Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Feature-based Multisensor Fusion Using Bayes Formula for Pedestrian Classification in Outdoor Environments

Laurence Ngako Pangop*, Frederic Chausse*, Sebastien Cornou** and Roland Chapuis*
*LASMEA, UMR 6602 CNRS
Universite Blaise Pascal
24 Avenue des Landais - 63177 Aubiere cedex, France
surname.name@lasmea.univ-bpclermont.fr
**Renault SAS
1 Avenue du Golf - 78288 Guyancourt cedex, France
sebastien.cornou@renault.com

Abstract— **Improvements on pedestrian classification reliability applying a Bayesian approach to multisensor data fusion is described in this paper. The proposed approach fuses information provided by a laser scanner and a monocular gray-level camera. The key is to combine in a probabilistic framework, the detecting capabilities of these sensors to classify pedestrians located along the vehicle trajectory. The approach comprises three processes: sensor data processing, tracking and classification. This work emphasizes the idea of redundancy and complementarity due to the different nature of the information provided by the laser scanner (a priori static outline and dynamic constraints of the pedestrian motion) and camera (patterns) to address pedestrian classification. Two contributions are presented: 1) estimation of likelihoods, p(feature|class), which is defined as the likelihood that a detected object belongs to a class (pedestrian or non-pedestrian) according to an observed feature; 2) likelihood combinations as well as past knowledge integration using Bayes formula.**

The performance of vision, laser and combined feature-based classifier through the application of a Receiver Operating Characteristics (ROCs) analysis is included. It was found that the combination of features results in an optimized system. Experimental results using real data (performed off-line) suggest that a Bayesian combination of features is an essential factor to enhance performance of pedestrian detection systems.

I. INTRODUCTION

Currently, other than energy considerations, the emphasis in automotive research is on active safety systems to provide the means to reduce the number of accidents involving vulnerable road users (such as pedestrians) as well as advanced means to assist the drivers before a collision occurs. Accidentology statistics indicate that despite recent advances in safety due to the introduction of passive safety systems, tighter legislation, etc. Pedestrian accidents still represent the second largest source of traffic-related injuries and fatalities, after accidents involving passenger cars. These numbers are a major concern among legislators and most likely directives will be decreed to legislate the use of safety means to reduce pedestrian casualties, particularly in Europe [6].

The detection and classification of pedestrians is a complex process. The data captured by onboard sensors will be searched in order to find features that indicate the presence of entities that might be pedestrians within the observed area. These features are then analyzed using different techniques in order to determine whether or not these represent a pedestrian, despite the multiple shapes, color or texture that these might have or the distance that they might be within the sensor field of view (FOV). Once their presence is detected, it is important to determine their position and distance to the ego-vehicle. Pedestrian information (i.e. presence, speed, direction of motion, etc.) and their relative location provide sufficient information for the driver or an assistive system to gain understanding of the relationship between the vehicle and the pedestrians and thus infer some information such as collision risk. The whole process has to be done with the vehicle in motion, hence in real time.

There are several pedestrian detection systems integrated as part of vehicle demonstrators, most use video cameras as the main sensor [12], [3] and [2]. Although pedestrians could be detected using different sensors, vision-based systems have several advantages that include a large FOV, good resolution, texture, etc. If used like a camera pair, it can also provide depth information. In addition camera cost and its potential uses for other perception functions for driving assistance systems give it a strong advantage over other types of sensors. Nevertheless, there are some drawbacks on the use of video cameras only for pedestrian detection. For this purpose the resolution of the images taken needs to be high. However,this implies higher computational costs due to the increased number of pixels to be processed. Vision depends very much on light conditions, when driving in urban areas these can change very much and consequently the degradation of image quality would exist. As a result, features will not be distinguishable and detection algorithms might fail. The onboard available computer power is also constrained, this limits the number of pixels that can be processes in real-time and consequently the camera resolution. The use of video camera is not sufficient for reliable pedestrian detection. There are conditions where the physics of the whole process means that cameras will fail, therefore if a pedestrian detection system is to be fielded in mass market vehicles, better reliability is needed.

In order to compensate for the short coming, from single sensors, by combining the capabilities of multiple sensors and fusing their data its possible to provide deployable solutions by car suppliers [5]. The evaluation of the state of the art in sensor-based pedestrian detection made by Gavrila [6] found that by combining a camera and laser range finder, it is possible to attain good reliability for pedestrian protection systems. Video cameras provide large FOV and high resolution, while laser scanners give good range information. Both sensors complement each other and hence their combined information should enhance overall detection performance [1], [11] and [4]. In [9] the authors exploit this complementarity, a laser scanner helps to locate obstacles in camera images. The detected objects are classified using vision-based algorithms. This association allows the development of a real-time pedestrian detection system. The time required to segment video frame is dramatically reduced as laser scanner segmentation is fast, due to the low resolution and fast frame rates. However their system performance is highly related to the reliability of the vision-based classifier. Indeed, the key component of any detection system is the classifier that makes the final decision. Its performance fixes the system accuracy.

This paper presents a Bayesian approach to combine laser- and vision-based features in order to enhance reliability of pedestrian classification in a dynamic environment. Laser scanner information is used both for segmentation and classification based on a priori known rules. The vision feature is the output of an Adaboost learning algorithm performed using Haar-like features like in [10]. Feature extracted from each sensor is converted into likelihoods, $p(\text{feature}|\text{class})$, which is defined as the likelihood a detected object belongs to a class (pedestrian or non-pedestrian) according to the considered feature. The probability of being a pedestrian is computed with respect to Bayes formula using these likelihoods and prior probabilities. For each detected object, our classifier outputs the probability of the detected being a pedestrian knowing the sensor features. A given object is considered as pedestrian if its probability is higher than a predefined threshold. Two sets of experiments have been made, first, simulations present the relevance of likelihood probability density functions proposed for modeling sensors. The ROCs plotted show that feature combination enhances the performance of the pedestrian classification. Finally experiments using real data are presented, these have demonstrated that the proposed approach with the integration of past knowledge improves the reliability of the pedestrian classification system.

A description of the primary processes needed before achieving pedestrian classification is presented in the next section. Section 3 addresses our Bayesian pedestrian classification approach. Section 4 presents results of simulations and real data experiments. Finally, section 5 presents the conclusions and future work.

II. FEATURES EXTRACTION PROCESSES

The aim of our endeavor is to design a system capable of detecting and tracking pedestrian crossing in front of the ego-vehicle, for deployment in mass market vehicles. The system should also estimate whether or not the trajectories of the detected pedestrians might intercept with the ego-vehicle trajectory.

Thus pedestrian classification is an essential part of such a system. Its accuracy is related to the reliability and rapidity of pedestrian protection system. Before performing our classification approach, some primary modules are required. They allow the computation of features used for classification. The fusion architecture used in our approach is shown in Fig.1. It is a feature-level fusion architecture as the sensor data is processed to compute likelihood which are combined by the mean of Bayes formula. Sensors are calibrated which enables a low-level attention control of the video by the laser scanner. Thus, the search space of the image processing can be significantly reduced to the ROI and scaling factor given by the object locations extracted from laser scanner data. Previous classification is performed from both sensor information. However, tracking is applied in laser scanner space only.

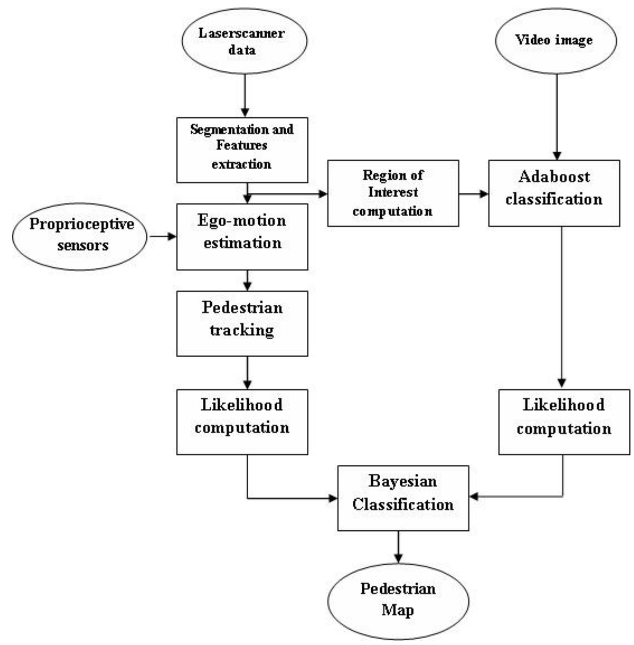

Fig. 1. Feature-level fusion architecture using a laser scanner and a video camera.

A. Primary Modules

The functions of the primary modules are : laser scan segmentation, feature extraction, tracking and Adaboost classification. They are described in detail in this section.

Segmentation and Feature Extraction

The segmentation step processes laser scanner data to identify sets of points (clusters) that probably belong to the same object according to their distance and spatial correspondence. The segmentation stage is not responsible for detecting targets but only suggests candidate objects for further classification. The separation between segmentation and classification permits the system to be quite permissive in the kind of depth data it accepts as input. In our case, this data is provided by the laser scanner, but we can consider to use stereo video frames.

Stable feature extraction is fundamental for the reliability and robustness of the whole system. The purpose is to extract relevant information from the scan points that constitute clusters. Thus assuming that the surrounding environment can be approximated by polygonal shapes, the line fitting is a suitable choice for object outline approximation. Good object characterization and data reduction can be achieved, the results are line segments per cluster. The Line fitting process is detailed in [8]. It can also be assumed that a pedestrian-like object would give one short (up to a set value) line segment, line fitting will help to reject objects over the set value. These are given long line segments or more than one segment (outlines with corners). The retained objects can be pedestrians, lamp posts or trees. Henceforth, the considered objects are only pedestrian-like objects. The features extracted are segment width and the middle point (x,y) of the segment. The segment width is used for classification task and the middle point is considered as object location. Fig.2 shows a result of the feature extraction with pedestrian-like and rejected objects from the laser scanner captured data.

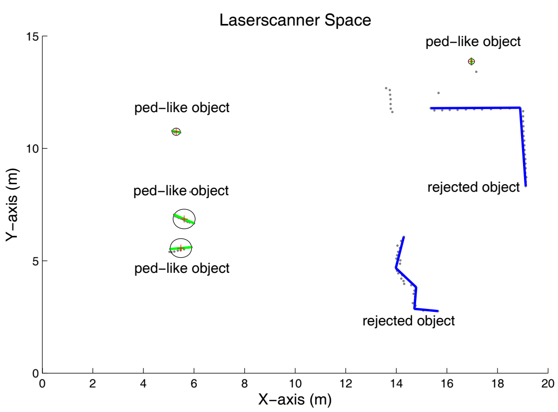

Fig. 2. Obstacles extracted from laser scanner data

Data Association and Tracking

In order for the tracker to measure the change in object position over time, we must determine which new segment corresponds to which existing track. For this purpose a data association algorithm is used. This is composed of two elements: a test to determine the compatibility between detected object and all tracked objects, and a selection criterion to choose the best matching among the set of compatible matchings. There have been some approaches to data association. The simplest method is the nearest neighbor (NN) algorithm which is a classical technique in tracking problems [7]. Here, the normalized squared innovation test is used to determine compatibility, and then the NN rule (smallest Mahalanobis distance) is used to select the best matching.

Pedestrian tracking is performed by a Kalman filter, it is assumed that a pedestrian model has constant velocity and white noise acceleration. The object location (x,y), previously computed, is used as the characteristic-point, i.e. the dynamic behavior of the object is described with respect to this point. The tracking algorithm estimates the state (position and velocity) of the object from the current observation and the predicted state (from the previous estimated state).

Adaboost Classification

This is a statistical model classifier, obtained by analyzing training set images which are multiple instances of pedestrians and ”negative” samples, i.e. images that do not contain pedestrians. Several features are extracted from each training sample. Adaboost learning algorithm is used to select relevant features to construct a pedestrian model. Then we also use Adaboost to classify each candidate window (ROI) by using these selected features. The key idea for the Adaboost algorithm is to build a (*strong*) classifier by combining the response of a set of simple (*weak*) classifiers, improving the performance that a complex classifier alone would have, details of the learning process are given in [10] for face detection. In this paper as we are more interested in showing that the combination of laser- and vision-based classifier leads to better results than the use of an individual classifier, then a cascade of classifier was not implemented like Viola and Jones [10]. The multiplicity of features suggested for different object detection tasks demonstrates that no one single type of feature might suffice for all possible detectors. Nevertheless as we are interested in showing the relevance of our Bayesian combination approach, we chose to use Haar-like feature only. This is due to its simplicity and efficiency on pedestrian detection [10]. Haar filter measures the contrast between two to four neighboring areas in the sub-window. See Fig.3 for illustrations of Haar filters.

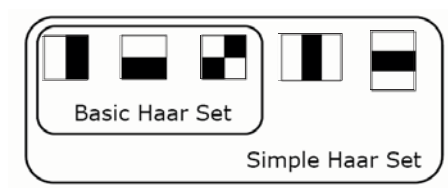

Fig. 3. Subset of the Haar-like features used in pedestrian detection

III. BAYESIAN PEDESTRIAN CLASSIFICATION

The classification stage identifies pedestrians from among the segmented objects by using feature vectors extracted from each candidate object, labeling them with a probability of being a pedestrian.

For a given feature vector $X = (x_1, x_2, \cdots, x_n)^T$, the Bayesian classifier computes $P(\text{ped}|X)$, the posterior probability that the feature vector X represents a pedestrian. The use of the Bayes formula, when all features are considered to be independent, yields the following decomposition :

$$P(\text{ped}|X) = \frac{1}{\beta} P(\text{ped}) p(x_1|\text{ped}) p(x_2|\text{ped}) \cdots p(x_n|\text{ped}) \quad (1)$$

Where $P(\text{ped})$ is a prior probability and β is a normalizing constant.

In our application $X = (width, speed, score)^T$, features extracted by the primary modules. Assuming $C \in \{\text{ped}, \text{no-ped}\}$, which represents pedestrian and non-pedestrian classes, we have to model the probability density functions (pdf), $p(\text{width}|C)$, $p(\text{speed}|C)$ and $p(\text{score}|C)$, to perform $P(\text{ped}|X)$.

Here $p(\text{score}|C)$ is estimated from the training process of Adaboost algorithm. Score follows a normal distribution that fits a histogram of scores obtained while characterizing the Adaboost learning results.

$$Score \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

Where μ and σ represent respectively mean and standard deviation of that normal distribution.

$p(\text{width}|C)$ and $p(\text{speed}|C)$ are defined respectively from a priori known static and dynamic restrictions of the objects under consideration, e.g. a $0.35m$ width object moving up to $2m/s$.

- **Pedestrian class**

The width of a pedestrian follows a normal distribution based on observation of some sequences of walking pedestrian.

$$Width \sim \mathcal{N}(\mu_w, \sigma_w^2) \tag{3}$$

The speed of a pedestrian follows a uniform distribution as a pedestrian can move at any speed within the bound $[V_{min}, V_{max}]$. Here $V_{min} = 0m/s$ represents standstill pedestrians and $V_{max} = 2m/s$ is the maximum speed authorized for the considered pedestrians.

$$Speed \sim \mathcal{U}(V_{min}, V_{max}) \tag{4}$$

- **Non Pedestrian class**

The non-pedestrian widths follow a uniform distribution as no a priori assumption is made. The minimum width (W_{min}) is given by the resolution of the laser scanner, in our case $0.05m$; the maximum width (W_{max}) is the maximum pedestrian length step measured when he is walking, here it is set to $1m$.

$$Width \sim \mathcal{U}(W_{min}, W_{max}) \tag{5}$$

The non-pedestrian speeds follow a normal distribution with zero-mean value as most of the objects of this class are static (e.g. Trees, lamp posts). The standard deviation, σ, is set to $0.1m/s$ to take into account the velocity estimation errors.

$$Speed \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

The conditional probabilities for the referred classes are plotted in Fig.4.

At each sample period, an object is classified as pedestrian, if a pedestrian probability $P(\text{ped}|X)$, is higher than a predefined threshold set according to the desired performance of the final classification. In practice, a false positive rate is defined and regarding ROC curve (false positive rate versus good detection rate) the related good detection rate is deducted as well as user-supplied threshold required to reach that performance.

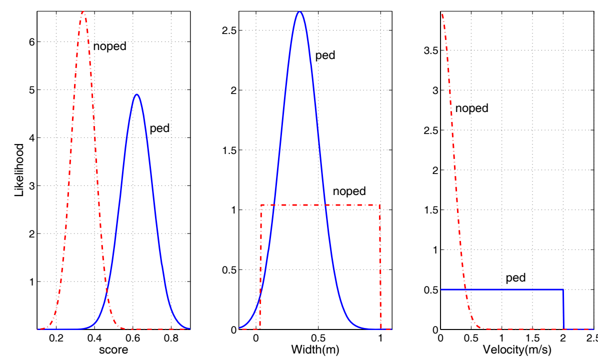

Fig. 4. Class-conditional probabilities density functions for the pedestrian and non-pedestrian classes, considering each feature.

Past knowledge is taken into account by replacing for each tracked object, the prior probability $P(\text{ped})$ by a posterior probability $P(\text{ped}|X)$ computed in the previous iteration. This improves the classification reliability, indeed this allows the reduction of spurious detection effects.

IV. EXPERIMENTAL RESULTS

Our experiments were based on real data sequences. In this Section we report our observations from experiments conducted on data obtained in a university campus. We used standard SICK planar laser scanner placed at $0.35m$ above ground level, targeting the leg height of a walking person; and a gray-level camera. A basic calibration of both sensors has been achieved which enables to project the location of the objects detected by the laser scanner onto the camera space. These experiments were performed in a semi-structured environment. One pedestrian was walking in front of the ego-vehicle which was stopped. The aim of this was to measure the relevance of our combination approach.

As part of the experiment, a basic Adaboost algorithm was constructed from 100 ”weak” classifiers. This detector has been trained on a database of 516 frontal and rear images of people. People are centered and are of approximatively the same size, in windows of 128×64 pixels. The non-pedestrian training set counted for 2550 natural scenes images of size 128×64 pixels, these did not have any people.

Experiments with simulated data were performed to compare the performance of the classifier using combined features (width, speed and Adaboost score) with the classifiers based on vision or laser data only. The resulting ROC curves are shown in Fig.5. In the first experiment we used only laser features (width and speed) for classification. The second curve represents the vision-based classifier. Regarding the first two curves, vision features are more discriminative than laser features for pedestrian classification. This is probably because the laser models are based on heuristics and the vision model is obtained through a training process. The final experiment used both vision and laser features. The combined features improve classification results. For a false positive rate of 1%, good detection rate is 75% for laser-based classifier, 96% for vision-based classifier and 98% for classifier using both laser and vision features for the set of data used.

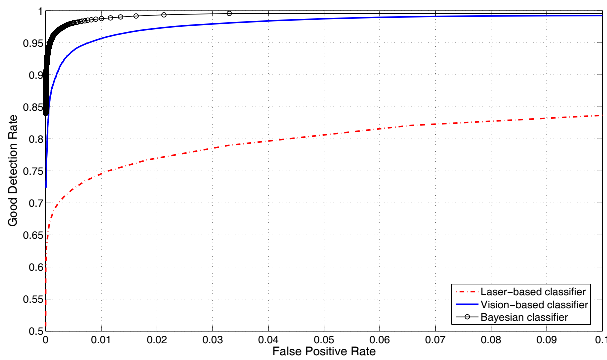

Fig. 5. Optimal ROC curves. They illustrate that feature combination improve pedestrian classification result.

Our Bayesian classifier has been tested on real data and the results compared with those from single sensor classifiers. The results from a classification are shown in Fig.6. The box to the left of the pedestrian represents some of the false alarms generated by the other classifiers. In Fig.7, the pedestrian trajectory is given relative to the vehicle location $(0,0)$ in the laser scanner space. It is to note that for this scenario, the laser sensor had mistakenly taken a lamp post as a pedestrian (See Fig.8).

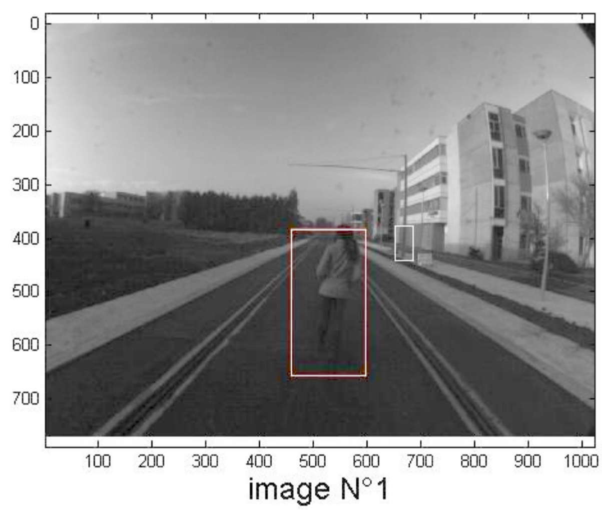

Fig. 6. A snapshot of the scene. Classified pedestrian and non-pedestrian (lamp post) are surrounded by the white boxes.

For each object, the probability of being a pedestrian is computed at each time using the Bayes formula. Prior probability is set equal for the two classes, for the single sensor classifier, this prior probability is constant throughout all the experiment. It means that no update is performed. In

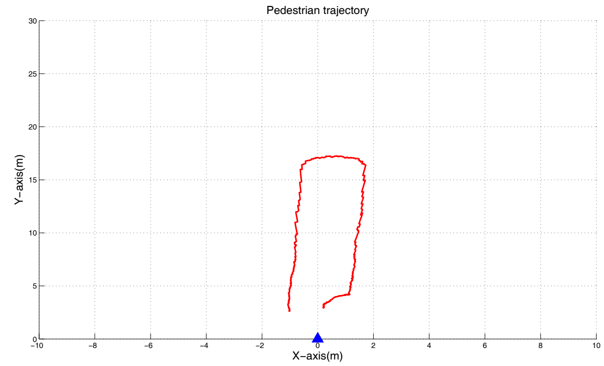

Fig. 7. Laser scanner space. Pedestrian trajectory

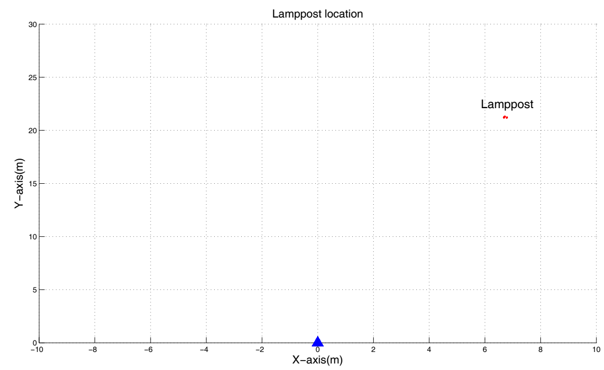

Fig. 8. Laser scanner space. Lamp post location

our approach, the integration of past knowledge involves the replacement of the prior probability by the posterior probability of the previous iteration. The features extracted and probabilities of being a pedestrian computed are shown in Fig.9 for a pedestrian and on Fig.10 for a non-pedestrian. The calcultated pedestrian probability must always be the high as the possible. However, we can notice that according to the vision or the laser based classifier, that probability is too noisy. This is because our implementation of the Adaboost algorithm is basic, the positive training set contains only front and rear pedestrians. Hence, the images of the side view pedestrian are hardly classified by this algorithm. Moreover, the laser classification is based on heuristics. We considered a pedestrian, an object which moves. Then when pedestrian speed decreases or reaches zero, the laser classifier considers it as a non-pedestrian. However, individual classifiers do not work very well, it is important to note that our Bayesian classifier is robust enough and gives very good results. This is confirmed in the case of non-pedestrians. The probability computed with the Bayesian classifier is almost 0 for lamp post, whereas when using the laser classifier response we can take it as a pedestrian. This is because the estimation of its velocity is erroneous. Finally, the drops which occur sometimes on the Bayesian result as shown in Fig.9 is because the pedestrian is too far from the experimental vehicle and hence the vision classifier does not take enough into account scale variation.

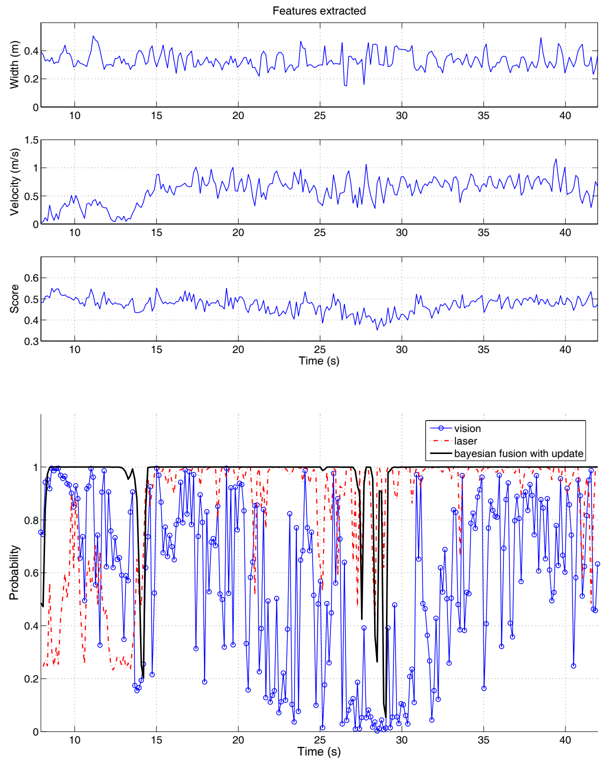

Fig. 9. The pedestrian classification result. Features extracted from sensors are too noisy. This involves the sensor-based classifiers to be not really efficient. However, our Bayesian classifier is more robust and reliable.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a Bayesian approach combining features extracted from a laser scanner and a monocular camera to classify pedestrians for a vehicle perspective. We have demonstrated how feature-level fusion and integration of past knowledge improve the reliability of classification. For each object, our classifier outputs probability of this object being a pedestrian. This classifier tested on real data provides very good preliminary results.

We are working towards the enhancement of the performance of our Bayesian classifier by improving the Adaboost learning process and the laser scanner data processing. Next, we are going to address the problem of data unavailability or delay. The whole method resides from the fact that combined features are independent due to the different nature of the information they provide. However, if we have to combine features from only one sensor, correlation problems can occur. We have to manage it by taking it into account in Bayes formula.

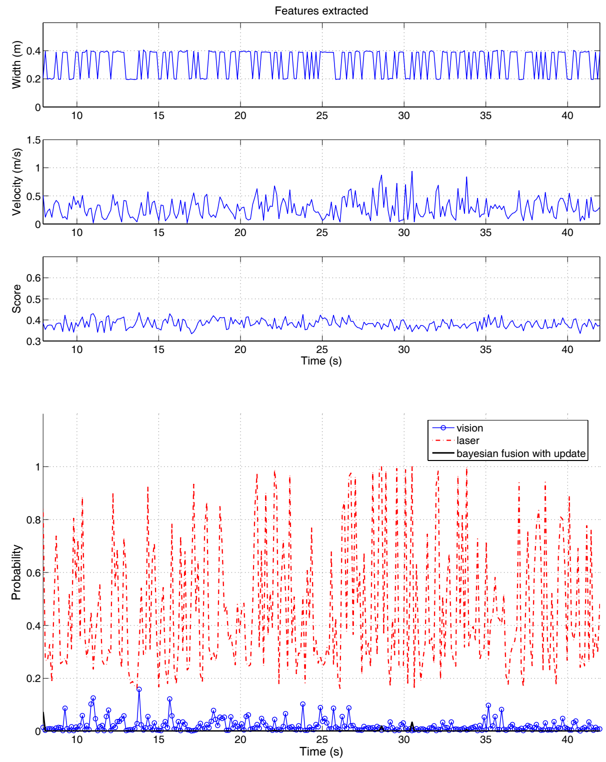

Fig. 10. The non-pedestrian classification result. Features extracted from sensors are too noisy. This involves the sensor-based classifiers to be not really efficient. However, our Bayesian classifier is more robust and reliable.

REFERENCES

[1] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, Shape-based pedestrian detection and localization. *In Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, pp. 328-333, Shangai, China, Oct. 2003.

[2] A. Broggi, M. Bertozzi, A. Fascoli, and M. Sechi, Shape-based pedestrian detection. *In Proc. IEEE Intelligent Vehicle Symposium 2000*, pp. 215-220, Dearborn, USA, 2000.

[3] U. Franke, D.M. Gavrila, S. Grzig, F. Lindner, F. Paetzold, and C. Whler, Autonomous driving goes downtown. *IEEE intelligent Systems*, 13(6), pp. 40-48, 1998.

[4] KC Fuerstenberg and U. Lage, Pedestrian Detection and Classification by Laserscanners. *9th EAEC International Congress*, Paris, France, June, 2003.

[5] T. Gandhi and M. Trived, Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies. *In Proc. of the IEEE Intelligent Transportation Systems Conference*, pp. 976-981, Toronto, Canada,2006.

[6] D.M. Gavrila, Sensor-based Pedestrian Protection. *IEEE Intelligent Systems*, vol. 16, NR.6, pp. 77-81, 2001.

[7] J. Guivant, E. Nebot, and S. Baiker, Simultaneous localization and map building using natural features and absolute information. *Robotics and Autonomous Systems*, vol. 40, pp. 79-90, 2002.

[8] A. Mendes and U. Numes, Situation-based Multi-target Detection and Tracking with Laserscanner in Outdoor Semi-structured Environment. *In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004.

[9] M. Szarvas, U. Sakai and J. Ogata, Real-time Pedestrian Detection Using Lidar and Convolutional Neural Networks. *In IEEE Intelligent Vehicles Symposium 2006*, Tokyo, Japan,2006.

[10] P.Viola and M.Jones, Rapid object detection using a boosted cascade of simple features.*In IEEE Conference on Computer Vision and Pattern Recognition*,2001.

[11] P.Viola, M. Jones and D. Snow, Detecting Pedestrians Using Patterns of Motion and Appearance. *In Proc. of the International Conference on Computer Vision (ICCV)*. Nice, France,2003.

[12] l. Zhao and C. Thorpe, Stereo- and neural network-based pedestrian detection. *IEEE Transaction on Intelligent Transportation Systems*, 1(3), 2000.