Asynchronous Bayesian Algorithm for Object Classification: Application to Pedestrian Detection in Urban Areas

Laurence NGAKO PANGOP, Frederic CHAUSSE and Roland CHAPUIS
LASMEA, UMR 6602 CNRS, Universite Blaise Pascal
24 Avenue des Landais
63177 Aubiere cedex, France
Email: firstname.lastname@lasmea.univ-bpclermont.fr

Sebastien CORNOU
Renault SAS
1 Avenue du golf
78288 Guyancourt cedex, France
Email: sebastien.cornou@renault.com

Abstract—This paper considers the problem of the classification of objects observed by vehicle embedded sensors. We propose a general architecture and an algorithm to perform multisensor fusion for the classification purpose. The proposed solution has to be robust and flexible. The robustness is essential because this system is for safety applications. The flexibility is ensured by a modular architecture alongside with a feature-level data fusion algorithm. This algorithm is based on the Bayesian formalism. Therefore, the classification of a detected object becomes the computation of a recognition probability given a set of features. This formalism allows an asynchronous data fusion, that is to say each sensor information is taken into account as soon as it is available. The proposed approach is applied to combine information from a laser scanner and video camera in order to detect pedestrians moving in the equipped vehicle path. The experimental results present pedestrians detected in urban areas. These results confirm the effectiveness of the proposed approach.

Keywords: Multisensor fusion, Bayesian algorithm, Pedestrian detection and tracking, Perception system.

I. INTRODUCTION

The problem of object classification is defined as follows: given a set of observations of an object and a predefined object class, the classification involves determining whether or not the object belongs to this class. This problem is inherently different to “state” estimation problems [1], [2] because the class of objects is not a metric parameter; it is likely to be pedestrian, car, bicycle or truck. The non-metric nature of the object class has given a rise to several alternative approaches to the representation, quantification and manipulation of uncertainty in class identity [3].

Statistical methods, fuzzy set theory and adaptive neural network methods are some of the methods that have been applied to this problem. However two methods, both statistically based, remain the most popular: Bayesian methods [4], [3] and Dempster-Shafer methods [5]. These two approaches are both commutative and associative, so results are independent of the order in which data is received and incorporated. Dempster-Shafer method was developed to overcome a perceived inability of the Bayesian formalism to adequately represent a lack of knowledge as opposed to a reasoned ambivalence. However this is at the expense of a computationally more intensive algorithm and has been shown [3], in general, to be ineffective at overcoming this deficiency. We maintain that with the correct and reasoned formulation, Bayesian method offers the most efficient, scalable, and theoretically justifiable method of managing uncertainty in situations such as object classification. We, therefore, propose to use such a formalism to develop our multisensor fusion algorithm.

This paper presents a modular multisensor architecture and develops an asynchronous Bayesian algorithm for object classification. Although designed for a general purpose, the eventual implementation of this algorithm is for an on-board pedestrian protection system. The requirements of such a system are robustness and flexibility. The robustness is essential because the system is to be deployed in vehicle for safety applications. It is obtained by the use of a variety of different sensor types [6] ensuring the complementarity and redundancy in the sensing capability of the classification system. To obtain the full benefit of using several sensors, a feature-level fusion architecture is employed. This architecture enables at the same time to have a flexible system. Indeed our feature-level fusion architecture is a network of perception modules connected to a supervisor. Locally computed features are sent to the supervisor which computes one probability for each tracked target. This probability represents the confidence of the system to assign a class to a target. The supervisor manages the perception modules and processes tasks which are not directly related to low-level data processing, namely the attention area computation, multi-target tracking and data fusion. This decoupling is made possible by a standardization of the inputs and outputs of the perception modules. The inputs of each perception module are the positions of the tracked targets. Its outputs are the likelihoods representing the features extracted from the sensor data associated with this perception module. Therefore, the data fusion issue becomes the combination of likelihoods done by the mean of the Bayes theorem. To reduce latency, each sensor data is incorporated as soon as it arrives. This data fusion algorithm is characterized as asynchronous, thus.
An overview of the proposed feature-level multisensor system is presented in the next section. The module roles and the links (I/O) between data-based modules and the supervisor are detailed. Section 3 introduces the asynchronous Bayesian algorithm for the purpose of object classification. For clarity, we initially develop the algorithm for a hierarchically organized multisensor system and then show how it can be extended for an asynchronous configuration. An implementation of the proposed approach for pedestrian classification is presented in section 4. The experimental results, performed on real data, confirm the effectiveness of our approach. Pedestrians are well detected even in cluttered environments. Finally, the main conclusions are summarized in section 5.

II. FEATURE-LEVEL DATA FUSION ARCHITECTURE

This object classification system is likely to be deployed for safety applications in vehicles, therefore the reliability and robustness to partial system failures are of prime importance. One way to increase the reliability is to introduce a redundancy in the number of sensors employed, thereby increasing the ability of the system to withstand loss of individual sensing units [7]. Moreover, a multisensor system also allows a more complete description of scenes thanks to the combination of information from different types of sensors [8], [9]. However, the use of multiple sensors gives rise to two related problems: how to connect all the sensors together (organization and architecture) and how to fuse the data obtained from these disparate sources (data fusion). In the present case, the data fusion mechanism results directly from a development of an object classification algorithm based on a Bayesian formalism. The system constraints are: flexibility ensuring the deployment of the same architecture "easily" in any vehicle regardless its sensing system; and robustness to a partial system failure. To fulfill these constraints, we propose a modular organization. Fig.1 shows the module layout. The perception modules convert the discriminative features into likelihoods which are transmitted to the supervisor for the calculation of classification probabilities. This architecture decouples low and high level processing as each data-based module performs locally its own processing.

A. Object segmentation module

This module serves to detect and localize the objects or targets in the scene. It is composed of a sensor data together with a segmentation algorithm. The sensors often used are radar, laser scanner or stereovision sensors. The segmentation algorithms are sometimes based on clustering processes [10]. The input of this module is an attention area which delimits the zone where relevant objects can be detected. For instance the attention area can be a road lane in case of the tracking of a preceding vehicle for an ACC (Automotive Cruise Control) application; or for a pedestrian detection system an area in front of the host vehicle where dangerous pedestrians can be located (the shape and size of this zone vary with the host vehicle width, speed and yaw rate, a detection time and the maximum speed of pedestrians under consideration). This attention area help speed up the system and reduce noise sources. The output of this module are the positions of the segmented objects located in the attention area. The output of this module is the list of detected objects which is transmitted to the supervisor for the classification task. Several sensor data can be used for the segmentation to refine it.

B. Perception module

Various features can be extracted from sensor data for the classification purpose: size and velocity from laser scanner data [11]; symmetry from video image [12]; a score, output of a learning algorithm, from the video image [13]. Basically, a perception module is composed of a sensor data and a feature extraction algorithm alongside the likelihood function associated to this feature. Several discriminative features can be extracted from the same data, therefore several perception modules can share the same sensor data. A major specificity of our architecture is that the perception modules inputs and outputs are standard. That ensures the flexibility of the overall system. The inputs are positions of the relevant targets computed by the segmentation module. While the outputs are the likelihoods obtained from each feature. These likelihoods are sent to the supervisor for the classification task. This standardization enables to add or remove "easily" sensors.

C. Supervisor

The supervisor is the core of our classification system. It processes high-level tasks (not directly processes from sensor raw data). Our motivation was to designed a multisensor architecture with a central part linking up the data-based modules. For the object classification concern, these high-level processes are: the attention area computation, the object tracking and the Bayesian fusion. This module can also manage the perception modules in case that the computing resources are limited (as usual with the automotive computers). The management notion will not be discussed in this work because our first objective is to propose a multisensor fusion approach (architecture and
data fusion formalism) which enables object classification. The
details on the calculation of the attention area can be found
in [14]. The object tracking can be performed by the mean
of the well known Kalman filter. In our pedestrian detection
system, a multi-target tracking is implemented. The Bayesian
fusion results in the combination of likelihoods (features)
provided by the perception modules. The data fusion algorithm
is developed in the following section.

III. ASYNCHRONOUS DATA FUSION ALGORITHM

The fusion problem addressed here is basically an object
classification issue. The objective is to collect observations
on objects in a scene and estimate the probability of these
objects to belong to a considered object class (e.g. pedestrian,
car, bicycle or truck). A hierarchically organized multisensor
system is first developed. While this in itself is not a new
idea [3] it allows us to introduce basic concepts, define
notation, and build a framework from which we can derive
the asynchronous data fusion. The algorithm is derived from
the Bayes’ theorem.

Let’s consider a network of n sensors or in our case n
perception modules, each connected to the supervisor (Fig.1).
The classification system attempts to determine if a detected
object belongs or not to an object class, \( H_a \in H \). H is a set
of classes. For our purpose, H has only two possible values
H = \{class, \overline{class}\}. Each perception module i takes an observation
\( Z_i \) and then calculates the likelihoods \( p(Z_i|H_a) \) of the object
being of the type \( H_a \). These likelihoods are locally computed.
The local likelihoods are passed to the supervisor which fuse
them with any prior probability, \( P(H_a) \), to arrive at a global
consensus posterior probability, \( P(H_a|Z) \), where \( Z = \bigcup_i \{Z_i\} \)
with \( i = 1, \ldots, n \). This posterior can then become the prior
used by the supervisor next to assimilate new likelihoods of
the same target.

A. Hierarchically organized data fusion algorithm

Relating the probabilities to the observations through Bayes’
theorem we obtain:

\[
P(H_a|Z) = \frac{p(Z|H_a)P(H_a)}{P(Z)}
\]  

(1)

where \( P(H_a|Z) \) is the posterior probability of a target being
of the class type \( H_a \) given the observations \( Z \).

\( p(Z|H_a) \) is the probability of a particular set of observations
\( Z \) knowing that the object is of class type \( H_a \). This term can
be also considered the likelihood of \( H_a \) being correct given
the set of observations \( Z \).

\( P(H_a) \) is the prior probability of the class \( H_a \) being correct.
\( P(Z) \) serves as a normalizing function, ensuring the poste-
rior probabilities sum to one over the set \( H \).

The crucial term in equation 1 is the likelihood \( p(Z|H_a) \).
Since each perception module i will only be able to supply
locally constructed likelihoods conditioned on its own ob-
ervation, \( Z_i \), evaluation of this term by the supervisor implies
combining the information received from all the perception
modules.

For a sensor system it is reasonable to assume that the
likelihoods from each informational sources \( p(Z_i|H_a) \), \( i = 1, \ldots, n \), are independent since the only parameter they have
in common is the state.

\[
p(Z|H_a) = p(Z_1|H_a)p(Z_2|H_a)\ldots p(Z_n|H_a)
\]  

(2)

Thus, equation 1 can be rewrittten:

\[
P(H_a|Z) \propto P(H_a) \prod_{i=1}^{n} p(Z_i|H_a)
\]  

(3)

\[
\sum_{a} P(H_a|Z) = 1
\]  

(4)

Update of Prior: The new posterior set becomes the priors
for the next data fusion cycle, ensuring continuity between
cycles of the algorithm. The prior can begin as noninformative
(in this case a uniform prior is noninformative) and be built
up at each time step \( k \):

\[
P^k(H_a) = P^{k-1}(H_a|Z)
\]  

(5)

B. Asynchronous data fusion algorithm

Equation 3 assumes that all the perception system likeli-
hoods are received before their assimilation by the supervisor.
We can say that the information are hierarchically organized.
The problem with this organization is that the results are
available only at the frequency of the lowest perception
module. However, for safety applications the lowest latency is
wished. In order to reduce the latency, we propose to assimilate
each likelihood at its arrival under the following assumption:
Assumption: at a given time the likelihoods from a perception
module are assumed to be temporally independent from all the
previous likelihoods provided by this perception module.

Under this assumption, equation 3 becomes

\[
P(H_a|Z) \propto P(H_a)p(Z_i|H_a)
\]  

(6)

In general, the previous assumption holds when in case of
important changes in the scene. For instance a change of the
point of view of the sensors or a variation of the scale factor
(for classified images). In our implementation, we assume that
this assumption holds because either the classified objects are
in motion or the host vehicle is moving. In other words, we
are in presence of dynamic scenes.

In case of static scenes and if the temporal correlations
are known, they have to be considered in the calculation of
posterior probabilities.

IV. PEDESTRIAN CLASSIFICATION SYSTEM

A. Implementation

The system architecture used in our approach is shown in
Fig.2. Sensors available here are: one laser scanner, and one
video camera. Recognition task is solved by fusing features
extracted from both sensors using Bayes’ theorem. Detection
and tracking are performed within the laser spawned space.
The two sensors are mutually calibrated, that enables a low-
level attention control of the video by the laser scanner.
Figure 2. Multi-module architecture using laser scanner and vision information for pedestrian detection and classification.

The asynchronous Bayesian algorithm has been implemented to classify the pedestrians. This multisensor system is designed to track and classify pedestrians along the host vehicle path. The classification results will help assess dangerous situation in order to avoid pedestrian accidents. In this application, \( H = \{ \text{pedestrian, non-pedestrian} \} \). The output of the system is a list of the locations of the dangerous pedestrian-like objects, their velocity and the probability that a detected object belongs to pedestrian class given a set of observations.

The sensing system used consists of a mono planar laser scanner and a gray-level video camera. Both are mounted in front of the host vehicle. Their field of view overlaps in the attention area. The laser scanner is approximately at 0.45m above the ground level, targeting at the pedestrian legs height. While the camera is behind the windscreen at the rearview mirror height.

In order to be fast, object segmentation is performed only in the laser scanner space. A calibration of both sensors allows objects detected by the laser scanner to be projected into the camera image.

The discriminative features used are the width, velocity and score (output of a learning algorithm achieved on video image), therefore \( Z = \{ \text{score}, \text{width}, \text{velocity} \} \). The width is calculated from the cluster extracted from the laser scanner. The velocity is estimated thanks to Kalman filter algorithm. The score is the output of an Adaboost algorithm using Haar-like features. The details of the features extraction processes can be found in [15].

In our implementation, equation 3 becomes:

\[
p(H_a|Z) \propto p(\text{score}|H_a)p(\text{width}|H_a)p(\text{velocity}|H_a)P(H_a)
\]  

\[ p(H_a|Z) \propto p(\text{score}|H_a)p(\text{width}|H_a)p(\text{velocity}|H_a)P(H_a) 
\]  

\[ \text{with } H_a = \text{pedestrian} \]

B. likelihood functions

Bayes’ theorem tells us how to manipulate probabilities, but it does not answer the question of how to assign these probabilities, which is discussed in this section.

The score-based likelihoods are estimated from the training process of the Adaboost algorithm. Score follows a normal distribution \( \mathcal{N} \) that fits histogram of scores obtained while characterizing the Adaboost algorithm.

\[
\text{Score} \sim \mathcal{N}(\mu, \sigma^2)
\]

(8)

where \( \mu \) and \( \sigma \) are respectively the mean and standard deviation of that normal distribution.

While width and velocity-based likelihood functions are based on heuristics.

The width of a pedestrian follows a normal distribution \( \mathcal{N} \) parametrized from observation of some sequences of pedestrians in motion.

\[
\text{Width} \sim \mathcal{N}(\mu_w, \sigma_w^2)
\]

(9)

The non-pedestrian widths follows a uniform distribution \( \mathcal{U} \) as no a priori assumption is made. The minimum width, \( W_{min} \) is the resolution of the laser scanner, in our case \( 5 \times 10^{-2} \)m; the maximum width, \( W_{max} \) is the maximum pedestrian length step measured, here it is set to 1m.

\[
\text{Width} \sim \mathcal{U}(W_{min}, W_{max})
\]

(10)

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

\[
\text{Velocity} \sim \mathcal{U}(V_{min}, V_{max})
\]

(11)

The non-pedestrian velocity follows a distribution plotted in Fig.3. This is chosen because the most of the objects one can confuse with the pedestrians are static.

Figure 3. Class-conditional probabilities density functions for the two classes considering each feature.
C. Experimental results

Gain of fusion

Experiments with simulated data were performed to compare the performance of the classifier using combined features (width, speed and score) with the classifiers based on vision or laser information only. The resulting ROC curves are shown in Fig.4. 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.

![ROC Curves](image)

Figure 4. Optimal ROC curves. They illustrate that feature combination improve pedestrian classification result.

Pedestrian detection results

The proposed pedestrian classification algorithm has been tested on a sequence with three pedestrians in the scene. The vehicle was in motion during this experiment. One of the pedestrian was in motion and the others were standstill. The first pedestrian was crossing the road in front of the vehicle. Fig.6 shows a snapshot of the scene at the beginning of the sequence, and Fig.5 is the corresponding top view of the laser scanner. Only one of the pedestrian was present in the attention area (the red zone) at that moment. In Fig.7 the features extracted when this pedestrian was located in the attention area are shown. Probabilities computed from these features in case of an hierarchical and asynchronous data fusion configuration are the same (Fig.8). In all the experiment, a target is considered as a pedestrian when its probability is greater than 0.9. The pedestrian in Fig.6 is not detected only at the first iteration mainly because of the false estimation of his speed. The system performance is not deteriorated by the use of the proposed asynchronous algorithm compare to the hierarchical data fusion algorithm. Fig.9 shows an image with all the pedestrians well classified by the system. The hard point remaining is the discrimination of motionless targets. In some configuration, the system fails in classification of static target like part of a vehicle (figure.10).

![Scene Observation](image)

Figure 6. A snapshot of the scene observed with the pedestrian located in the attention area surrounding by a rectangular box.

V. CONCLUSIONS

This paper introduces a feature-level multisensor fusion architecture and develops an asynchronous Bayesian data fusion algorithm for the purpose of object classification. The proposed architecture is flexible. We demonstrate that this architecture enables the classification system to enhance its recognition capability for instance by addition of new perception modules. This flexibility is made possible by the use of an asynchronous Bayesian algorithm for the data fusion. The results demonstrate that this algorithm does not deteriorate the optimal classification results, obtained when using an
hierarchically organized data fusion algorithm. Moreover, the proposed approach is sufficiently generic to enable any object classification (e.g. pedestrian, car, bicycle, truck). Only the feature processing and tracking have to be adjusted.

A real-time implementation of this approach for the pedestrian classification is in progress. It will help quantify the system performances. We are also working on the risk assessment to have a complete pedestrian protection system which will alert in case of detection of pedestrian in danger.

REFERENCES

Figure 9. The output of the multisensor system with asynchronous data fusion.

Figure 10. Hard point remains the classification of static targets. The left box (blue) is an example of false positive generated by the system.