SVM-Based Pedestrian Recognition on Near-InfraRed images

Luisa Andreone[a], Francesco Bellotti[b], Alessandro De Gloria[b], and Roberto Lauletta[b]

[B] *FIAT Research Centre*
Strada Torino, 50
10043 Orbassano (TO) – Italy
+39 011 9083071
luisa.andreone@crf.it

[A] *DIBE – University of Genoa*
Via Opera Pia 11/a
16145 Genova – Italy
+39-010-3532780
franz@elios.unige.it

Abstract

This paper describes the algorithms we developed for a new automotive night vision system for pedestrian detection based on Near Infrared (NIR) illuminators and sensors. The system applies in the night domain the SVM technique, which has already been successfully implemented in day-light applications. In this project we have developed optimizations in order to meet accuracy and time performance requirement for in-vehicle deployments. In particular, we present a novel pre-SVM processing technique, which performs pixel-level and multi-resolution analysis in order to discard portions of the frame that are not likely to contain pedestrians. This procedure allows exploiting the SVM as a very accurate classifier focused on the most critical cases.

1. Introduction

Advanced Driver Assistance Systems (ADAS) are getting ever more widespread and important. Huge public and private funds have been allocated to such systems in order to meet important targets in terms of reducing the number of accidents. For instance, the European Union's 6th Framework Programme aims at reducing road accident fatalities by 50% by 2010. Accidents occurring at night represent a significant part of the total number of road fatalities. Several studies have pointed out that reduced visibility is one of the major impairment factor [1]. "Motorists fail to appreciate the limitations of their visual functions at night" [2]. Consequently, drivers "may routinely behave in a manner that is not supported by their visual ability at night" [3]. Thus, reduced visibility crashes may be alleviated by systems that compensate for the drivers' inability to see adequately [4]. Thus, intelligent systems for supporting the driver during night promise to have a significant impact on traffic safety [5].

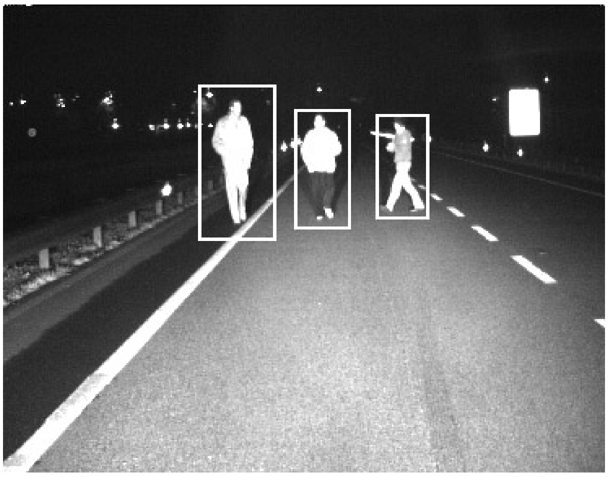

Fig. 1 Recognition of pedestrians with the EDEL NIR Night Vision System.

This paper presents the latest developments of the EDEL project (Enhanced Driving in Poor Visibility Conditions) [6], co-funded by the European Commission within the 5th Framework programme. EDEL involves CRF, Bosch, Jaguar, Hella, OSRAM and the Universities of Genoa, Siena and Karlsruhe. It is aimed to develop a semantic night vision system based on Near Infrared (NIR) sensors and a novel illumination system based on multi-element array laser technology integrated in a newly designed vehicle projector headlamps (Fig. 1).
EDEL has focused on pedestrian detection, as they are frequent victims of night accidents. The implemented system is being integrated in three car demonstrators at FIAT, Jaguar and Bosch.
The EDEL system involves a variety of aspects, which have been addressed in several papers, such as: NIR

Illumination system [7], User requirements and customer-benefit analysis [8,9], and overall system [10,11]. In this paper we present the latest advances in the artificial vision module, which we had introduced in [12].
The main contribution of this paper consists in providing a description of the system under development, presenting the design approach and discussing early results from the implemented pedestrian recognizer. In particular, our semantic scene interpreter relies on a Support Vector Machine (SVM), which is a machine-learning technique well established in the scientific community [13, 14]. This technique has been proposed for pedestrian recognition at the MIT, where Poggio et Al. successfully exploited it in daylight vision applications [15]. We have extended this approach to the night vision, drawn by the fact that NIR images have a similar spectrum as daylight images. The SVM approach is very general and the system can be trained to recognize a variety of classes (e.g. obstacles, animals, etc.), even if, due to real-time performance limitations, our system just implements a pedestrian recognizer, at present. Moreover, a learning machine adapts its behaviour according to its input, in order to continuously improve expected performance [16]. Night vision is an application which will be exploited in a huge variety of situations and contexts of use, which cannot all be foreseen and taken into consideration a-priory. Thus, incremental learning and adaptation is a fundamental requirement for such an application.
To the best of our knowledge, this paper is the first one which discusses in-depth a concrete working experience on the item, which may be of interests for researchers and practitioners in order to extend their knowledge on issues, possible solutions and trade-offs concerning implementation of night-vision techniques for automotive applications. In particular, we present a novel pre-SVM processing technique, which allows exploiting the SVM as a very accurate classifier focused on the most critical cases.

2. Experimental settings

The findings we present in this paper concern processing of a number of video sequences taken from a prototype vehicle equipped with the EDEL NIR illumination and sensor system [10,11]. The sequences involve 2 main scenarios, extra-urban and urban roads.
We have selected over 8,000 frames, and extracted sample images for pedestrians and non-pedestrians. Such samples have been used for statistical analysis and to build the models onto which our algorithms rely. The current database, which also constitutes the SVM's training set, includes about 3,000 pedestrians and 6,000 non-pedestrians. We foresee to further increase the database up to a reasonable size of about 6,000 – 10,000 pedestrians and non-pedestrians. Training the system with ever more images is very useful to improve precision. However, this involves increasing the number of support vectors (i.e. the vectors that define the hyper-plan separating the pedestrian and non-pedestrian spaces in the SVM) and consequently the frame processing times.
Thus, we have investigated and implemented several additional solutions in order to increase the overall system recognition accuracy without lowering time performance.
The performance results we present here have been obtained on a 1.4 GHz Pentium 4 PC, which is the computer that will be installed in the car prototype.

3. The algorithms

In our technique each frame, which is sized 372 X 287, is subdivided into 5 rectangular regions, or stripes (Fig.1). Each region corresponds to a specific range of target distances, as reported in table 1. Every stripe is subdivided into windows of different sizes, according to the distance from the camera (64x128, 48x96, 32x64, 24x48, 16x32 pixels). According to our empirical analysis, this granularity allows pinpointing pedestrians in a distance range from 6 to 80 meters. The total number of windows is 13116. In each window we check the presence of three different classes of pedestrians: frontal pedestrians and right and left lateral pedestrians. This distinction is important since the three different positions show different patterns that cannot be effectively captured in a single model.

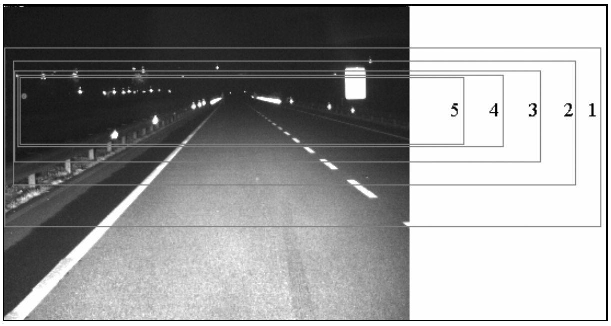

Fig. 2 Subdivision of a frame in regions

Region Number	Min Distance (mt.)	Max Distance (mt.)	Window size (HxW)	Number of windows
1	6	9	128x64	532
2	10	18	96x48	954
3	19	29	64x32	1640
4	30	50	48x24	2640
5	51	100	32x16	7350

Table 1 – Characterization of the frame regions

With such a number of windows, even if our SVM used just 30 input coefficients per each window, which is a very minimum number, we would obtain a frame-processing time of about 1400 ms, which compromises real-time performance. Since SVM computation is highly resource consuming, the idea driving our approach is to filter a number of windows before SVM processing. Thus, we use simpler methods in order to discard windows that are very likely not to contain pedestrians and leave the SVM computation just to solve the critical cases. This allows exploiting the SVM as a very precise classifier, considering a larger number of input coefficients in order to improve accuracy.

In order to discard windows that do not contain pedestrians, we have thought of and implemented a heuristic pre-processing phase, which consists of two steps.

The first step is based on a pixel-level analysis. In day-light conditions, the high variability of shape and colours of human bodies and of the background makes such an analysis not useful [17]. However, pixel analysis is suitable in night vision applications since properly-lit night scenes captured through NIR sensors show higher contrast levels [11], that make shape detection effective. The pixel-level pre-processing we implemented relies on a voting system based on the matching between the current window and a pedestrian model built on statistical data and geometrical features. The pedestrian model is at very rough granularity since the aim of the test is to discard windows that do not contain pedestrians, rather than to identify pedestrians. The test for frontal pedestrians involves 66 rectangular areas for each window. The areas are of different size according to the size of the window. For lateral pedestrian, a lower number of rectangles are needed: just 33.

While very conservative, this technique allows discarding an average of 70% windows in urban roads and 95% in extra-urban roads.

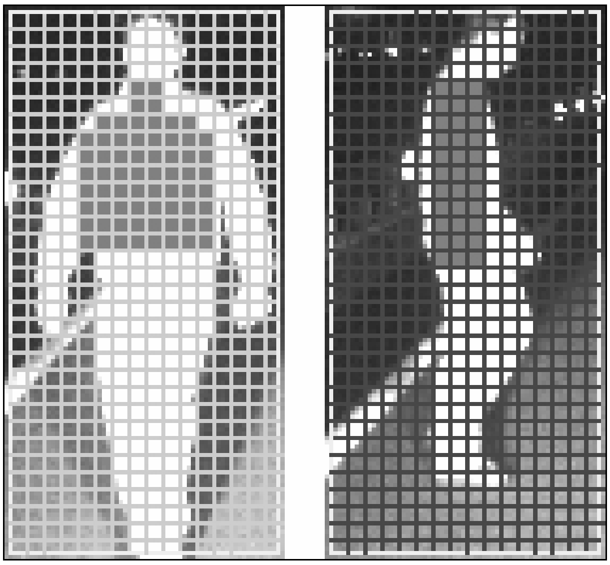

Fig. 3 The pixel-level model of frontal and lateral pedestrians

One of the key issues in developing an object detection system concerns the type of object representation, which has to guarantee low intra-class variability and high interclass variability. As mentioned above pixel-based representations are not suited for actual pedestrian recognition. Poggio et Al. [15] proposed a "template-ratio" approach based on base functions that are able to encode the difference in the average intensity between adjacent region. In particular, the Haar wavelet transform, which considers the differences between blocks of an image at different levels of detail [18], is efficient in extracting few information units that encode the differences between pedestrians and non-pedestrians images [15].

Thus, after using a pixel-level technique for discarding windows that clearly do not contain pedestrians, we proceed with a wavelet-based analysis.

The second step of the pre-processing phase thus consists in performing the Wavelet transform of the "survived" windows in two different resolutions and for three orientations: vertical, horizontal and diagonal. For each orientation, we obtain 13x29 high-resolution coefficients and 5x13 low-resolution coefficients. Such coefficients are obtained with a supersampling of the images through a redundant set of base functions (quadruple density dictionary) [19]. The total number of coefficients is 1326.

We perform the Wavelet transformations on the original windows, using different basis function domain sizes according to the actual size of each window. Differently from what presented in [15], this prevents bilinear filtering to resize all the windows to the standard 128*64 size. We then normalize the obtained coefficients in order to make the quantities comparable across the various window sizes and to implement just one SVM's training set.

With 1326 coefficients, computation times are in the order of the seconds, which makes the analysis impractical. Thus, the number of coefficients to be considered has to be reduced. Our statistical analysis showed that about 250 are sufficient to characterize a pedestrian pattern. Considering just these coefficients allows achieving a good trade-off between accuracy and real-time performance.

The most significant coefficients are those that typically distinguish pedestrians from non-pedestrians. Pedestrians are characterized by high coefficient values (representing discontinuities in the image) close to the sides of the body, low values in the center of the body, and typically random values in the background. This is a clear pattern, which is not present in most non-pedestrian images, which tend to have quite different shapes.

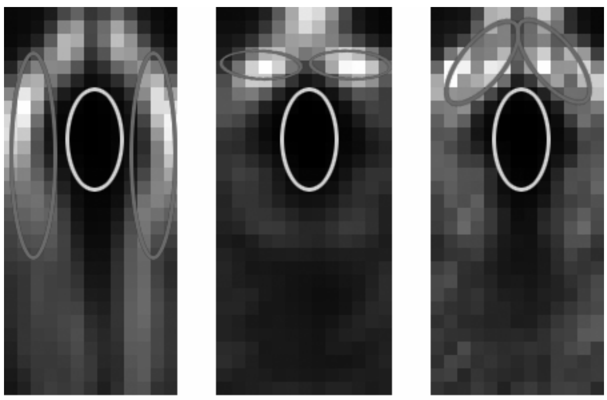

Fig. 4. Distribution of the Wavelet coefficients for frontal pedestrians (vertical, horizontal and diagonal basis functions)

We thus built a pedestrian model which considers three classes of coefficients: one internal, corresponding to the body, and two lateral, corresponding to the edges of the pedestrian. Fig. 4 shows such classes of significant coefficients for the three different orientations of the Wavelet transform. The images represent the average values of the coefficients of frontal pedestrian images, normalized in a 0 (black) – 255 (white) range. The model relies on statistical data extracted from the image database, with correction factors we introduced in order to keep into account geometrical features such as symmetry with respect to the vertical axis. Table 2 provides a detailed characterization of the coefficients. For non-frontal pedestrians, we consider just one model, whose coefficients are flipped for the left and the right side.
A threshold criterion has been introduced in order to compare, during the actual execution of the algorithm, the candidate image's coefficients with the model's coefficients.

Pedestrian type	Wavelet coefficients transform		Number	Threshold on normalized scale 0 - 255	Minimal percentage
Frontal pedestrian	Vertical	Right side	36	> 80	60%
		Left side	36	> 80	60%
		Centre	44	< 10	80%
	Horizontal	Right side	15	> 80	60%
		Left side	15	> 80	60%
		Centre	41	< 10	80%
	Diagonal	Right side	13	> 80	60%
		Left side	13	> 80	60%
		Centre	37	< 10	80%
Lateral pedestrian	Vertical	Right side	24	> 80	60%
		Left side	24	> 80	60%
		Centre	24	< 10	60%
	Horizontal	Right side	25	> 80	60%
		Left side	25	> 80	60%
		Centre	21	< 10	80%
	Diagonal	Right side	21	> 80	60%
		Left side	21	> 80	60%
		Centre	25	< 10	80%

Table 2 – Characterization of the Wavelet coefficients. For both frontal and lateral pedestrians, we perform the Wavelet transform with vertical, horizontal and diagonal basis functions. In each transform, we consider three classes of significant coefficients (centre, corresponding to the body of the pedestrian, and right and left sides, corresponding to the lateral edges of the pedestrian)

In order to speed-up the runtime processing time, we proposed an iterative procedure which checks the result at each step before going on to the next phase. So, we first compute the "black" coefficients of the vertical transform. If at least 60% of them are below the threshold, then the other black coefficients are computed and so on.

At the end of this wavelet analysis, we manage to reduce the number of candidate windows by a further 99% on average, as detailed in table 3, which summarizes the performance of the various techniques we implemented.
Only the windows that have passed this pre-processing phase are then fed into the SVM classifier for the final decision. The final stage is performed by two SVM classifiers, one for lateral pedestrians and one for frontal pedestrians. Each classifier processes the 48 most significant Wavelet coefficients from each candidate window.

Technique	Number of processed windows per frame	Time	miss rate	false recognition rate
Pixel	13116	<1ms	0%	-
Wavelet	655	<2ms	< 0.02%	-
SVM	7	<100 ms	<14%	<0.001%

Table 3 – Performance of the techniques applied to extra-urban driving scenarios. Rates are relative to the number of windows processed at each stage by the relevant technique. Pixel-level techniques are effective to discard a wide number of windows that are not likely to contain a pedestrian. Wavelet analysis further reduces the number of candidate windows by a factor 20. The final result from the SVM is highly accurate.

Technique	Number of processed windows per frame	Time	Miss rate	False recognition rate
Pixel	13116	<1ms	0%	-
Wavelet	3930	<2ms	< 0.03%	-
SVM	196	<100 ms	<15%	<0.13%

Table 4 – Performance of the techniques applied to urban driving scenarios. Average (not peak) processing times are higher in this case, due to the higher complexity of the scene.

Results in table 3 come from a test with 1000 frames selected as representative of night driving conditions in extra-urban roads, with and without pedestrians.

The recognition's system overall performance is very high (a false recognition rate below 0.001% with a miss rate below 0.02%). Results are worse in the urban scenario, where scenes are much more complex.
In both scenarios, most of the recognition failures are due to pedestrians which are about 80-100 meters far from the camera and are not apparent because they are not well lit by the car's NIR illuminators. On the other hand, false recognitions are mostly due to shadows and far road signs having shapes similar to pedestrians

4. Conclusions and future work

Development of intelligent systems to support safe driving is an important priority in the automotive research. The EDEL project has developed a NIR-based Night Vision System aimed at detecting pedestrians.
The procedure we have developed integrates well established signal processing and machine learning techniques, which we optimized for NIR night vision. Our recognition system performs a frame-by-frame analysis. After optimizing this fundamental step, we are going to implement a tracking module which exploits inter-frame temporal and spatial locality in order to improve accuracy of detection and reduce false recognition rates so to allow the future deployment into night vision systems on vehicles.
In the design approach adopted by EDEL project, technical tests will check the validity of the interpolation technique we have implemented in order to present the driver with a smooth animation (25 frames per second) despite the latency time which has peaks of up to 100 ms.

Acknowledgments

The authors thank Carlo Bellotti and Daniele Stellardo for their important theoretical and practical contribution in developing the algorithmic procedures.
The authors thank the EDEL project consortium, the project has been co-funded by the European Commission under the IST 5th Framework Programme, and to acknowledge the work of all the partners of the project, including U. Beutnagel-Buchner and D. Grimm, of Robert Bosch GmbH, M. Wueller of Hella, D. Kooss of the University of Karlsruhe, M. Mariani of the University of Siena, and R. Shelton of Jaguar.

References

[1] http://nncf.unl.edu/eldercare/info/seniordriving/nightdriv.html. *Tech Rep 6*., Nebraska Highway Safety Progr.

[2] Leibowitz H. W., Owens D. A, *We drive by night*, Psychology Today, January, pp. 55-58, 1986.

[3] Ward N.J., Wilde G.J.S, *Driver approach behaviour at an unprotected railway crossing before and after enhancement of lateral sight distances: An experimental investigation of a risk perception and behavioural compensation hypothesis*, Safety Science, 1-2, nn. 1-3, pp. 63-75, 1995.

[4] Tijerina L., Browning, N., Mangold S.J., Madigan E.F. and Pierowicz J.A., *Analysis of Reduced Visibility Crashes andPotential IVHS Countermeasures*, NHTSA Technical Report DOT HS 808 201, 1995.

[5] P. Knoll, *A NIR Based System for Night Vision Improvement*, Convergence 2002.

[6] EDEL project website: www.edel-eu.org

[7] D. Kooß (University Karlsruhe), F. Bellotti (University Genua), C. Bellotti (University Genua), L. Andreone (Centro Ricerche Fiat), *EDEL - Enhanced Driver's perception in poor visibilità*, Progress in Automobile Lighting (PAL) Symposium 2003, Darmstadt, September 2003.

[8] M. Mariani L. Andreone, S. Palmieri F. Tango, *User requirements and customer benefit analysis in the design of a novel driver support system for night vision,* Human Computer Interaction International (HCII), Crete, July 2003

[9] M. Mariani, L. Andreone, *Customer benefit in user centered design of a novel driver's vision enhancement support system,* Intelligent Transport Systems (ITS) 2004, Madrid, November 2003

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

[11] L. Andreone, P. C. Antonello, F. Bellotti, D. Grimm, M. Mariani, M. Wüller, J. Heerlein, *Development of new generation of driving support system for night vision enhancement the EDEL project,* Intelligent Transport Systems ITS 2005, Praha.

[12] C. Bellotti, F. Bellotti, L. Andreone, and M. Mariani, *Developing a Near Infrared Based Night Vision System,* IEEE Intelligent Vehicle Symposium, Parma, June 2004.

[13] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[14] Anguita, D., Ridella, S., Rivieccio, F., Zunino, R., "Quantum optimization for training support vector machines". *Neural Networks* 16(5-6), 2003

[15] C. Papageorgiou and T. Evgeniou and T. Poggio, *A trainable pedestrian detection system*, In Intelligent Vehicles, pp. 241--246, October 1998.

[16] T. M. Mitchell, *Machine Learning*, McGraw-Hill. 1997

[17] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio, *Pedestrian Detection Using Wavelet Templates*, Conference on Computer Vision and Pattern recognition, CVPR 1997: 193-199

[18] Mallat S.G., *A theory for multiresolution signal decomposition: the wavelet representation,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7), pag. 674-693. 1989.

[19] Z.R. Struzik, *Oversampling the Haar Wavelet Transform*, Technical report INS-R0102, CWI, Amsterdam, The Netherlands, March 2001.