

# Pedestrian Detection using SVM and Multi-feature Combination

M. A. Sotelo, I. Parra, D. Fernández, E. Naranjo

**Abstract**—This paper describes a comprehensive combination of feature extraction methods for vision-based pedestrian detection in the framework of Intelligent Transportation Systems. The basic components of pedestrians are first located in the image and then combined with a SVM-based classifier. This poses the problem of pedestrian detection in real, cluttered road images. Candidate pedestrians are located using a subtractive clustering attention mechanism based on stereo vision. A by-components learning approach is proposed in order to better deal with pedestrians variability, illumination conditions, partial occlusions, and rotations. Extensive comparisons have been carried out using different feature extraction methods, as a key to image understanding in real traffic conditions. A database containing thousands of pedestrian samples extracted from real traffic images has been created for learning purposes, either at daytime and nighttime. The results achieved up to date show interesting conclusions that suggest a combination of feature extraction methods as an essential clue for enhanced detection performance.

## I. INTRODUCTION

Pedestrian detection is a challenging problem in real traffic, cluttered environments, as long as pedestrian detection should perform robustly under variable illumination conditions, variable rotated positions and pose, and even if some of the pedestrian parts or limbs are partially occluded. In order to ease the pedestrian recognition task in ITS vision-based systems a candidate selection mechanism is normally applied. The selection of candidates can be implemented by performing an object segmentation either in the 3D scene or in the 2D image plane. The first solution requires the use of stereo vision [1] [2], while the second one tackles the problem of candidate selection using a single camera. Not many authors have tackled the problem of monocular pedestrian recognition [3] [4] [5]. In some cases, they have even provided crowd detection algorithms [6]. The advantages of the monocular solution are well known. On the one hand, it constitutes a cheap solution that makes mass production a viable option for car manufacturers. On the other hand, monocular systems are less demanding from the computational point of view, and ease the calibration maintenance process. On the contrary, the main problem with candidate selection mechanisms in monocular systems is that, in average, they are bound to yield a large amount of candidates per frame in order to ensure a low false negative ratio (the number of pedestrians that are not selected by the attention mechanism). Another problem in monocular

systems is the fact that depth clues are lost unless some constraints are applied, such as the flat terrain assumption, which is not always applicable. The variability of scale and size of candidates can derive in misdetection of small persons, especially children, if depth information is not available. In addition, as indicated in [2], objects similar to humans but belonging to larger structures, such as the photo of a person in a publicity advertisement, would be mistakenly considered as a pedestrian by a monocular system. These problems can be easily overcome by using stereo vision systems, although other problems arise such as the need of maintaining calibration and the high computational cost required to implement dense algorithms. Nonetheless, with recent hardware advances, real-time dense stereo vision becomes increasingly feasible for general-purpose processors [7]. Likewise, remarkable efforts have been done in order to ease the problem of calibration maintenance in stereo vision systems [8].

In this work, we present a full solution for SVM pedestrian detection at daytime, also applicable, although constrained, to nighttime driving. Concerning the various approaches proposed in the literature, most of them are based on shape analysis. Some authors use feature-based techniques, such as recognition by vertical linear features, symmetry, and human templates [4] [9], Haar wavelet representation [10] [11], hierarchical shape templates on Chamfer distance [5] [1], correlation with probabilistic human templates [12], sparse Gabor filters and SVM [13], graph kernels [14] and principal component analysis [15]. Neural network-based classifiers [16] and convolutional neural networks [17] are also considered by some authors. The use of motion and appearance has been previously used in [18]. In [19] an interesting discussion is presented about the use of binary or gray level images, as well as the use of the so-called hotspots in infrared images versus the use of the whole candidate region containing both the human body and the road. Using single or multiple classifiers is another topic of study. As experimentally demonstrated in this paper, and supported by other authors [3] [19] [2], the option of multiple classifiers is definitely needed. Multiple classifiers can cope with specialization at different depths, for the appearance of pedestrians varies significantly as a function of distance. Another crucial factor, not well documented in the literature, is the effect of pedestrians bounding box accuracy. Candidate selection mechanisms tend to produce pedestrian candidates that are not exactly similar to the pedestrian examples that were used for training, in the sense that on-line candidates extracted by the attention mechanism may contain some part of the ground or, on the contrary, may cut pedestrians feet,

M. A. Sotelo, I. Parra and D. Fernández are with the Department of Electronics, Escuela Politécnica Superior, University of Alcalá, Alcalá de Henares, Madrid, Spain [sotelo,parra,llorca@depeca.uah.es](mailto:sotelo,parra,llorca@depeca.uah.es)

E. Naranjo is with the Informatics Department at the Industrial Automation Institute. CSIC, Madrid, Spain [jnaranjo@iai.csic.es](mailto:jnaranjo@iai.csic.es)

arms or heads. This results in significant differences between candidates and examples. Some authors have demonstrated that the recognition of pedestrians by components is more effective than the recognition of the entire body [11] [20]. A key problem is still to find out the most discriminating features in order to significantly represent pedestrians. For this purpose, several feature extraction methods have been implemented, compared, and combined. While a large amount of effort in the literature is dedicated to developing more powerful learning machines, the choice of the most appropriate features for pedestrian characterization remains a challenging problem nowadays.

The rest of the paper is organised as follows: section II provides the system description, including the candidate selection mechanism. The implementation and comparative results achieved up to date are presented and discussed in section III. Finally, section IV summarizes the conclusions and future work.

## II. SYSTEM DESCRIPTION

This section includes the candidate selection mechanism, the pedestrian recognition system using SVM, and the set of feature extraction methods tested in the experiments.

### A. Candidate Selection

An efficient candidate selection mechanism is a crucial factor in the global performance of the pedestrian recognition system. On the one hand, the candidate selection method must assure that no miss-detection occurs, i.e., that all real pedestrians are effectively detected. On the other hand, candidates, that are usually described by a bounding box in the image plane, must be detected as precisely as possible, since the detection accuracy has a remarkable effect on the performance of the recognition stage. In order to extract information from the 3D scene most authors use disparity map techniques [16] as well as segmentation based on v-disparity [2] [21]. The use of disparity-based techniques is likely to yield useful results in open roadways, as long as pedestrians can be distinguished using depth disparity clues. However, depth disparity clues are unlikely to be useful for segmenting out pedestrians in city traffic due to the heavy disparity clutter. In addition, the computation of accurate disparity maps requires fine grain texture images in order to avoid noise generation. Otherwise, disparity based methods are prone to produce many outliers that affect the segmentation process. Concerning the v-disparity image, the information for performing generic obstacles detection is defined with a vertical line. This implies managing very little information to detect obstacles, which may work well for big objects detection, such as vehicles, but might not be enough for small, thin objects detection, such as pedestrians. Conversely, we propose a candidate selection method based on the direct computation of the 3D coordinates of relevant points in the scene. Accordingly, a non-dense 3D geometrical representation is created and used for candidate segmentation purposes. This kind of representation allows for robust object segmentation whenever the number of relevant points in the

image is high enough. A major advantage is that outliers can be easily filtered out in the 3D space, making the method less sensitive to noise. we propose a *Subtractive Clustering* method [22], which is a well known approach in the field of *Fuzzy Model Identification Systems*. The implementation of this method in our application is described in [20]. Figure 1 shows an example of candidate selection in different urban and extraurban environments at daytime and nighttime. Obviously, the number of candidates is bound to change depending on traffic conditions, as long as some cars are considered as candidates by the subtractive clustering algorithm.



Fig. 1. Candidate selection in a sequence of images.

In practice, a multiple candidate selection strategy has been implemented. The purpose is to produce several candidates around each selected cluster in an attempt to compensate for the effect of the candidate bounding box accuracy in the recognition step. Accordingly, five new candidates are generated for each candidate cluster, by slightly shifting the original candidate bounding box in the  $u$  and  $v$  axes in the image plane.

### B. Pedestrian Detection using SVM

The use of Support Vector Machine (SVM) in the field of pedestrian recognition has become a common approach for many researchers in the last years since it was first proposed in [10] and [11]. It is well known that SVM provides a method to calculate the hyperplane that optimally separates two high-dimensional classes of objects. In spite of that, there are other important aspects that need to be addressed when constructing a classifier, such as the global classification structure, the use of single or multiple cascaded classifiers, and the training strategy. The first decision to make implies the development of a holistic classifier against a by-components approach. In the first option, the complete candidate, described by a bounding box in the image plane, is applied to the input of the SVM classifier. On the contrary, the by-components approach suggests the division of the candidate body into several parts. Each pedestrian body part is then independently learnt by a specialized classifier in a first learning stage. The body local parts are then integrated by another classifier in a second learning stage. This allows

for detecting partially occluded pedestrians, as long as the system is modified to detect components of human body (e.g. head, torso or limbs). By using independent classifiers for each body part the learning process is simplified, due to the fact that a single classifier has only to learn individual features of local regions in certain conditions. Otherwise, it would be unrealistic to expect an acceptable learning result using a holistic approach, for the appearance of pedestrians in the scene presents a high intraclass variability (due to lateral and longitudinal movements, different shapes, pose, clothing, etc). After extensive trials, we propose a total of 6 different sub-regions for each candidate region of interest, which has been rescaled to a size of 24x72 pixels. This solution constitutes a trade-off between exhaustive sub-region decomposition and the holistic approach. The optimal location of the six sub-regions, empirically achieved after hundred of trials, has been chosen in an attempt to detect coherent pedestrian features, as depicted in figure 2. Thus, the first sub-region is located in the zone where the head would be. The arms and legs are covered by the second, third, fourth, and fifth regions, respectively. An additional region is defined between the legs, covering an area that provides relevant information about the pedestrian pose. This sub-region is particularly useful to recognize stationary pedestrians.

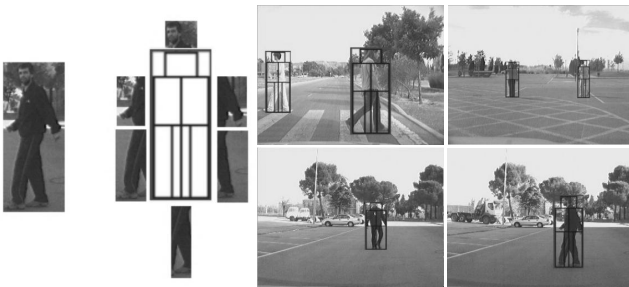


Fig. 2. Left: decomposition of a candidate region of interest into 6 sub-regions. Right: sub-regions examples in a set of images.

The use of a by-components approach demands the development of a cascaded classifier. A two-stage classifier is needed. In a first stage, separate SVM-based classifiers are trained using individual training sets that represent a subset of a sub-region, as previously stated. Each SVM classifier produces a theoretical output between -1 (non-pedestrian) and +1 (pedestrian). This stage provides classification of individual parts of the candidate sub-regions. In a second step, the outputs of all classifiers are merged in a simple classifier that makes a decision based on a distance criterion, in order to yield the final classification result. A major issue is the effect of illumination conditions. It is clear that daytime and nighttime samples must be compulsorily separated in order to create multiple specialized classifiers. The nighttime classifier can be reasonably expected to operate correctly only in very short distances (below 8 m) for non-illuminated areas, where pedestrians can be appropriately illuminated by the car beams (infrared images would be needed in order

to achieve long range detection, as mentioned in section I). Nonetheless, nighttime pedestrian detection can be done up to 20-25 m in illuminated areas. The effect of pose must also be taken into account as a significant source of variability in pedestrians appearance. Most pedestrians appear standing in the surrounding of the vehicle, either walking or simply in a stationary position. The differences between walking and stationary pedestrians are clear. There are even some remarkable differences between pedestrians moving laterally, with regard to the vehicle trajectory, and those moving longitudinally. Pedestrians intersecting the vehicle trajectory from the sideways are usually easier to recognize since their legs are clearly visible and distinguishable. In fact, some authors have proposed two separate SVM classifiers according to this statement [2]. A more complicated case occurs when a pedestrian crouches or bends down. The appearance of pedestrians in those cases is dramatically different from that of standing pedestrians. Changes due to different clothing also contribute to further complexity in the variability problem. Thus, large skirts and coats make pedestrians look very different from those in trousers and suits. Likewise, pedestrians bringing trolleys or bags make the recognition problem even more difficult. Had it not been enough, pedestrians legs are not always visible in the image, specially when pedestrians are very close to the vehicle. This is a critical case of great importance for pre-crash protection systems. In our approach, three factors are considered key: development of specialized SVM classifiers, design of an appropriate training strategy, and selection of sufficiently discriminant features. Trade-offs have to be made between accuracy of classification and processing time. For this purpose, we decided to create two separate SVM classifiers performing pedestrians learning at daytime and nighttime, respectively. Variability due to pose, clothing, and other artifacts is handled in two ways. On the one hand, adequate training databases are created containing as many representative cases as possible. In this stage, pedestrians in different pose (standing, walking, ducked) and clothing (coats, jeans, etc) are included in the database, as well as pedestrians with handbags and other artifacts. On the other hand, an appropriate selection of discriminant features is carried out, as explained in the next section. Additionally, we have performed an optimal kernel selection for the SVM classifiers. For this purpose, we used a small training set for which 4 different kernels were tested. The gaussian kernel was finally chosen as the optimal one after the trials.

### C. Feature Extraction Methods

The optimal selection of discriminant features is an issue of the greatest importance in a pedestrian detection system, considering the large variability problem that has to be solved in real scenarios. A set of features must be extracted from each sub-region and fed to the classifier. The choice of the most appropriate features for pedestrian characterization remains a challenging problem nowadays, since recognition performance depends crucially on the features that are used to represent pedestrians. Legs and arms are long

elements that tend to produce straight lines in the image, while the torso and head are completely different parts, not so easy to recognize. This statement, although based on intuition, suggests the combination of several feature extraction methods for the different subregions into which a candidate is divided. Accordingly, we have tested a set of 7 different features extraction methods; Canny image: the result obtained after applying a Canny filter to the region of interest is directly applied to the input of the classifier; Haar Wavelets, originally proposed for pedestrian recognition in [10]; Gradient magnitude and orientation; Cooccurrence matrix [23]: The resulting size of the feature vector depends on whether the cooccurrence matrix is computed over the original gray level image or over a binary one (after applying the Canny operator). In this work, we have tested both; Histogram of intensity differences: it computes the relative frequencies of the intensity differences between neighbouring pixels along four orientations. The image is normalized to 128 gray levels. This generates four 128-lengthed vectors per sub-region; Normalized histogram (HON); Number of Texture Unit [24]. The local texture information for a pixel can be extracted from a neighbourhood of 3x3 pixels, that represents the smallest complete unit of the texture. The corresponding texture unit is defined by a set containing the eight elements that surround the pixel under consideration. The NTU process generates a feature vector the same size as the region of interest.

### III. EXPERIMENTAL RESULTS

The system was implemented on a Pentium IV at 2.4 Ghz running the Knoppix GNU/Linux Operating System. Using 320x240 pixel images, the complete algorithm runs at an average rate of 20 frames/s, depending on the number of pedestrians being tracked and their position. The average rate has a strong dependency on the number of pixels being matched because of the correlation computational cost, which consumes, in average, 80% of the whole processing time. The candidate selection system has proved to be robust in various illumination conditions, different scenes and distances up to 25 m. The quality of the classification system is mainly measured by means of the detection rate / false positive ratio. These two indicators are graphically bounded together in a Receiver Operational Curve (ROC). Once candidates are validated by the SVM classifier, a kalman filter tracks the position of each pedestrian in the 3D scene. We created several training and test sets containing thousands of positive and negative samples (pedestrians and non-pedestrians, respectively) in different situations. Positive samples (pedestrian samples) were extracted from recorded images acquired in real traffic conditions. Training sets were created both at daytime and nighttime using the TSetBuilder tool [25], specifically developed in this project for this purpose. By using the TSetBuilder tool different candidate regions were manually selected from the image on a frame-by-frame basis. Special attention was given to the selection of non-pedestrian samples. Accordingly, negative samples (non-pedestrian samples) in the training sets were neither

randomly nor manually selected. The candidate selection mechanism proposed in section II-A was used instead to automatically produce the negative training samples. The use of this mechanism endows the process with a strong realistic component. In total, the training sets contain up to 30.000 samples, while the test sets amount up to 13.050 samples.

#### A. Combination of feature extraction methods

For this experiment, the training and test sets were designed to contain 10.000 and 3.670 samples, respectively. All samples were acquired in daytime conditions. Classification performance can be improved by combining different feature extraction methods for different candidate sub-regions. The best performing features for each sub-region are then combined in a second classifier, instead of applying the same feature extractor to all 6 sub-regions. Figures 3.a to 3.f show the ROC curves for each separate sub-region after computing the 7 pre-defined features. A detailed observation of this figure reveals that NTU and Histogram prove to be the most discriminant features for head and arms, while HON and Canny seem to perform the best for legs. Although their performance is similar, HON is preferred as long as it provides smaller feature vectors that result in faster execution time. Similarly, the area between-the-legs is best recognized by NTU. In principle, there seems then to be an optimal feature extraction method for each candidate sub-region. The combined used of optimal features leads to a clear increase in the overall classifier performance with regard to individual feature extractors, as depicted in figure 4.a, where a Detection Rate (DR) of 99.1% is achieved for a False Positive Rate (FPR) of 2%. These results improve the performance of Canny's detector, which is the best performing feature extractor, exhibiting a DR of 95% at a FPR of 2%.

Another comparison has been studied in order to analyze the influence of the second-stage classifier that combines the information delivered by the six specifically trained SVM models. In a first approach, we have used a simple distance criterion (i.e. distance to the hyperplane separating pedestrians from non-pedestrians) that computes the addition of the 6 first-stage SVM outputs and then decides the classification by setting a threshold. Another option has been tested by training a two-stage SVM, using the same training and test sets described at the beginning of this section. The results achieved up to date show that the simple-distance classifier outperforms the two-stage SVM classifier, as depicted in figure 4.b. For this purpose, the results attained for the 2 best performing feature extractors (Canny and Gradient) are compared in 4.b for either the simple-distance classifier and the 2-stage SVM classifier. In either case, the performance exhibited by the simple-distance classifier is superior. Thus, a DR of 97% at a FPR of 2% is achieved using Canny's detector and simple-distance classification, while a DR of 95% at a FPR of 2% is obtained in the same conditions with the 2-stage SVM. As a conclusion, the combined use of by-components optimal feature extraction methods in a second distance-based classifier is proposed as a reliable

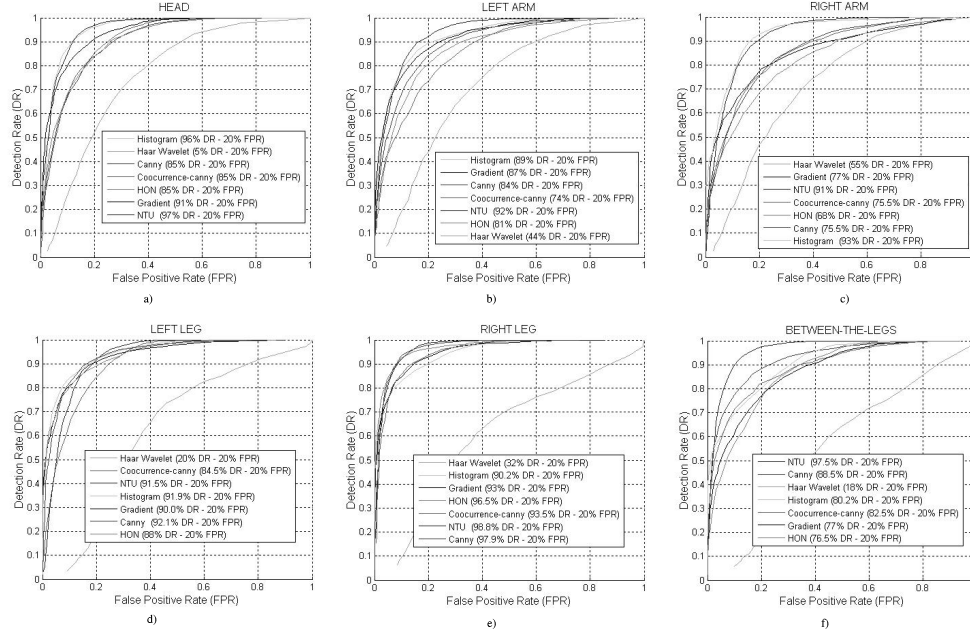


Fig. 3. ROC curves. a) Head. b) Left arm. c) Right arm. d) Left leg. e) Right leg. f) Between-the-legs.

solution for single frame pedestrian classification. Additional experiments were carried out in order to prove the need for separate training sets for day and night-time, as well as to test the influence of the candidates bounding box accuracy in classification performance.

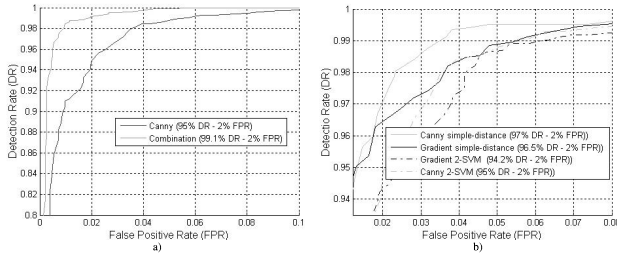


Fig. 4. ROC curves. a) Comparison between features combination and Canny's extractor. b) Comparison between simple-distance classifier and 2-stage SVM.

### B. Global performance

The performance of the global system is evaluated in a set of sequences recorded in real traffic conditions. Some of the sequences were acquired in urban environments and others in extraurban areas. The purpose of this evaluation is to assess the combined operation of the attention mechanism and the SVM-based classifier, including a multiframe validation stage using kalman filtering. The results obtained in the experiments are listed in Table I. For each row in the table the following information is provided: type of environment (urban or extraurban; the extraurban sequences were recorded in open roads as well as in the Campus of the University of Alcalá) duration of the sequence, number of detected

pedestrians (only pedestrians below 25 m are considered), number of missed pedestrians, and number of false alarms (F/A) issued by the system. Let us remark that the generation of false alarms is also subject to multiframe validation in order to avoid glitches. Accordingly, a false alarm takes place only when a false positive persistently occurs in time. The global system was implemented according to the following features: subtractive clustering candidate selection; by-components SVM using the 6 sub-regions described in section II-B; combination of features according to the description provided in section III-A; multiple SVM for day and night-time classification; multi-candidate generation to compensate for the bounding box accuracy effect; multiframe validation using kalman filtering.

The analysis of results reveals that performance is quite different in urban and extraurban environments. Thus, the pedestrian detection system exhibits a ratio of 11 false alarms in 20 minutes of operation in urban scenarios. This yields a ratio of 33 false alarms per hour. Similarly, the Detection Rate is 93.24% in urban environments, where 10 pedestrians were missed by the system. Let us clarify the fact that all missed pedestrians were partially occluded or completely out of the vehicle path. Concerning the 11 false alarms produced by the system, they were caused by 3 motorbikes, 2 trees, 4 lampposts and other urban furniture, 1 dustbin, and 1 fence. In all false alarm cases, there was a misclassified real object causing the false alarm. Conversely, all pedestrians were properly detected and tracked by the system in extraurban sequences, where images are not so heavily corrupted with clutter. Similarly, there were nor missed pedestrians neither false alarms in the tests carried out in these types of environments. Although these figures are still unacceptable

TABLE I  
GLOBAL PERFORMANCE EVALUATED IN A SET OF SEQUENCES

Environment	Duration	Detected	Missed	False alarms
Urban	20 min	138	10	11
Extraurban	10 min	24	0	0

for the deployment of a real pedestrian detection system in urban environments, the results described in this paper point to the possible application of a robust pedestrian protection system in roads and other open environments.

#### IV. CONCLUSIONS AND FUTURE WORK

In this work, we have carried out a comparative study of feature extraction methods for vision-based pedestrian detection. Candidates are first selected by an attention mechanism based on subtractive clustering and stereo vision. This helps reduce the number of false candidates and enhance the performance of the recognition stage. In order to reduce the variability of pedestrians, the learning process has been simplified by decomposing selected candidates into 6 local sub-regions that are easily learned by individual SVM classifiers. The so-called optimal features have been identified for each subregion and combined in a more discriminant by-components classifier. Likewise, the effects of illumination conditions and candidate size have been studied. Another important factor, usually disregarded by most authors, is the effect of candidates bounding box accuracy. Experimental results support the use of a multicandidate generation strategy, in order to assure the issuing of some well fitted candidate that matches the samples used for training. Finally, we have presented the global performance of the system described in this paper. Further improvement needs to be made before deploying a really robust vision-based pedestrian detection system for assisted driving in real traffic conditions. For this purpose, motion-based and position dependant features will be incorporated, aiming at enhancing the shape-based pedestrian detection algorithm developed in this work. Thus, additional classifiers are being introduced so as to detect motorbikes, urban furniture and so on, aiming at decreasing the false alarm rate.

#### ACKNOWLEDGEMENTS

This work has been supported by grants DPI2002-04064-C05-04 and DPI2005-07980-C03-02 from the Spanish Ministry of Education and Science and FOM2002-002 from the Spanish Ministry of Public Works.

#### REFERENCES

- [1] D. M. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: The protector system," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 13-18, Parma, Italy, June, 2004.
- [2] G. Grubb, A. Zelinsky, L. Nilsson, and M. Rilbe, "3d vision sensing for improved pedestrian safety," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 19-24, Parma, Italy, June, 2004.
- [3] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: single-frame classification and system level performance," in *Proc. IEEE Intelligent Vehicles Symposium*. pp. 1-6, Parma, Italy, June, 2004.
- [4] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proc. IEEE Intelligent Vehicles Symposium*. Dearborn, Michigan, USA, October, 2000.
- [5] D. M. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*. ISBN: 0-7695-0164-8, 1999.
- [6] P. Reisman, O. Mano, S. Avidan, and A. Shashua, "Crowd detection in video sequences," in *Proc. IEEE Intelligent Vehicles Symposium*. Parma, Italy, June, 2004.
- [7] H. Sunyoto, W. Mark, and D. M. Gavrila, "A comparative study of fast dense stereo vision algorithms," in *Proc. IEEE Intelligent Vehicles Symposium*. Parma, Italy, June, 2004.
- [8] T. Dang and C. Hoffmann, "Stereo calibration in vehicles," in *Proc. IEEE Intelligent Vehicles Symposium*. Parma, Italy, June, 2004.
- [9] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape-based pedestrian detection and localization," in *Proc. IEEE ITS Conference*. Shanghai, China, October, 2003.
- [10] C. Papageorgiou and T. Poggio, "A trainable system for object detection," in *Intl J. Computer Vision*. Vol. 38, No. 1, pp. 15-33, 2000.
- [11] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 23, No. 4, 2001.
- [12] H. Nanda and L. Davis, "Probabilistic template based pedestrian detection in infrared videos," in *Proc. IEEE Intelligent Vehicles Symposium*. Versailles, France, June, 2002.
- [13] H. Cheng, N. Zheng, and J. Qin, "Pedestrian detection using sparse gabor filter and support vector machine," in *Proc. IEEE Intelligent Vehicles Symposium*. Las Vegas, USA, June, 2005.
- [14] F. Suard, A. Rakotomamonjy, A. Bensrhair, and V. Guigue, "Pedestrian detection using stereo-vision and graph kernels," in *Proc. IEEE Intelligent Vehicles Symposium*. Las Vegas, USA, June, 2005.
- [15] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Puetzold, and C. Wohler, "Autonomous driving goes downtown," in *IEEE Intelligent Systems and Their Applications*. Vol. 13, pp 40-48, 1998.
- [16] L. Zhao and C. E. Thorpe, "Stereo and neural network-based pedestrian detection," in *IEEE Transactions on ITS*. Vol. 1, No. 3, September, 2000.
- [17] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian detection with convolutional neural networks," in *Proc. IEEE Intelligent Vehicles Symposium*. Las Vegas, USA, June, 2005.
- [18] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. ICCV*. pp 734-741, 2003.
- [19] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," in *IEEE Transactions on ITS*. Vol. 6, No. 1, March, 2005.
- [20] D. Fernández, I. Parra, M. A. Sotelo, L. M. Bergasa, P. Revenga, J. Nuevo, and M. Ocana, "Pedestrian recognition for intelligent transportation systems," in *Proc. ICINCO*. Barcelona, Spain, September, 2005.
- [21] R. Labayrade, C. Royere, D. Gruyer, and Aubert, "Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner," in *Proc. International Conference on Advanced Robotics*. pp. 1538-1543, 2003.
- [22] S. Chiu, "Fuzzy model identification based on cluster estimation," in *J. of Intelligent and Fuzzy Systems*. Vol. 2, No. 3, pp. 267-278, 1994.
- [23] R. M. Haralick, "Statistical and structural approaches to texture," in *Proceedings of the IEEE*. Vol. 67, No. 5, May, 1979.
- [24] L. Wang, "Texture unit, texture spectrum and texture analysis," in *IEEE Transactions on Geosciences and Remote Sensing*. Vol. 28, No. 4, pp. 509-512, 1990.
- [25] J. Nuevo, "Tsetbuilder tutorial, technical report," 2005, <ftp://www.depeca.uah.es/pub/vision/SVM/manual.pdf>.