Pedestrian Detection on a Moving Vehicle: an Investigation about Near Infra-Red Images

A. Broggi, R.I. Fedriga, A. Tagliati
Dipartimento di Ingegneria dell'Informazione
Università di Parma
Parma, I-43100, Italy
{broggi, tagliati, fedriga}@ce.unipr.it

T. Graf, M. Meinecke
Electronic Research
Volkswagen AG
Wolfsburg, D-38436, Germany
{thorsten.graf, marc-michael.meinecke}@volkswagen.de

***Abstract*— The aim of this paper is to investigate the employment of a NIR (Near Infra-Red) camera on a moving vehicle for pedestrian detection, considering the NIR radiation as a suitable waveband for a single sensor device. In particular, we studied a system with special attention to cost, in order to make it affordable for all vehicles. We used as equipment an analog NIR CCD camera and examined the system behaviour in three different lighting conditions: daytime, nighttime using dipped-beam headlights and nighttime using full-beam headlights. Subsequently we chose the lighting situation which proved, in the long run, to be the most suitable for pedestrian detection. The algorithm proved to be efficient in common exurban scenarios.**

I. Introduction

Pedestrian detection on urban and exurban roads is one of the most challenging tasks in the artificial vision field. In recent years, several researches have been focused on the detection of pedestrians, because of the great impact on road safety. Recent works on pedestrian detection using Far Infra-Red (FIR) thermal images have achieved promising results, thanks to the specific appearance of human beings in such domain: effective pedestrian detection algorithms have been proposed [1], [2], [3], [4]. However, besides the previous positive topics, there are a number of disadvantages in the use of FIR radiation in automotive applications. Thermal images have a lower spatial resolution than images formed with NIR radiation, reducing sensitivity to textures. They generally have a lower grey level dynamic range than NIR ones. The energy emitted by the objects in the FIR band depends on their temperature. In an outdoor scene there are a number of factors that affect temperature and hence thermal contrast such as cloud cover, humidity, precipitation and wind. All of these effects alter and constantly change thermal contrast making road scenes viewed in the FIR domain difficult to interpret. Furthermore the fluctuating thermal contrast can approach zero [5], making large objects, e.g. cars, sometimes difficult to see. NIR images instead have a better contrast: the Near Infra-Red radiation, the wavelength of which can be mesured in the range 0.7-1.1μm, is reflected from the obstacles with almost the same behaviour as visible light. In fact a NIR system is named *active* because it demands illumination in order to acquire the radiations reflected by the surrounding objects. To provide such illumination in night scenarios many systems use NIR illuminators [6], [7], [8] as their visual range is greater (up to 150m) than that of standard dipped or full-beam headlights. These illuminators, though, introduce a relevant negative aspect: if two oncoming vehicles were equipped with NIR light sources, they would dangerously blind each other, as the human retina is vulnerable to thermal burns like those emitted by the Near Infra-Red radiation. For these reasons, we thought of investigating a NIR vision system which uses only the standard lighting equipment of a vehicle, dipped or full-beam headlights.

Our work presents a pedestrian detection algorithm for night vision conditions developed at the Artificial Vision and Intelligent System Laboratory of the University of Parma in collaboration with Volkswagen AG. This paper provides the description of the system developed and presents the results obtained together with some ideas for future improvements.

Section II presents an overview on the state of the art about the employment of NIR images for pedestrian detection. In Section III some preliminary considerations about the reasons for which we chose to develop a NIR vision system are illustrated. Section IV describes the algorithm structure. Results obtained are discussed in section V. Section VI contains some final considerations regarding the overall system behaviour and the guidelines for future developments.

II. State of the Art

Several studies presented in the literature describe systems for pedestrian detection in video sequences taken by NIR cameras mounted on a vehicle. Some of them use a single camera [9] or a hybrid system with both a NIR and a FIR camera [10] as sources. NIR illuminators are also very often employed [11], [12].

In order to reduce equipment expenses and to avoid eye damage caused by NIR radiations, we studied a low-cost acquisition system that employs only the vehicles headlights for night vision. Our approach proves to be efficient under common exurban scenarios.

III. Preliminary considerations

Given the system constrains (low cost, eye safety), we used as equipment a NIR CCD analog camera and, as illumination

sources, the dipped and full-beam headlights of the vehicle. Then we analyzed the images acquired with our system under different lighting conditions, in order to find which ones were the most suitable for pedestrian detection.

A. Daytime

The NIR radiation is emitted together with the visible light and has the same spatial distribution. Therefore, due to contiguity between Near Infra-Red and visible spectrum, natural objects can reflect both radiations very similarly. Thus images taken under daytime lighting conditions proved to be poor in informative content (fig. 1).

a *b*

Fig. 1. A daytime scenario in visible (*a*) and NIR (*b*) spectrum.

B. Nighttime using dipped-beam headlights

During nighttime, in order to acquire meaningful images, it is necessary to use external light sources. Having discarded specific NIR illuminators, for the previously listed reasons, we firstly employed the car standard dipped-beam headlights. The underlying idea was to develop a detection algorithm that was able to localize pedestrians just by searching for their legs. The dipped-beam headlights in fact illuminate only the lower area in front of the vehicle, leaving the upper one more darkened. The advantage of such kinds of images, besides the avoidance of oncoming vehicles and pedestrian dazzling, is that, being a smaller enlightened area, other kinds of disturbances, e.g. traffic signals, do not appear in the NIR images or, in case if they do, their grey level is lower than that of the pedestrians legs, easing their discarding procedure. Unfortunately the illuminated area is strongly asymmetrical (the left headlight is orientated towards the ground and the right one points upwards), jeopardizing the individualization of pedestrians appearing on the left side of the vehicle, who are in the shade (fig. 2*a*).

C. Nighttime using full-beam headlights

Our next experiments dealt then with the employment of the vehicle full-beam headlights. The main problem presented in this case was that the light cone produced by the headlights was not high enough to lighten pedestrians' heads as well as their torsos and legs (fig. 2*b*). Because of the special proportion existent between head and shoulders in the human being, this element very often is of great importance to differentiate pedestrians from other objects in the urban environment, like traffic signals or trees.

a *b*

Fig. 2. Employment of dipped-beam (*a*) and full-beam (*b*) headlights.

Nonetheless, the resulting lightened area proved to be quite extensive and uniform, allowing generic obstacles and pedestrians to appear in good contrast to the background.

IV. System description

Having produced the previous considerations while analyzing our system capabilities and characteristics, we chose to further investigate the third listed case with images acquired at nighttime using as the illumination source the vehicle full-beam headlights. All images represented in objects from here on have been acquired using the previously cited illumination system. Not having employed specific NIR illuminators, we still used NIR cameras instead of visible ones as they proved to be more efficient in detecting radiations emitted by many materials tipically used for clothes manufacturing (e.g. jeans), giving images with better colour contrast. We considered as a detection range the distances between $9m$ and $35m$, as pedestrians outside that range are not sufficiently illuminated in order to be detected. Another relevant topic of the algorithm is that it has been developed without using calibration data. This choice let us deal with some implementation problems, but also made easier the system setup phase. Calibration data can be used in a successive phase to refine the algorithm output.

A. Preliminary data processing

The first low-level operation consists in the enhancement of intense areas in the image. High brightness values individuate objects, such as traffic signals, pedestrian skin, some clothing fabrics, that greatly reflect the NIR radiation. The enhancement ameliorates contrast between such areas and their background. These areas are consequently surrounded by a rectangular perimeter that we will refer to as *bounding box*. From now on our research will be limited to evaluate such areas instead of the image in its entirety, letting the algorithm be computationally more efficient. As *bounding box* perimeters happen to be imprecise compared to a pedestrian's occupied area, a further espansion of such areas is carried out, evaluating, for each one pixel large row and column surrounding the *bounding box*, if there is a sufficient number of pixels over threshold, where the threshold is empirically obtained as half of the average grey level of the pixels contained in the original *bounding box*.

B. Bounding boxes evaluation

Bounding boxes are then binarized using the previously computed threshold (preliminary data processing, expansion phase). The evaluation of the entire body and legs is carried out using a group of statistical describers which allow to measure some image parameters, retained invariant: the *geometrical moments*. They let some global properties estimates of an object (the human body, in our case) and thus carry out the normalization of said object, making its recognition insensitive to position, dimension, deformation, rotation and small quantities of noise. For a generic discrete function in two variables, the moments are defined as:

$$M_{jk} \doteq \sum\sum X^j Y^k I(X,Y) \tag{1}$$

where function $I(X,Y)$ represents pixel's binarized value at coordinates $X,Y)$. Of particular importance are the zero and first-order moments: analysing eq. 1 we find that:

$$M_{0,0} = \sum\sum I(X,Y) \equiv Area \tag{2}$$

$$M_{1,0} = \sum\sum X \cdot I(X,Y) \tag{3}$$

$$M_{0,1} = \sum\sum Y \cdot I(X,Y) \tag{4}$$

from which we obtain:

$$C_x = \frac{M_{1,0}}{Area} \tag{5}$$

$$C_y = \frac{M_{0,1}}{Area} \tag{6}$$

where C_x and C_y represent the center of mass coordinates of the binarized *bounding box*. It is then possible to compute the moments of superior order having as datum point the center of mass coordinates. The expression of eq. 1 becomes:

$$M_{jk} \doteq \sum\sum (X - C_X)^j \cdot (Y - C_Y)^k \cdot I(X,Y) \tag{7}$$

invariant to image translations. Moreover, dividing eq. 7 by the value of *Area*:

$$M_{jk} \doteq \frac{1}{Area} \cdot \sum\sum (X - C_X)^j \cdot (Y - C_Y)^k \cdot I(X,Y) \tag{8}$$

we obtain invariance in respect to the scaling concerned with superior order moments computation. From this expression, it is possible to calculate the second-order moments for row, column and both:

$$M_X = M_{2,0} \doteq \frac{1}{Area} \cdot \sum\sum (X - C_X)^2 \cdot I(X,Y) \tag{9}$$

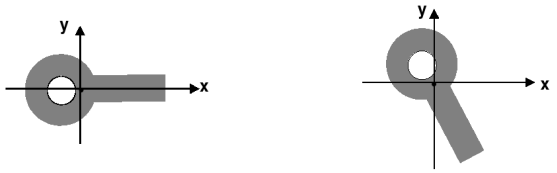
$$M_Y = M_{0,2} \doteq \frac{1}{Area} \cdot \sum\sum (Y - C_Y)^2 \cdot I(X,Y) \tag{10}$$

$$M_{XY} = M_{1,1} \doteq \frac{1}{Area} \cdot \sum\sum (X - C_X) \cdot (Y - C_Y) \cdot I(X,Y) \tag{11}$$

where:

- M_X is the inertial moment, relative to Y axis, in respect to the center of mass;
- M_Y is the inertial moment, relative to X axis, in respect to the center of mass;
- M_{XY} is the inertial moment, relative to both X and Y axis, in respect to the center of mass (it is worth 0 if X or Y are the main axis of the object).

These parameters are very important to understand object's orientation on the plane (fig. 3).



a *b*

Fig. 3. Second-order moments for a generic object: in (*a*) $M_X > M_Y$ and $M_{XY} = 0$. In (*b*) $M_X < M_Y$ and $M_{XY} \neq 0$.

Employing moments M_X, M_Y and M_{XY} it is possible to estimate exactly the inclination angle of an object with respect to its main axis or *axis of minimum inertia*[1]:

$$\tan 2 \cdot \alpha = \frac{2 \cdot M_{XY}}{M_X - M_Y} \quad \Longrightarrow \quad \alpha = \frac{1}{2} \arctan \frac{2 \cdot M_{XY}}{M_X - M_Y} \tag{12}$$

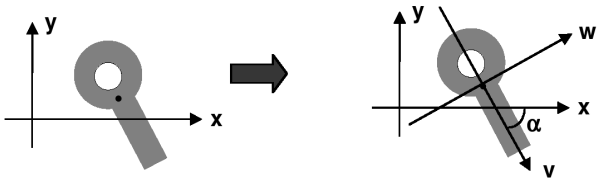


Fig. 4. Axis of minimum inertia for a generic object.

Another significant parameter which is possible to evaluate using second-order moments is the *eccentricity* of an object. It allows to estimate how similar to a circle a generic object is, and it is defined as:

$$Ec \doteq \frac{1}{Area} \cdot [(M_Y - M_X)^2 + 4 \cdot M_{XY}] \tag{13}$$

It assumes increasing values as its difference with a circle increases.

For each evaluation phase every parameter validity interval is established: this removes from successive elaboration phases *bounding boxes* external from identified intervals. So, after the previous considerations concerning geometrical moments, the parameters that we recognized as discriminant in the evaluation phase are the following:

[1]Axis inertia: quadratic distance sum between object's axis and each pixel.

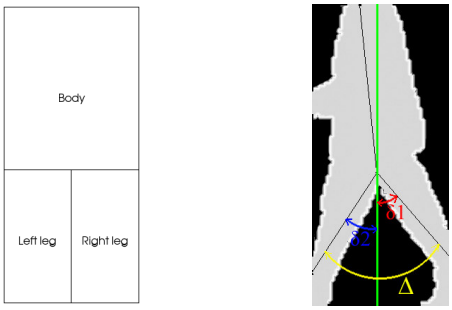
1) Eccentricity: from the observation of many image sequences, we found that it was possible to establish an interval of values between which the eccentricity of pedestrians is included, thus enabling the algorithm to discard quite simply many objects belonging to the urban environment. For example, circular traffic signals have low eccentricity values, while other objects belonging to the urban environment, such as poles, have high eccentricity values.

2) Objects inclination: for each *bounding box*, after having localized the principal axis of the objects included in it, the angle of such axis has been considered. As for the previous parameter, the inclination of the axis allows a quite easy filtering of objects different from pedestrians (e.g. trees).

3) Legs inclination: last parameter to evaluate, we searched for pedestrians legs, limiting the searching area to the lower part of each *bounding box* which survived the preceding filterings, and then further subdividing this reduced area in two other areas for the legs individuation process (fig. 5*a*).

C. Matching phase

Once the objects basic structure and a qualitative hint about pedestrians posture have been obtained, a more accurate control is necessary. We thus introduced a matching operation between each survived *bounding box* and a group of precalculated models. To choose the matching algorithm, much attention has been paid to computational costs, being the real-time application. The models belonging to the group are classified using the angle Δ formed by their legs while in different (walking or standing) positions (fig. 5*b*).



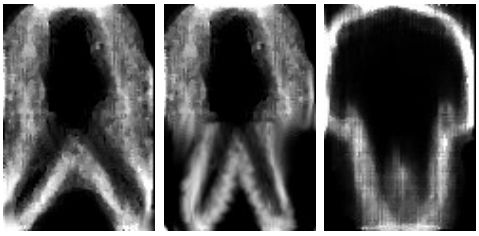
a *b*

Fig. 5. Example of *bounding box* subdivision for legs research (*a*) and Δ angle formed by legs (*b*).

Three classes of pedestrians, based on Δ angle validity intervals, have been created (fig. 6). During the matching phase, only the pedestrian's contour will be considered.

The global matching process occurs applying two different methods and then combining the results of such methods with a weighted average.

1) I^{st} method: models for the matching process are computed as probability maps of the pixels belonging to pedestrians contours, in different postures. The set of models is obtained as follows:



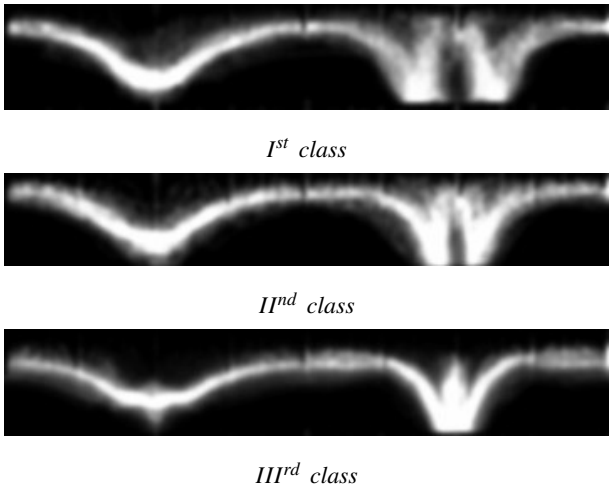
I^{st} *class* II^{nd} *class* III^{rd} *class*

Fig. 6. Pedestrian classes for the I^{st} *method.*

- pedestrians are extracted from a manually annotated set of images;
- pedestrians are classified, by the angle Δ formed by their legs, in 3 groups;
- edges of previous pedestrians are extracted;
- for each group a probabilistic template is computed.

The matching algorithm works simply as follows: for each pixel belonging to the contour of the object under evaluation, the value of the corresponding pixel in the selected model (that representing the class characterized by the same Δ angle) is assigned. These values are then added: if the resulting sum is larger than a threshold, computed as percentage of the total contour of the object, the match returns a positive result.

2) II^{nd} method: as the previous matching algorithm proved to be simple but not always effective, a second matching method has been developed. In this case (a sort of extension of the I^{st} method), we consider objects and models contours in their polar coordinates configuration. The set of models is obtained using the same process as in the I^{st} *method*, leading to the three classes shown in fig. 7.



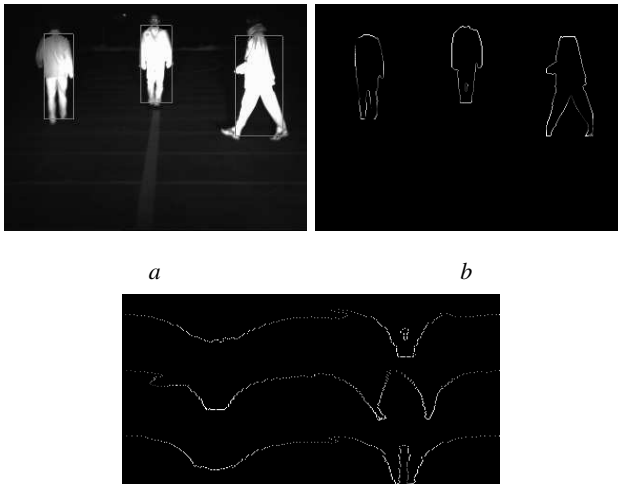
I^{st} *class*

II^{nd} *class*

III^{rd} *class*

Fig. 7. Pedestrian classes for the II^{nd} *method* (polar models).

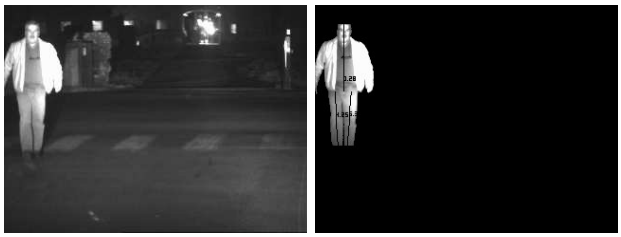
This evaluation system proved to be very effective in many situations. It works in fact as a good complementary method to the first one, so we combined both methods resulting in a weighted average.



a *b*

c

Fig. 8. *Matching phase* results: original image with detected *bounding boxes* (*a*), I^{st} (*b*) and II^{nd} *method* (*c*) applied to each *bounding box*. Higher matching values are visualized with brighter grey level pixels.



a *b*

Fig. 9. Best case: original input image (*a*) and respective output one (*b*) with pedestrian's main *axes of minimum inertia* visualized on it.

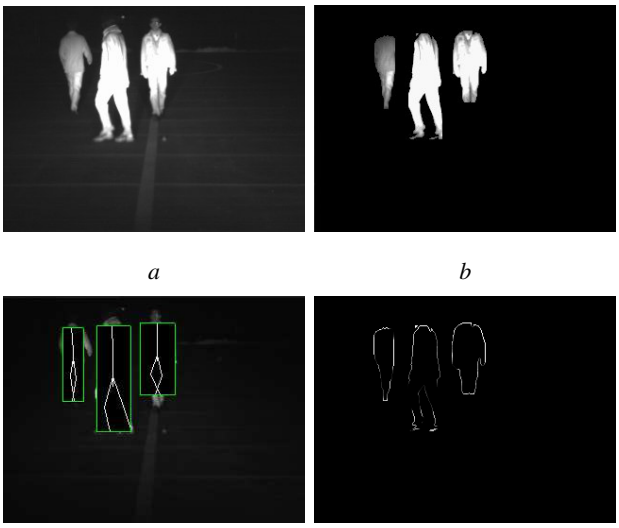
V. Results

The algorithm has been developed on a system with the following characteristics: AMD Athlon XP 2200+ processor, 1.8GHz, 256KB of cache memory, 512MB of RAM. The computation time required is very dependent on the number of *bounding boxes* found in the image, which determines the number of comparisons demanded by the matching phase. The average computational time required is about 20-25*ms*. As far as concerns the correct detection rate, best performances, obtained in images with high contrast between pedestrians and background (fig. 9), move around 90%, whereas worst ones (fig. 10) can bring the detection percentage down to 30%[2]. No performance comparisons with images acquired using dipped-beam headlamps have been shown as our approach requires good illumination of both legs and upper body for a pedestrian to be detected. This situation, if using the dipped-beam headlamps (thought not to blind oncoming vehicles), never occurs. The algorithm results are shown in following images (fig. 11 and 12).

[2]All results have been obtained processing sets of about 700 images.

a *b*

Fig. 10. Worst case: original input image (*a*) and respective output one (*b*) with detected objects main *axes of minimum inertia* and surrounding *bounding boxes* visualized; the green *bounding box* highlights the only pedestrian detected.



a *b*

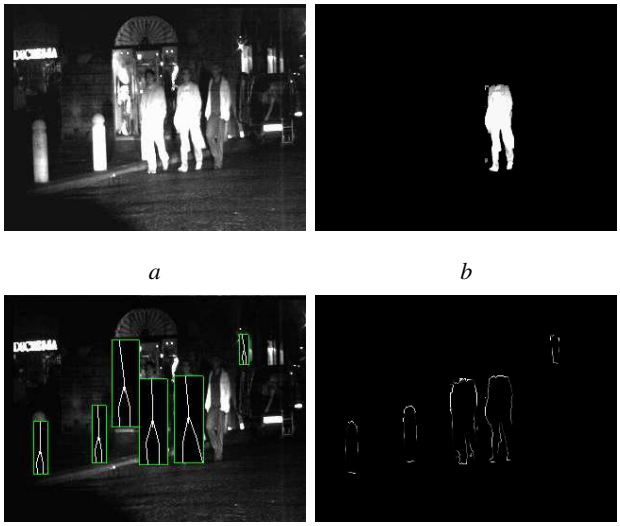
c *d*

Fig. 11. Good contrast between pedestrians and background: all pedestrians are well detected. Original input image (*a*) and respective output one (*b*) with two other images showing intermediate processing results: pedestrians main *axes of minimum inertia* visualization (*c*) and matching process output (*d*).

VI. Conclusions

In this paper we have evaluated how convenient it is to use an economical NIR vision system for pedestrian detection in automotive applications during nighttime, employing as illumination sources the full-beam headlights of a vehicle. During problem's analysis, a lot of attention has been addressed to individuate NIR images most important and distinctive characteristics. The developed algorithm firstly enhances bright areas in the images and surrounds them with a rectangular perimeter (*bounding box*), in order to reduce the areas that will pass to the second phase evaluation. It then estimates the content of such *bounding boxes* with a weighted combination of two matching methods.

The algorithm produced satisfactory results in good lighting conditions and in situations where the environment or traffic illumination did not create disturbance in the acquired images. The NIR image in fact may contain large areas of very low grey level where there is little or no information at night,



a *b*

c *d*

Fig. 12. Bad contrast between pedestrians and background: only one pedestrian is detected. Original input image (*a*) and respective output one (*b*) with two other images showing intermediate processing results: pedestrians main *axes of minimum inertia* visualization (*c*) and matching process output(*d*).

if the energy of the illuminators is not reflected back to the sensor, but may also contain very intense regions e.g. lights from oncoming vehicles or traffic signals. Difficulties aroused during the development phase have been both software and hardware. In the case of pedestrians overlapping, for example, it is difficult to enclose them correctly with a *bounding box* each: for these sort of problems a *tracking* system would allow the algorithm correct detection to rise. Another simple but effective improvement would be obtained introducing calibration data in order to eliminate all those *bounding boxes* which do not obey the perspective constrains (e.g. dimensions). for future developments, as far as the hardware is concerned, the employment of a High Dynamic Range CMOS NIR camera instead of a CCD one (fig. 13), should allow images with less saturated (bright) areas and thus with a higher informative content.

a *b*

Fig. 13. NIR images acquired with a CCD (*a*) and a CMOS (*b*) camera.

ACKNOWLEDGMENT

This work has been funded by Volkswagen AG.

REFERENCES

[1] H. Nanda and L. Davis, "Probabilistic Template Based Pedestrian Detection in Infrared Videos," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Paris, France, June 2002.

[2] X. Liu and K. Fujimura, "Pedestrian Detection using Stereo Night Vision," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 6, pp. 1657–1665, Nov. 2004, iSSN 0018-9545.

[3] Y. Fang, K. Yamada, Y. Ninomiya, B. K. P. Horn, and I. Masaki, "A Shape-independent Method for Pedestrian Detection with Far-infrared Images," *IEEE Trans. on Vehicular Technology*, vol. 53, no. 6, pp. 1679–1697, Nov. 2004, iSSN 0018-9545.

[4] A. Broggi, M. Bertozzi, R. Chapuis, F. C. A. Fascioli, and A. Tibaldi, "Pedestrian Localization and Tracking System with Kalman Filtering," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004, pp. 584–589.

[5] W. L. Wolfe and G. J. Zissis, *The Infrared Handbook*, Washington, USA, 1985.

[6] D. Kooß, F. Bellotti, C. Bellotti, and L. Andreone, "EDEL - Enhanced Driver's Preception in Poor Visibility," in *Progress in Automobile Lighting (PAL) Symposium 2003*, Darmstadt, Germany, Sept. 2003.

[7] D. Kooß, F. Bellotti, C. Bellotti, L. Andreone, and M. Mariani, "EDEL - Developing a Second Generation Night-Vision System," in *eBusiness and eWork*, Vienna, Austria, Oct. 2004.

[8] L. Andreone, P. C. Antonello, F. Bellotti, D. Grimm, M. Mariani, M. Wüller, and J. Heerlein, "Development of a New Generation of Driving Support System for Night Vision Enhancement: the EDEL Project," in *Intelligent Transport Systems ITS 2005*, Praha, Czech Republic, 2005.

[9] H. Sun, C. Hua, and Y. Luo, "A Multi-stage Classifier Based Algorithm of Pedestrian Detection in Night with a Near Infrared Camera in a Moving Car," in *Procs. Third Intl. Conf. on Image and Graphics (ICIG'04)*, Hong Kong, China, Dec. 2004.

[10] F. Jahard, D. A. Fish, A. A. Rio, and C. P. Thompson, "Far/Near Infrared Adapted Pyramid-Based Fusion for Automotive Night Vision," in *Image Processing and its Applications (IPA97), IEE Conference Publication No. 443*, Dublin, Ireland, July 1997.

[11] C. Bellotti, F. Bellotti, A. D. Gloria, L. Andreone, and M. Mariani, "Developing a Near Infrared Based Night Vision System," in *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004.

[12] L. Andreone, F. Bellotti, A. D. Gloria, and R. Lauletta, "SVM-Based Pedestrian Recognition on Near-InfraRed Images," in *Procs. 4th Intl. Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, Sept. 2005.