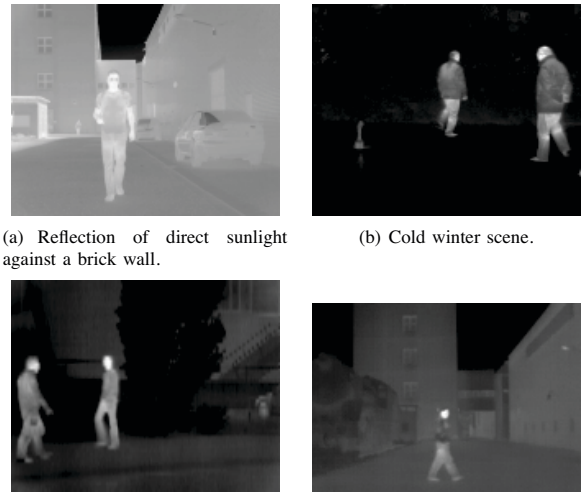
In this kind of application, pedestrians can be at any distance. Objects with a temperature close to the expected of a human body are searched for at different radiance scales.

Pedestrians in far infrared images presents some very distinctive characteristics, such as 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 doesn't match those of the human body.

### A. Sensor calibration

Since the system only looks for pedestrians, the sensor has 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 micro-bolometer 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 succesfully driving at night and within an environment temperature between $5C$ and $20C$. 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 1 shows infrared images under different ilumination and temperature conditions.



(a) Reflection of direct sunlight against a brick wall.

(b) Cold winter scene.

(c) Scene under heavy clouds.

(d) Night scene.

Fig. 1. Infrared images under different illumination conditions.

### B. Camera model

The camera is modeled as a pin-hole, and the intrinsic and extrinsic parameter are known. The world system of coordinates is placed on the ground plane, moving along with the vehicle and so does the camera position (figure 2).

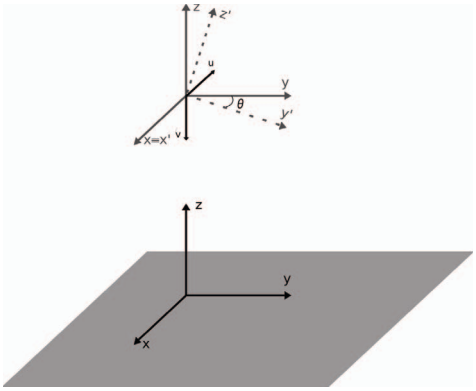



Fig. 2. System of reference of world and camera coordinates.

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 (equation 1). The rotation of the camera is known via a three degrees gyroscope.

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = P \cdot W \cdot \begin{bmatrix} X \\ Y \\ Z \\ S \end{bmatrix} \tag{1}$$

where $P$ is the intriniscs 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, being the true pixel coordinates $u = \frac{U}{S}$ and $v = \frac{V}{S}$.

$$P = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

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

Only the pitch angle ($\theta$) is considered so, a point on the ground plane is projected on the image as:

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

$$v = c_v - \frac{f_u \cdot t_y}{Y \cdot \sin(\frac{\pi}{2}+\theta)} - \frac{f_v}{\tan(\frac{\pi}{2}+\theta)} \quad (3)$$

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 images. These four parameters are measured in pixels.

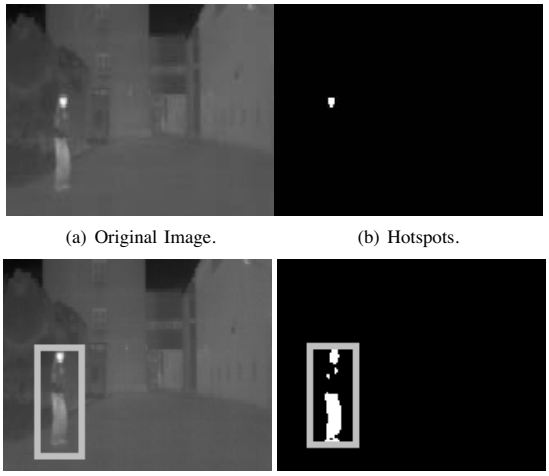
The intrinsics are obtained in a calibration process that involves the use of a special chessboard pattern, a matrix of incandescent lamps.

### C. 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. For each distance span an horizontal section is extracted from the image, and analyzed. The lower point of these section should correspond with the location of the ground plane. The highest point should guarantee that any pedestrian would fit inside, despite of its height.

Extraction of the warm areas is done by thresholding the image in two phases: the first one tries to extract the heads; the second one, 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 (figure 3(b)). 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 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 (figure 3(c)). 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.



(a) Original Image. (b) Hotspots.

(c) Regions of Interest. (d) Final ROI containing a pedestrian.

Fig. 3. Selection of regions of interest.

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. After this step, the whole shape of the pedestrian is selected in a new region of interest. The lower part of this bounding box rests on the ground, as it is located under the pedestrians feet (figure 3(d)). We assume the ground is flat in front of the vehicle.

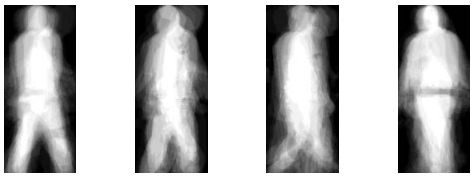
*1) Correlation with probabilistic models:* Final verification of the extracted regions is done by means of gray scale correlation with some precomputed models. In far infrared images the most recognizable feature of pedestrians is the silhouette of the body temperature against the background. So, the correlation takes place between the ROIs thresholded with the lowest temperature set for the human body (see figure 3(d)) and the models, whose creation process is explained as follows.

From several processed sequences, extracted ROIs containing pedestrians are manually classified. The models are created computing the mean of the value of each pixel for the training group.

Pedestrians have very different appearances depending on the their gait cycle. The main difference is due to the position of the legs. Thats why ROIs containing pedestrians are grouped into four different categories, attending to that position (figure 4). 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.

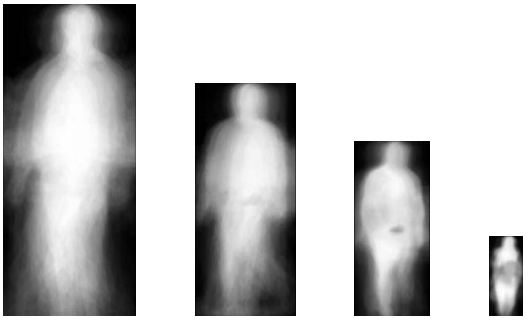
Pedestrians located close to the camera present very rich details on the image, as far as long wave infrared vision goes. However, as the distance increases, the contour gets softer and the overall appearance of the candidates changes. Manually

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 the particular range of distances in which the candidate is (figure 5). Models have been created for four distance ranges: 5 to $15m$, 15 to $25m$, 25 to $40m$, and over $40m$. Pedestrians closer than $5m$ to the camera are sometimes incomplete in the images, and a good classification is not possible.



(a) (b) (c) (d)

Fig. 4. Examples of pedestrians models for different appearences.



(a) 5 to 15m. (b) 15 to 25m. (c) 25-40m. (d) 40 to $\infty$.

Fig. 5. Examples of pedestrian models for different ranges of distances to the camera.

The correlation value between candidate and models is obtained by equation (4), [3].

$$c = \frac{\sum_{i=0}^{N} [(p(x,y)_i - 0.5)(M(x,y)_i - 0.5)]}{\sum_{i=0}^{N} [p(x,y)_i - 0.5]} \quad (4)$$

where $p(x,y)$ is each pixel of candidate ROI and $M(x,y)$ is each pixel of the model. If a ROI match with a probability over $65\%$ the tracker is initialized.

## IV. Tracking.

The Unscented Kalman Filter (UKF) [10] 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 [11]. This is particularly useful in the acquisition step of data through a visual system.

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 (equation 5).

$$\hat{x} = \begin{bmatrix} p_x & p_y & v_x & v_y \end{bmatrix} \quad (5)$$

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 are:

$$\chi_0 = \bar{x} \quad (6)$$

$$\chi_i = \bar{x} + \sqrt{(n+\lambda)P} \qquad i = 1, \ldots, n \quad (7)$$

$$\chi_i = \bar{x} - \sqrt{(n+\lambda)P} \qquad i = n+1, \ldots, 2n \quad (8)$$

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

### A. 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 is included in the update inside the error $Q$ of the covariance $P$ [12]. The detected pedestrian position is simplified as the projection of the center of gravity in the ground plane. This prediction stage takes into account both the movement of the pedestrian and that of the vehicle.

$$\hat{x}_{t+1} = M \cdot R \cdot x_t + t_r \quad (9)$$

$$P_{t+1} = M \cdot R \cdot P_t \cdot (M \cdot R)^t + Q \quad (10)$$

The pedestrian's movement model is expressed with matrix M, in equation 11, where for each measurement, the predicted state of the position for the next moment (equation 9) is the current plus the distance walked in the time between calculations.

$$M = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

The vehicle movement is modeled as a combination of a translation across the ground plane ($t_r$) 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 with the information of the gyroscopes. 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.

$$R = \begin{bmatrix} R_p & 0 \\ 0 & R_v \end{bmatrix} \tag{12}$$

$$R_p = R_v = \begin{bmatrix} cos(-\alpha) & -sin(-\alpha) \\ sin(-\alpha) & cos(-\alpha) \end{bmatrix} \tag{13}$$

*B. Measurement Update.*

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

$$\gamma_t^i = f(\chi_{t-1}^i) \tag{14}$$

The new propagated sigma points are used to obtain the predicted mean and covariance.

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.

*C. Matching.*

For each new pedestrian whose correlation with the models is over the threshold, 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. 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 ($B_{L_i}^{t-1}$). Position and velocity of the new pedestrian have to be within the limits that sets the covariance of the state for a positive match. If two or more pedestrian are found inside a single uncertainty area matching is done with the one that minimizes cost equation (15). That is, those with a more similar state vector.

$$C_{L_i} = \rho \left| X_L^{t-1} - X_i^t \right| + (1-\rho) \left| (V_L^{t-1} - V_i^t) \right| \tag{15}$$

where $X$ is the position in the ground plane, $V$ is the velocity, $L$ is the labeled pedestrian and $\rho$ is the importance granted to the position over the speed.

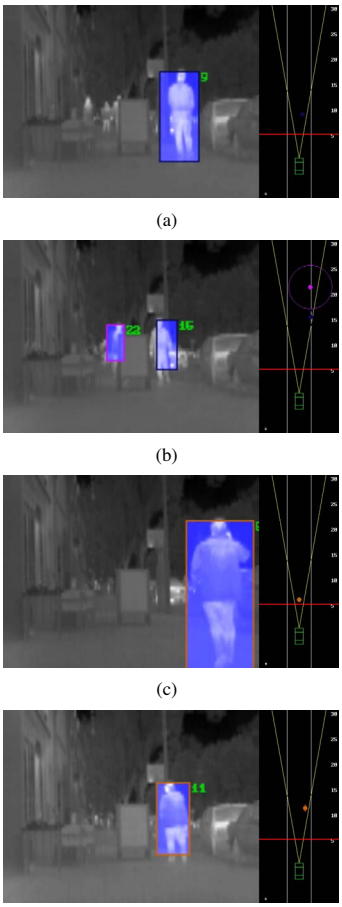
## V. Results

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

As of the current development state 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. Another option would be using a longer focal length lens. The algorithm have also been proven to be very solid against misdetections; that is, the rate of static objets classified as pedestrians is low.

The algorithm has been tested on an experimental platform and in real 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.

A processed sequence is shown in figure 6.



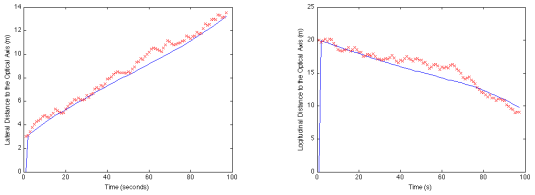
(a) (b) (c) (d)

Fig. 6. Pedestrians detected and tracked in a static sequence.

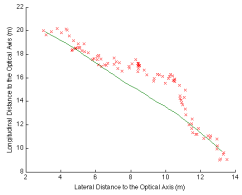
Figure 7 depicts the trajectory of a single pedestrian being followed. It includes the noisy measurements of its position, as well as the output state of the UKF. It shows that bad measurements, or outliers, do not affect the state of the position. As such, the false alarms are reduced and the driver 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 the first one it has been detected in. Driving at $50km/h$ and processing $19fps$ the alert sound is triggered

when the pedestrian is $35m$ ahead of the vehicle. The driver has then approximately $2.5s$ to react.




(a) Longitudinal distance to the vehicle. (b) Lateral distance to the vehicle.




(c) Movement of the pedestrian over the XY plane.

Fig. 7. Measurements and UKF states for a pedestrian's position.

## VI. Conclusions

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 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.

## Acknowledgment

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

## References

[1] K.Tro and M.Hubay andP.Stonyi and E.Keller *Fatal traffic injuries among pedestrians, bicyclists and motor vehicle occupants*. Forensic Science International, Volume 151, Issue 2, Pages 151-156

[2] M Bertozzi, A Broggi, P Grisleri and T Graf. *Pedestrian detection in infrared images* Intelligent Vehicles Symposium, 2003 Proceedings, IEEE, 2003

[3] M Bertozzi, A Broggi, Cristina Hilario, R Fedriga, G Vezzoni and M Del Rose. *Pedestrian Detection in Far Infrared Images based on the use of Probabilistic Templates*. IEEE Intelligent Vehicles Symposium, p327-332, May 2007.

[4] E Binell, A Broggi, A Fascioli, S Ghidoni, P Grisleri,T Graf and M Meinecke. *A modular tracking system for far infrared pedestrian recognition*. Intelligent Vehicles Symposium, 2005 Proceedings, IEEE. pp759 - 764, 2005

[5] H Nanda and L Davis. *Probabilistic template based pedestrian detection in infrared videos*. Intelligent Vehicles Symposium, 2002 Proceedings, IEEE, 2002

[6] M Bertozzi, A Broggi, M Rose and M Felisa *A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier.* Intelligent Transportation Systems Conference, 2007

[7] 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, pp 1-6, 2007

[8] Li Zhang, Bo Wu and R Nevatia *Pedestrian Detection in Infrared Images based on Local Shape Features*. Computer Vision and Pattern Recognition, 2007. CVPR '07, pp1-8, 2007

[9] F Xu and X Liu and K Fujimura. *Pedestrian Detection and Tracking With Night Vision*. IEEE Transactions on Intelligent Transportation Systems, 2005.

[10] S Julier and J Uhlmann. *A new extension of the Kalman filter to nonlinear systems*. Int. Symp. Aerospace/Defense Sensing. January 1997.

[11] Meng Wan and J Herve. *Adaptive Target Detection and Matching for a Pedestrian Tracking System*. IEEE International Conference on Systems, Man and Cybernetics, 2006. pp 5173-5178, vol. 6.

[12] M. Kohler. *Using the kalman lter to track human interactive motion - modelling and initialization of the kalman lter for translational motion*, tech. rep., Universitat Dortmund, 1997.