Pedestrian crossing detection based on evidential fusion of video-sensors

L. Boudet, S. Midenet

▶ **To cite this version:**

L. Boudet, S. Midenet. Pedestrian crossing detection based on evidential fusion of video-sensors. Transportation research. Part C, Emerging technologies, 2009, Vol17,n5, pp.484-497. ⟨hal-00506354⟩

HAL Id: hal-00506354

https://hal.science/hal-00506354v1

Submitted on 27 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

Transportation Research Part C: Emerging Technologies, Vol 17, Issue 5, October 2009, Pp 484-497
Artificial Intelligence in Transportation Analysis: Approaches, Methods, and Applications
doi:10.1016/j.trc.2009.04.009
Copyright © 2009 Elsevier Ltd All rights reserved.
http://www.sciencedirect.com/science/journal/0968090X

Pedestrian crossing detection based on evidential fusion of video-sensors

Laurence Boudet*,[1], Sophie Midenet

Université Paris-Est,The French National Institute for Transport and Safety Research, GRETIA, 2 rue de la Butte Verte, 93166 Noisy-le-Grand cedex, France

Abstract

This paper introduces an online pedestrian crossing detection system that uses pre-existing traffic-oriented video-sensors which, at regular intervals, provide coarse spatial measurements on areas along a crosswalk. Pedestrian crossing detection is based on the recognition of occupancy patterns induced by pedestrians when they move on the crosswalk. In order to improve the ability of non-dedicated sensors to detect pedestrians, we introduce an evidential-based data fusion process that exploits redundant information coming from one or two sensors: intra-sensor fusion uses spatiotemporal characteristics of the measurements, and inter-sensor fusion uses redundancy between the two sensors. As part of the EU funded TRACKSS project on cooperative advanced sensors for road traffic applications, real data have been collected on an urban intersection equipped with two cameras. The results obtained show that the data fusion process enhances the quality of occupancy

*Corresponding author.
Email addresses: laurence.boudet@cea.fr (Laurence Boudet), sophie.midenet@inrets.fr (Sophie Midenet)
[1]Present address: CEA, LIST, Laboratoire Intelligence Multi-capteurs et Apprentissage, F-91191 Gif-sur-Yvette, France

patterns obtained and leads to high detection rates of pedestrian crossings with multi-purpose sensors in operational conditions, especially when a secondary sensor is available.
Key words: Urban traffic management, Multi-purpose video-sensor, Pedestrian crossing detection, Multi-sensor fusion, Theory of evidence

---

1. Introduction

Considering the issue raised by the impact of road traffic on climate, urban traffic management systems have to evolve toward a better consideration of non-pollutant modes of transport. Solutions are being investigated to favor pedestrian mobility by improving safety and comfort (Hughes et al., 2000). These improvements often require infrastructure modifications, but can also be achieved through traffic management actions, such as pedestrian-oriented traffic light strategies like Puffin or Pelican (Catchpole, 2003). This paper addresses the detection of pedestrians on multi-camera equipped signalized intersections, and describes an online system that detects pedestrian crossing events on crosswalks. This system is to be part of an observatory system dedicated to pedestrian mobility in signalized intersections, with focus on the assessment of time sharing between pedestrians and road traffic. Our aim is to analyze the impact of traffic light strategies and to evaluate how green and red phases for pedestrians relate to demand and to pedestrian crossing practices (McLeod et al., 2004).

Video-sensors are becoming more and more widespread for urban traffic management systems, and provide usual and innovative traffic measurements such as flow, queue length or spatial occupancy. A big advantage of video

sensors in urban contexts is that the same cameras used for motorized traffic analysis can provide information on specific traffic like trucks or buses, and also on pedestrian flow. Video-sensors can be considered as potential multi-purpose sensors for urban traffic control systems.

INRETS-GRETIA is participating in the TRACKSS project which addresses the potential of video-sensors in such matters. As part of the Information Society policies of the European Commission, the TRACKSS project - Technologies for Road Advanced Cooperative Knowledge Sharing Sensors - aims to develop new systems for cooperative sensing and predictive flow, infrastructure and environmental conditions surrounding traffic, with a view to improving the safety and efficiency of road transport operations (Trackss, 2008). As part of this project INRETS-GRETIA is working with the TRACKSS partner Citilog on the potential of using existing non-dedicated video sensors for online detection of pedestrian crossing events; at the same time Citilog and the ITACA Institute are working on bus detection and tracking through cooperation between magnetic loop and video-sensor.

This paper reports the results obtained on pedestrian crossing event detection. The system developed for that purpose receives inputs from traffic-oriented video sensors that compute spatial occupancy rates on predefined regions over a pedestrian pathway. The system is made up of two modules that transform these occupancy rates into pedestrian crossing occurrence detection:

- a data fusion module, which improves the basic measurement using spatiotemporal information redundancy, and multi-sensor redundancy when two cameras are available for analysis;

- a pattern recognition module, which detects temporal patterns induced by pedestrians crossing the road.

Our idea is twofold: exploiting when possible existing sensors to develop pedestrian crossing detection ability, and using a data fusion model to address the potential weaknesses of pedestrian detection due to non-optimal camera positions.

The data fusion process concerns both inter-sensor and intra-sensor fusion: inter-sensor fusion takes advantage of two sensors observing one crosswalk from different angles, while intra-sensor fusion takes advantage of the spatiotemporal characteristic of spatial occupancy. Both fusion processes are defined within the transferable belief model framework.

The pattern recognition module detects pedestrian crossings in the spatiotemporal data obtained after data fusion. It aims to detect as many pedestrian crossing patterns as possible and before their ending. Depending on the type of data used, these principles apply when a small number of pedestrians move in the scene, but not in crowded scenes such as station accesses.

2. System architecture

Over the last few years INRETS-GRETIA has equipped a real intersection in the close suburbs of Paris with a multi-camera system for road traffic management research projects (Midenet et al., 2004; Boillot et al., 2006). Figure 1(a) depicts one view of this experimental site, which shows one double-lane outbound link. The crosswalk that goes over this link has been chosen for the experiments reported here.

TRACKSS partner Citilog has provided us with traffic-oriented video sensors based on their product MediaCity; it has been adapted to pedestrians by internal parameters tuned to take into account the size and speed of pedestrian movements in the image. The underlying image processing software is based on movement detection (Auber et al., 1996) and provides spatial occupancy rates on predefined regions every second. For the pedestrian crossing event detection application, we define at least two regions of interest (ROI) covering the pedestrian pathway on the pavement, one region per lane in the case of a larger link. Two additional ROI are considered on each sidewalk (see Figure 1).

place Fig. 1 about here

We define a pedestrian crossing event (PCE) as an event lasting several seconds characterized by the presence of at least one pedestrian on the pavement. Occupancy state patterns (where the state is empty or occupied) on ROI induced by PCE reflect pedestrian movements from one side of the road to the other. These patterns differ from those caused by other events which induce occupancy rate variations such as vehicle flow: vehicles clear the crosswalk perpendicularly whereas pedestrians follow the crosswalk direction.

PCE and other occupancy-inducing events are differentiated using a pattern recognition module on the basis of the occupancy state spatiotemporal patterns. Those that are consistent with the evolution of pedestrian movement on the crosswalk are detected as PCE based on an analysis of local occupancy dynamics. Occupancy state patterns are more appropriate than occupancy rate patterns as they do not depend on the number or apparent size of pedestrians.

PCE detection performance depends on the performance of the video sensors that compute occupancy rates (OR) on the ROI. Video sensor performance in turn is very much dependent on the position of the camera, but some general trends can be observed.

- Being based on movement detection, the spatial occupancy rate over the region of the image constitutes a coarse but robust basic measurement that can be exploited under a large variety of weather and lighting conditions.
- Pedestrian movement is better translated into occupancy rate when the crosswalk is positioned horizontally in the image - like in Figure 1(a) -, since pedestrian appearance remains comparable from one side to the other.
- A single pedestrian may not produce sufficient apparent movement and may not be detected over some ROI. This is because movement detection is intentionally thresholded to avoid noise.
- Other events than pedestrian movement also induce positive occupancy rates, perpendicular vehicle flow for instance, since movement is detected without pattern recognition.

The principle we apply consists in using data fusion techniques in order to enhance the quality of this non-dedicated and robust sensor: the redundancy of information can thus offset the non-optimality of multi-purpose video sensors. Firstly, we exploit the gradual occupancy transmission between adjacent ROI when pedestrians cross over on a crosswalk by defining an intra-sensor fusion process to rectify the gaps in the spatiotemporal OR

pattern. Secondly, we define an inter-sensor fusion process that takes advantage of the redundancy between video-sensors when crosswalks happen to be covered by two cameras. This is the case for most crosswalks of our experimental site, including the one we are focusing on. The view of the secondary video sensor that covers it is depicted in Figure 1(b). Movement detection on pedestrians with the secondary sensor is not as good as with the primary sensor because of the effects of perspective along the crosswalk. However, it provides information that can be useful for solving under-detection problems, or in the case of occluding lateral flow event. This is the purpose of the inter-sensor fusion process.

The system architecture is depicted in Figure 2. The first module is provided with occupancy rates given by the primary video sensor, and with those given by the secondary video sensor, if any. The second module receives the occupancy states given by the first module, and provides the final output.

place Fig. 2 about here

3. Data Fusion

The data fusion module is responsible for transforming an array of occupancy rates (OR) coming from one - or two - video-sensors, into an array of occupancy states (OS) that reflects as correctly as possible the true occupation of corresponding regions over the crosswalk. In order to exploit and combine the various sources of information about OR arrays, we use the transferable belief model framework that is briefly presented below.

3.1. The TBM framework

The belief functions stated in the Dempster-Shafer theory of evidence (Dempster, 1968; Shafer, 1976) provide a powerful tool for representing confidence levels and uncertainty. Smets (Smets and Kennes, 1994) has recently proposed justifications and innovative interpretation of the theory of evidence within the so-called transferable belief model framework (TBM). One of the most interesting aspects of this theory relies on its ability to represent ignorance and conflicting sources. Within the TBM the set Ω of all possible states of a system is called the frame of discernment. Basic belief assignments (bba) are defined on the powerset 2^{Ω} and make it possible to work with non-mutually exclusive evidence represented by subsets of 2^{Ω}:

$$\begin{aligned} m: \quad 2^{\Omega} &\rightarrow \quad [0,1] \\ A &\rightarrow \quad m(A) \end{aligned} \tag{1}$$

where $\sum_{A \in 2^{\Omega}} m(A) = 1$. Subsets A where $m(A) \neq 0$ are called focal elements, and $m(A)$ values are called basic belief masses (bbm). Mass $m(A)$ can be interpreted as the degree of belief given to A and to none of its subsets, given available evidence. Partial ignorance is represented by assigning a non-zero value to Ω, whereas total ignorance is represented by the bba with Ω as the only focal element. Basic belief masses are used to define belief function $Bel(A)$ which describes the level of belief given to A under a given belief structure:

$$Bel(A) = \sum_{B|B \subseteq A} m(B), \forall A \in 2^{\Omega} \tag{2}$$

The TBM framework provides several rules for combining sources of evi-

dence (Smets, 2007). The choice of a combination rule is a key point in data fusion modelling: the rules differ in the way they deal with conflict. The original combination rule, known as Dempster's rule, is a conjunctive one: it emphasizes the agreement between sources and ignores all the conflicting evidence through a normalization factor. The combination is calculated from the two bbas m_1 and m_2 in the following way:

$$\begin{aligned} m_{1,2}(C) &= \frac{1}{1-K} \sum_{A \cap B = C} m_1(A) m_2(B), \forall C \in 2^\Omega \backslash \emptyset \\ m_{1,2}(\emptyset) &= 0 \end{aligned} \tag{3}$$

where $K = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$ measures the amount of conflict between the two sources of information. The normalization factor $1 - K$ reallocates the amount of conflict to all the other focal elements. Some authors have proposed other conjunctive rules: Yager's rule (Yager, 1987) attributes the conflict to Ω, that is to total ignorance. Dubois and Prade's rule (Dubois and Prade, 1988) assigns each source of conflict to the immediate super-set, that is to the origin of the conflict. Dubois and Prade's rule can be formulated for all C in $2^\Omega \setminus \emptyset$ in the following way:

$$\begin{aligned} m_{1,2}(C) &= \sum_{A \cap B = C} m_1(A) m_2(B) + \sum_{A \cap B = \emptyset, A \cup B = C} m_1(A) m_2(B) \\ m_{1,2}(\emptyset) &= 0 \end{aligned} \tag{4}$$

Note that disjunctive or compromise rules exist which may be better suited for a high level of conflict between sources (Smets, 1990, 1993).

Another big advantage of the TBM framework is that the reliability of a source can be taken into account with a reliability factor α. A source characterized by its bba structure is affected by the discount factor $(1-\alpha)$ in the following way:

$$\begin{aligned} m'(A) &= \alpha m(A), \forall A \in 2^{\Omega} \backslash \Omega \\ m'(\Omega) &= (1-\alpha) + \alpha m(\Omega) \end{aligned} \tag{5}$$

Several transformations of belief structure into decision variables are available. One strategy consists in spreading bbm into singletons: the so-called pignistic probabilities $BetP$ (Smets, 1990) are computed and the hypothesis (i.e. the singleton C_i) that maximizes it is selected.

$$BetP(C_i) = \sum_{A|C_i \in A} \frac{m(A)}{|A|(1-m(\emptyset))} \tag{6}$$

3.2. Occupancy rate processing within the TBM framework

The overall schema concerning occupancy data processing within the TBM framework is the following. Each sensor measurement (OR) is considered as a piece of evidence characterizing the occupancy state (OS) of an ROI. This piece of evidence is framed in the TBM: occupancy rates are converted into basic belief masses. After being combined with other bbms corresponding to other sources of information, the basic belief mass is converted into occupancy state through a pignistic probability decision (Eq. 6).

The frame of discernment is composed of the two possible hypotheses on

the occupancy state of ROI: $\Omega = \{\mathrm{E}, \mathrm{O}\}$ with E stands for empty and O for occupied. Let us note n the number of ROI and $r^i_{t,k}$ the sensor measurement given by sensor i on the ROI k at time t where $1 \leq i \leq 2$ and $1 \leq k \leq n$. We define a basic belief assignment on 2^Ω in the following way:

$$\tilde{m}^i_{t,k} = \begin{bmatrix} 0 \\ \rho(r^i_{t,k})\alpha^i \\ (1-\rho(r^i_{t,k}))\alpha^i \\ 1-\alpha^i \end{bmatrix} \tag{7}$$

using the vector notation $m = \begin{bmatrix} m(\emptyset) & m(\mathrm{E}) & m(\mathrm{O}) & m(\Omega) \end{bmatrix}^T$.

The parameter α^i introduces a discount process and the function ρ converts the measurement into the degree of belief that this ROI is empty; ρ is chosen as an exponential function:

$$\begin{aligned} \rho: \quad [0,100] &\rightarrow \quad [0,1] \\ r^i_{t,k} &\rightarrow \quad \rho(r^i_{t,k}) = exp\left(\frac{-\left(r^i_{t,k}\right)^2}{\sigma^2}\right) \end{aligned} \tag{8}$$

The parameter σ tunes the sensitivity of sensors to movement detection. Since the occupancy rate may be rather low when a single pedestrian crosses the street, σ is set at a very low value ($\sigma = 4$). An example of sensor measurement is shown in Figure 3.

place Fig. 3 about here

3.3. Intra-sensor data fusion

The intra-sensor data fusion model is based on the assumptions that, when a pedestrian crosses the street, (i) the occupancy states last several seconds for each ROI and (ii) there is a spatial propagation of the occupancy between adjacent ROI. Thus, the proposed model uses (i) temporal information in order to extend the current OS and (ii) spatial information in order to model spatial propagation of occupancy. Temporal information is widely used in temporal filtering methods such as Kalman filters. Even if filters have already been studied in the context of the TBM (Ramasso et al., 2007; Smets and Ristic, 2007), this approach has not been kept here because measurement frequency (each second) is rather low compared to the duration of the events: information integration about state transition would be delayed for several seconds, which does not comply with online constraints.

The idea is to identify situation changes and to anticipate temporal conflict. The model is meant (i) to integrate spatiotemporal information that increases the degree of belief in the current state and (ii) to adapt the reaction time to a situation change thanks to an evolution model. As we are interested in favoring occupancy detection, the evolution models are chosen in order (i) to increase rapidly the degree of belief in the state O when there is evidence of the state transition from E to O, and (ii) to decrease slowly the degree of belief in the state O when there is evidence of the reverse transition.

The intra-sensor fusion model is composed of three main steps (Figure 4):

1. **Evolution model selection:** by comparing the previous bbms on ROI k and its neighbors with the new observation $r^i_{t,k}$, the system determines

the new context characterizing ROI k. The region may be "becoming occupied" (O.a), "being occupied" (O.b), "holding occupied" (E.a) or "being empty" (E.b). According to this context, an evolution model is selected that provides the evolution bba $me^i_{t,k}$.

2. **Update fusion:** the past bbms $m^i_{t-1,k}$ are updated by fusing them with the evolution bba. It gives updated bbms $mu^i_{t-1,k}$.
3. **Temporal fusion:** The new bbms $m^i_{t,k}$ are provided by the temporal fusion between the instantaneous bbms $\tilde{m}^i_{t,k}$ and the updated bbms $mu^i_{t-1,k}$.

place Fig. 4 about here

Both fusion steps use the combination rule of Dubois and Prade (Eq. 4) that transfers the conflict to the set of conflicting hypotheses. As the set of discernment is made up of two exclusive hypotheses, this rule is equivalent to Yager's conjunctive rule that transfers the conflict into the ignorance Ω. When the conflict is high, the rule assumes that the current belief on a state has to be reconsidered in the light of a new piece of evidence. As it is applied twice in our application, it enables state change.

Context O.a and O.b: If the sensor measurement $r^i_{t,k}$ is higher than σ, the context is either "becoming occupied" if an adjacent ROI was occupied at $(t-1)$, or "being occupied" if not. We aim at favoring a quick increase of the degree of belief of state O. A state change from E to O is performed when the bba at time $(t-1)$ better supports hypothesis E than hypothesis O for the region k. In that case, the model trusts the new measurement and forgets the past knowledge. Indeed, using Dubois and Prade's rule with the two fusion steps enable this state change. The fusion of the previous bbms

with the evolution bba creates a high level conflict during the update fusion step, which is transferred to Ω. Since the measurement is high enough, the level of conflict can be reallocated to O at the temporal fusion step.

Context E.a and E.b: If the sensor measurement $r^i_{t,k}$ is lower than σ, the context is either "holding occupied" if the previous degree of belief on state O is sufficiently high, or "staying empty" if not. We favor a slow decrease of the degree of belief of state O.

Details of the evolution bbm used in each of these four contexts are given in Boudet and Midenet (2008).

3.4. Inter-sensor data fusion

Figure 5 depicts the overall fusion process in the case of a single video-sensor and in case of two video-sensors. When a secondary video-sensor gives additional sensor measurements, a multi-sensor fusion step is added that provides a new bbm $m^{1,2}_{t,k}$ on the basis of the bbms outputted by the two intra-sensor fusion processes. The multi-sensor fusion step is performed with the Dubois and Prade's combination rule (Eq. 4). The pignistic decision that provides the OS $s_{t,k}$ is computed on the basis of the fused bbms $m^{1,2}_{t,k}$. Furthermore, the fused bbms $m^{1,2}_{t,k}$ are also used in the intra-sensor fusion steps of each sensor.

place Fig. 5 about here

3.5. Input-dependant discounting of sources

During the data fusion processes, discounting factors are introduced twice: in the bba computation step (see 3.2) and in the multi-sensor fusion step (see 3.4). Traditionally, the discount process (see Eq. 5) enables to take into

account sensor reliability and to minor the influence of a sensor considered as less reliable. In our case, we observed that pedestrian movement under-detection happens on both sensors from time to time. Thus, we defined input-dependant discount processes to weaken the influence of a sensor only when it may have failed to detect pedestrian movement.

The input-dependant discount process $\alpha(r^i_{t,k})$ of the bba (see Eq. 7) is set to a value α^i when the measurement $r^i_{t,k}$ is higher than σ and is reduced to $\alpha^i - \gamma$ otherwise. Regarding the multi-sensor fusion step, the discount process is applied when only one sensor detects movement on a ROI. If the sensor a is the one that does not detect movement, the bba $m^a_{t,k}$ taken as input of the multi-sensor fusion is discounted by a factor $(1 - \alpha^a) = 0.3 + 0.2 m^a_{t,k}(\mathrm{E})$. Thus, the discounting factor is higher when the hypothesis E is more strongly supported.

4. Pattern recognition

The goal of the pattern recognition step is to distinguish pedestrians from other items moving on the crosswalk. It is based on spatiotemporal occupancy state pattern recognition and classifies the event in progress either as a pedestrian crossing event (PCE) or not as a PCE (noted as $\overline{\mathrm{PCE}}$).

Evolution of pedestrian movement on the crosswalk is quite typical: a pedestrian takes the crosswalk from one sidewalk to the other. Occupancy patterns induced by pedestrian crossings depend on crossing features such as direction and walking speed. They are highly variable since several pedestrians may cross the street at the same time or successively, in the same or opposite directions.

Correlation-based pattern matching could have been used to recognize pedestrian crossings with occupancy patterns. However, this technique requires listing a set of pattern examples that has to contain all possible patterns. Thus, the learning set has to be big enough, especially as occupancy may not be detected for a few seconds. Instead, we choose to recognize the local dynamics induced by pedestrian crossings: occupancy is temporally shifted between adjacent ROI and usually lasts a few seconds on each of them. Occupancy patterns induced by vehicle flows are different: occupancy begins quasi-simultaneously on the pavement regions and lasts for a longer or shorter period of time; occupancy on the sidewalk regions may occur due to the perspective effects (occlusions) of video imaging or vehicle shadows.

In order to characterize these properties, we convert the occupancy state sequences into occupancy duration states. Thus, the proposed pattern recognition method is based on a double-level process: at a local level, the class of the occupancy source is inferred by considering the occupancy duration states of two adjacent ROI; at a global level, a decision on the event in progress is taken based on the local level.

4.1. Occupancy state coding: fuzzy occupancy duration states

Fuzzy functions are used to convert occupancy state sequences into fuzzy occupancy duration (FOD) states. We use two fuzzy functions per OS: the OS is either "recent" or "long". The transition value between them is fixed at 3 seconds: it corresponds to the mean time that a pedestrian takes to cross a crosswalk region; fuzziness makes it possible to obtain pedestrian speed variability. The set of fuzzy functions $\mathcal{F} = \{f_{\mathrm{RE}}, f_{\mathrm{LE}}, f_{\mathrm{RO}}, f_{\mathrm{LO}}\}$ shown in Figure 6 converts an OS sequence into a FOD array $\delta \in [0,1]^4$ that cor-

responds to the fuzzy values of each state in $\mathcal{D} = \{d_{\mathrm{RE}}, d_{\mathrm{LE}}, d_{\mathrm{RO}}, d_{\mathrm{LO}}\}$. An FOD state is considered as active if its value is strictly positive. Depending on the fuzzy functions used, only one state is active each second except at transitions where there are two.

place Fig. 6 about here

4.2. Local pattern recognition

The recognition of occupancy source is based on analysis of the local occupancy dynamics. We consider the FOD arrays of two adjacent ROI: the active FOD states are usually different in the case of pedestrian crossings whereas they are usually the same in the case of vehicle flow.

Bayesian inference is a simple and effective way to address this recognition problem. The conditional probability of observing a pair of active FOD states (d_k^i, d_{k+1}^j) in $\mathcal{D}^2$ on ROI $(k, k+1)$ given the class c of a local occupancy source is computed from a learning set following the frequentist approach; posterior probabilities are computed by applying Bayes' theorem which reverses the conditional probabilities (9).

$$P(c|d_k^i, d_{k+1}^j) = \frac{P(d_k^i, d_{k+1}^j|c)P(c)}{\sum_c P(d_k^i, d_{k+1}^j|c)} \tag{9}$$

The FOD arrays are taken into account for the computation of frequency occurrence and posterior probabilities: the probability that the local occupancy source belongs to a class c given a pair of FOD arrays (δ_k, δ_{k+1}) becomes (10).

$$P(c|\delta_k, \delta_{k+1}) = \sum_{(i,j)\in[1,4]} \delta_k^i \delta_{k+1}^j P(c|d_k^i, d_{k+1}^j) \tag{10}$$

The set Ω_L of classes learnt represents the possible sources of the local occupancy. It contains three classes:

- c_{N}, a class for the local event "no occupancy",
- c_{PC}, a class for the local event "pedestrian crossing", and
- c_{VF}, a class for the local event "vehicle flow".

The class c_{VF} is considered because these events are very frequent with a characterizable occupancy dynamics. Bayes'inference principle is shown as a bayesian network in Figure 7(a); Figure 7(b) depicts a short illustrative sequence and shows the influence of the FOD states on the local occupancy sources inferred.

place Fig. 7 about here

For the generation of learning set, the OS have been labeled each second for each ROI according to the video records. However, errors in occupancy state estimation have to be learnt as well: if an OS is empty whereas an event occurs on the video, it is labeled as "no occupancy". When the two labels of a pair of ROI are different, we keep the instance only if one label of them is c_{N}; otherwise we discard it from the learning set. In addition, we select events that are well separated from others; concerning pedestrian, we discard bi-directional simultaneous crossings. Our objective is to provide the system with "pedagogical" examples.

4.3. Global pattern recognition

The aims of the global pattern recognition process is to determine the event in progress on the pavement part of the crosswalk. It accumulates

the local inferences computed for all the inference nodes over several seconds in order to determine whether the spatiotemporal occupancy pattern is consistent with a pedestrian crossing or not.

Inference nodes are treated differently if they are linked to a sidewalk region (outer nodes) or if they are linked to pavement regions (inner nodes). The former are used to accumulate evidence of the beginning or end of pedestrian crossing, while the latter are used to accumulate evidence of the occupancy source on the pavement. Formally, the time during which the most probable occupancy source remains the same is computed on each node, and then duration thresholds are used to confirm the occupancy source. These thresholds set the trade-off between the delay for taking a decision and its robustness. They are set at 3 consecutive seconds for vehicle flow, and at 5 consecutive seconds for pedestrian crossing (including the accumulation on outer nodes). An additional condition, the detection of a beginning or an end, is required for pedestrian crossing confirmation (see Figure 8).

place Fig. 8 about here

Once an occupancy source is confirmed, the beginning of the corresponding event is looked for backward. If the occupancy source is a pedestrian crossing, the event in progress is classified as PCE; it is classified as $\overline{\text{PCE}}$ otherwise. For each PCE decision, different time variables are saved (Figure 8): T_b, the time of the PCE beginning on the pavement; T_e the time of PCE end on the pavement and T_d the time of the decision.

5. Experiments

5.1. Experimental data

The two views shown in Figure 1 have been used in the experiments: they cover the same crosswalk from two different points of view but are primarily positioned for road traffic measurements. In this scene, the main classes of event are the pedestrian crossing events (PCEs) and the vehicle flow events (VFEs) that occur when the vehicles clear the crosswalk perpendicularly. A third class of event exists on this site: an occluding lateral flow event (OLFE) occurs when high vehicles (like buses and trucks) move in front of the crosswalk and occlude it from the primary sensor only.

The two internal ROI have been used to label the events and determine their limits: a PCE begins when a pedestrian steps onto the pavement and ends when the last pedestrian steps onto the second sidewalk. Two events belonging to the same class are distinct when there is a break longer than 2 seconds between them. Events of different classes are not exclusive, for instance a PCE is in conjunction with a VFE when a pedestrian is on the pavement while a vehicle is still on the other lane.

Two 40-minute sequences have been recorded at two different dates. The learning set has been generated based on a 15-minute sequence: it is composed of 15 PCEs (1'37) and 24 VFEs (6'16"). The test set is composed of 87 PCEs (9'37), 260 VFEs (21'03") and 17 OLFEs (1'18).

5.2. Illustration of the data processing

Figure 9 shows a 90-second sequence of sensor measurement on the four ROI that cover the crosswalk. In the first 30-second period, there are two

pedestrian crossings. The pedestrian patterns show a typical occupancy propagation from one sidewalk to the other one. Then, there are several vehicle flow events detected on the pavement (the two inner ROI). Figure 9 shows the bbms obtained ($m^{12}_{t,k}$) on the states E and O when the data fusion process[2] is applied on these data. At the beginning of a new occupancy in an ROI, the bbm on O increases gradually when only one sensor detects some movement and increases very quickly when both do. The bbm on O decreases to a low value after two seconds without movement detection. The bbm on Ω, defined as $(1 - m^{12}_{t,k}(\mathrm{O}) - \mathrm{m}^{12}_{\mathrm{t,k}}(\mathrm{E}))$, is high mainly at state transitions.

place Fig. 9 about here

Figure 10 shows the system results on the same 90-second sequence. Graphs 1 to 4 show the pignistic decisions on the occupancy states derived from the bbms. Graph 5 shows the classes of local occupancy sources obtained through the local pattern recognition process; the height of the vertical bars gives the posterior probability on the class obtained on the inner node linked to ROI 2 and 3. As shown in graphs 6 and 7, the system succeeds to detect the two pedestrian crossings and to discard as $\overline{\mathrm{PCE}}$ the following events.

place Fig. 10 about here

5.3. Evaluation protocol and metrics

Evaluation objectives are twofold: firstly, to evaluate the whole pedestrian crossing detection system, and secondly to evaluate the benefit of using a secondary sensor in the system as well as the benefit of using the data fusion

[2]The reliability coefficient is defined with $\alpha^i = 0.9$ and $\gamma = 0.2$ for both sensors.

process proposed.

The real events and the detected events need to be matched within the evaluation process. We consider that a real PCE is detected as soon as a detected PCE overlaps and that a detected PCE is a false alarm if no real PCE overlaps. Let us note $\mathcal{T}$ the set of real PCEs and $\mathcal{R}$ the set of detected PCEs, the evaluation criteria are:

- the PCE detection rate (DR) defined by $\text{DR} = \frac{\|\mathcal{R} \cap \mathcal{T}\|}{\|\mathcal{T}\|}$,
- the false alarm rate (FAR) defined by $\text{FAR} = 1 - \frac{\|\mathcal{R} \cap \mathcal{T}\|}{\|\mathcal{R}\|}$.

These evaluation criteria are compared for different test configurations: with or without intra-sensor fusion, and with or without inter-sensor fusion. The details of the six test configurations are given in Table 1 with the measurement used.

place Table 1 about here

In order to assess the quality of PCE detection, we use another criteria that measures how well the real PCE time intervals are matched by the detected PCE time intervals. We define the time percentage of PCE detections by $TP = \frac{dur(r_{e*} \cap t_e)}{dur(t_e)}$, where a real PCE t_e is detected by the PCE r_{e*} if any, and $dur(e)$ computes the duration of an event e. This criteria is computed on all the real events and is given as cumulative distribution.

Additional evaluation criteria are computed to estimate the performance of the online detection system. They are made up of:

- the error on the beginning time: time difference between the beginning time T_b of a detected PCE and the beginning time T_b^* of the real PCE (if any), defined by $T_b - T_b^*$;

- the error on the end time: time difference between the end time T_e of a detected PCE and the end time T_e^* of the real PCE (if any), defined by $T_e - T_e^*$;
- the delay for detecting a PCE according to the real PCE (if any) defined by $T_d - T_b^*$.

Figure 8 gives an example of an event with T_b, T_d and T_e.

5.4. Evaluation results

All the results given in this section relate to the test set.

PCE detection results are given in Table 2 according to the test configuration. These results are quite satisfactory when the primary sensor is used with and without data fusion ($\mathbf{H_1}, \mathbf{S_1}$): the detection rate of real PCEs is high (81%) even if the FARs are quite high. Note that the PCEs represent only one fourth of real events in the test set. The application of the intra-sensor fusion ($\mathbf{H_1}$) enables a 10% drop in the FAR: this process extends the occupancy a few seconds and fills the gaps between very close occupancy sequences. The FAR is reduced because fewer occupancy state patterns are consistent with pedestrian crossing patterns.

place Table 2 about here

As foreseen, the performance of the secondary sensor is quite poor. Nevertheless, the results show that a secondary sensor is a good complement to a primary optimal sensor and improves its performance in terms of detection rate and reduction in the number of false alarms. The best test configuration is the one that uses the double fusion process ($\mathbf{F_{12}}$): the FAR is the lowest obtained for all test configurations. A lot of false alarms are due (i) to side-

by-side vehicles that are slightly shifted when they clear the crosswalk and (ii) to pedestrians that come into a sidewalk region of the crosswalk whereas a VFE is on-going.

Figure 11 depicts the cumulative distribution of PCE detection time percentage (TP) for the configurations that use data fusion. It makes possible to compare the PCE detection quality of the different configurations. For instance, 40% (resp. $45\%, 16\%, 46\%$) of real PCEs are detected during at least 80% of their duration for $\mathbf{F_{12}}$ (resp. $\mathbf{H_1}, \mathbf{H_2}, \mathbf{G_{12}}$). The best configuration is the one whose graph is at top left, which is $\mathbf{G_{12}}$ or $\mathbf{H_1}$. This figure shows that the benefit in false alarms obtained by $\mathbf{F_{12}}$ (see Table 2) is not at the expense of the detection quality.

place Fig. 11 about here

Table 3 gives the detection rates obtained on real PCEs according to type: whether or not the pedestrian is alone, whether or not the crossing is isolated from other events. A crossing is isolated from other events if there is a gap longer than 2 seconds between the crossing and the previous and next events. The detection rate of pedestrian groups and isolated crossings are very high: OS patterns induced by these crossings are complete and disjoined from vehicle flow-inducing OS patterns. The intra-sensor fusion improves the detection of single pedestrians and isolated crossings. The poor performance of the systems using the secondary sensor ($\mathbf{S_2}, \mathbf{H_2}$) comes from their failure in detecting single pedestrians.

place Table 3 about here

The evaluation results of the online detection system based on the double fusion process ($\mathbf{F_{12}}$) are shown as distributions in Figure 12 and 13. Figure

12 shows the errors made on the beginning and end of the detected events, whereas Figure 13 shows the detection delays. Figure 12 shows that the PCE decisions are good when they relate to a real PCE: their beginning time and end time are accurate ($\pm$ 2 seconds) for around 70% of them. The error on the beginning time is centered at zero. The detected events last mostly one second longer than the real ones.

place Fig. 12 and 13 about here

Figure 13 shows that the events are detected mostly 4 seconds after the beginning of the real PCE on the pavement. This delay is acceptable as it corresponds to the mean time taken by a pedestrian for crossing one lane. Let us note that the few cases of negative delays are due to false alarms occurring right before the real PCEs.

6. Conclusion

We have introduced an online pedestrian crossing detection system supplied with traffic-oriented video-sensors that provide coarse measurements on areas along a crosswalk. One of its components is a pattern recognition module that detects pedestrian crossings as soon as possible in temporal occupancy state sequences. This module recognizes the occupancy patterns compliant with pedestrian evolution on a crosswalk based on the analysis of local occupancy dynamics. The other component is a data fusion module that fuses the measurements provided by two sensors and that transforms them into occupancy states. It has been devised to exploit spatiotemporal characteristics of the measurements in order to correct the under-detection of pedestrians by video-sensors and to remain usable with only one sensor.

The results obtained with real operational data show that the fusion process enhances the quality of occupancy state patterns used for pattern recognition and leads to significant improvements in pedestrian detection as well as in false alarm reduction. This shows that the same cameras fixed on the infrastructure can be used for multi-purpose traffic scene analysis once efficient post-processing is provided.

The next step in data fusion developments will deal with enhanced inter-sensor conflict management in order to solve pedestrian detection issues in the case of occluding lateral flow events. New traffic scenes collected as part of the TRACKSS project will enrich our data base for further development and assessment processing. The pedestrian crossing detection system is planned to be used on INRETS experimental site for traffic management studies aiming at analyzing and improving pedestrian mobility and safety.

Acknowledgment

The authors would like to thank the European Commission for funding this work within the TRACKSS project.

References

Aubert, D., Bouzar S., Lenoir, F., Blosseville J.M., 1996. Automic vehicle queue measurement at intersections using image-processing. Proceedings of the 8th International Conference on Road Traffic Monitoring and Control, 422, London, pp. 100-104.

Boillot, F., Midenet, S., Pierrelée, J.C., 2006. The real-time urban traffic

559 control system CRONOS: algorithm and experiments. Transportation Re-
560 search C 14, 18-39.

561 Boudet, L., Midenet, S., 2008. A spatiotemporal data fusion model for occu-
562 pancy state estimation: an evidential approach. Proceedings of the 11th
563 International Conference on Information Fusion, Cologne, 30 June-3 July
564 2008, pp. 1333-1339.

565 Catchpole, J., 2003. Win-win outcomes for pedestrians and drivers by opti-
566 mizing traffic signal timing. Road and Transport Research Journal 12(3),
567 74-82.

568 Dempster, A., 1968. A generalization of bayesian inference. Journal of the
569 Royal Statistical Society 30, 205-247.

570 Dubois, D., Prade, H., 1988. Representation and combination of uncertainty
571 with belief functions and possibility measures. Computational Intelligence
572 4, 244-264.

573 Hughes, R., Huang, H., Zegeer, C., Cynecki, M., 2000. Automated detection
574 of pedestrians in conjunction with standard pedestrian push button at
575 signalized intersections. Transportation Research Record 1705, 32-39.

576 McLeod, F.N., Hounsell, N.B., Rajbhandari, B., 2004. Improving traffic sig-
577 nal control for pedestrians. IEE International Conference on Road Trans-
578 port Information and Control 12, London, 20-22 April 2004, pp. 268-277.

579 Midenet, S., Boillot, F., Pierrelée, J.C., 2004. Signalized intersection with
580 real-time adaptive control: on-field assessment of CO2 and pollutant emis-
581 sion prediction. Transportation Research D 9, 29-47.

582 Ramasso, E., Rombaut, M., Pellerin, P., 2007. State filtering and change
583 detection using TBM conflict. Application to human action recognition in
584 athletics videos. IEEE Trans. Circuits Syst. Video Techn., 17(7), 944-949.

585 Shafer, G., 1976. A mathematical theory of evidence. Princeton University
586 Press.

587 Smets, P., 1990. Constructing the Pignistic Probability Function in a context
588 of uncertainty. Uncertainy in Artificial Intelligence 5, 29-39.

589 Smets, P., 1993. Belief functions: The disjunctive rule of combination and
590 the Generalized Bayesian theorem. International Journal of Approximate
591 Reasoning 9(1), 1-35.

592 Smets, P., 2007. Analyzing the combination of conflicting belief functions.
593 Information Fusion 8(4), 387-412.

594 Smets, P., Kennes, R., 1994. The Transferable Belief Model. Artificial Intel-
595 ligence 66, 191-243.

596 Smets, P., Ristic, B., 2007. Kalman filter and joint tracking and classification
597 based on belief functions in the TBM framework. Information Fusion, Spe-
598 cial Issue on the Seventh International Conference on Information Fusion,
599 Part II, 8(1), pp. 16-27.

600 TRACKSS project, 2008, www.trackss.net

601 Yager, R.R., 1987. On the Dempster-Shafer framework and new combination
602 rules. Information Sciences 41, 93-137.

List of Figures

(a) Primary sensor: horizontal view (b) Secondary sensor: lateral view

Figure 1: An example of the two views with the underlined regions of interest on the same crosswalk.

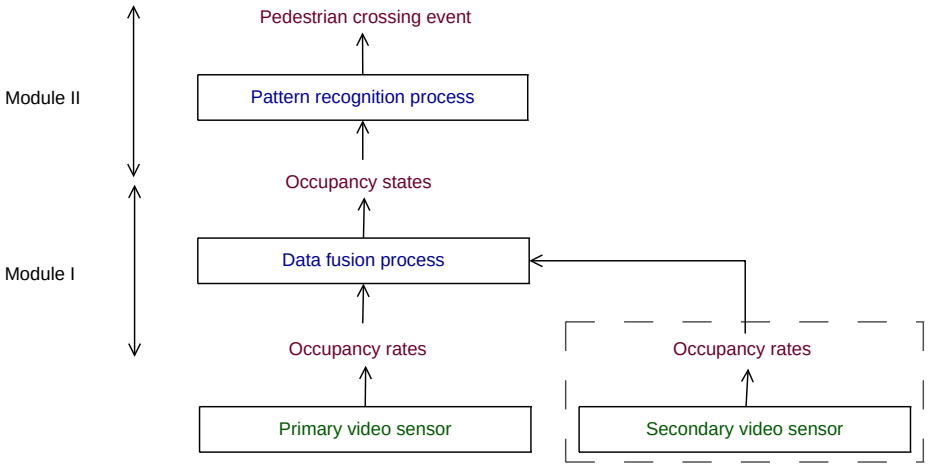


Figure 2: System architecture with a primary sensor and an optional secondary sensor.

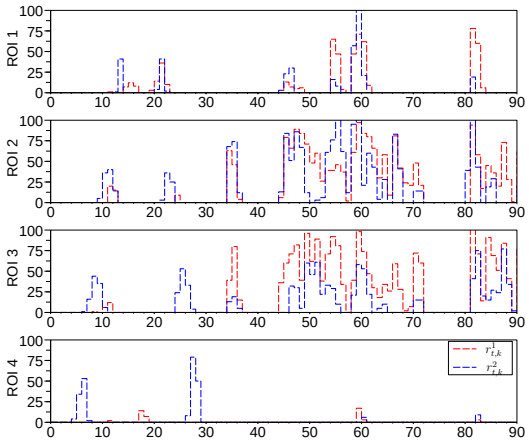


Figure 3: An example of sensor measurements on a 4-ROI crosswalk; the time scale is in seconds.

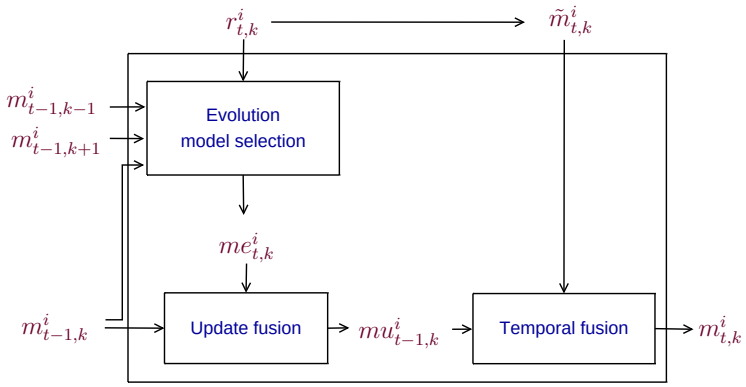
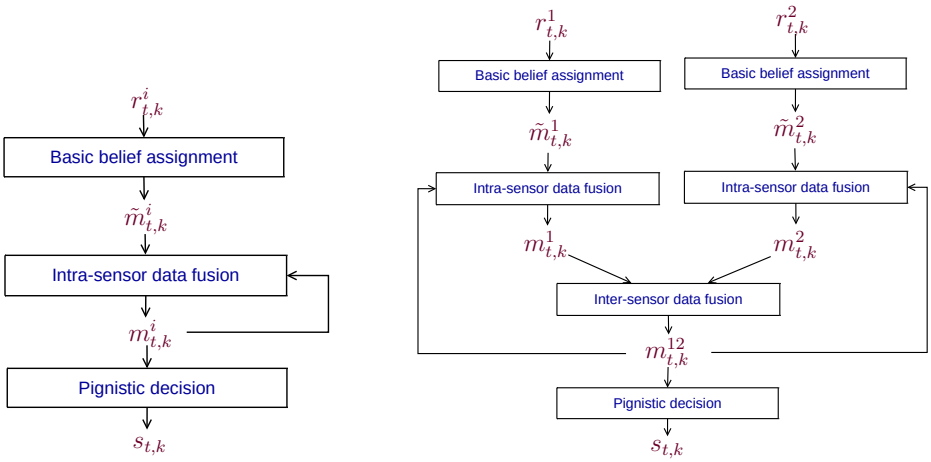


Figure 4: Intra-sensor data fusion process.



(a) The mono-sensor case (b) The multi-sensor case

Figure 5: Data fusion processes: the mono-sensor and the multi-sensor cases.

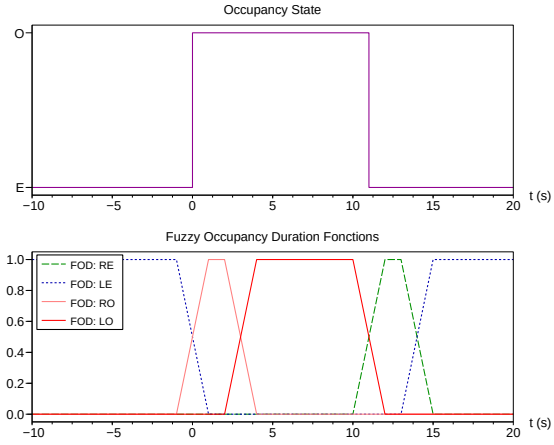
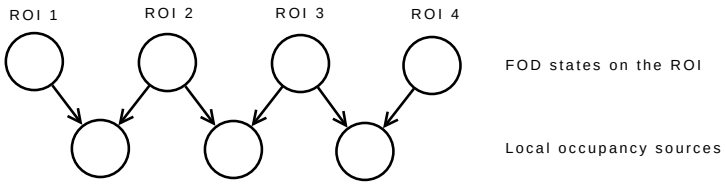
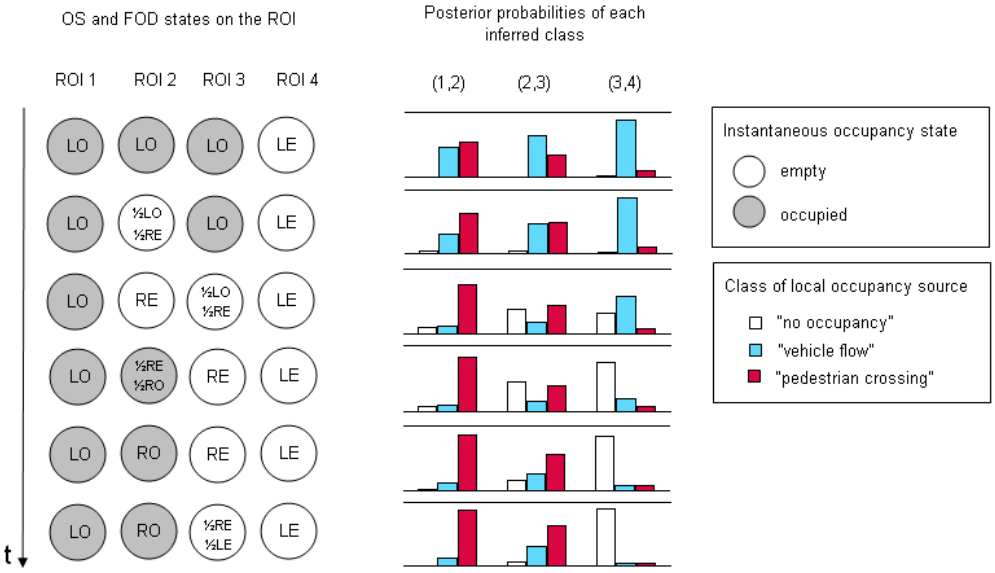


Figure 6: Conversion of an Occupancy State sequence into Fuzzy Occupancy Duration states where R: Recently, L: Lengthily, E: Empty and O: Occupied.



(a) Bayesian architecture



(b) Bayesian inference on a sequence with state transitions

Figure 7: Bayesian inference of local occupancy souces based on FOD states of pairs of adjacent ROI, where R: Recently, L: Lengthily, E: Empty and O: Occupied.

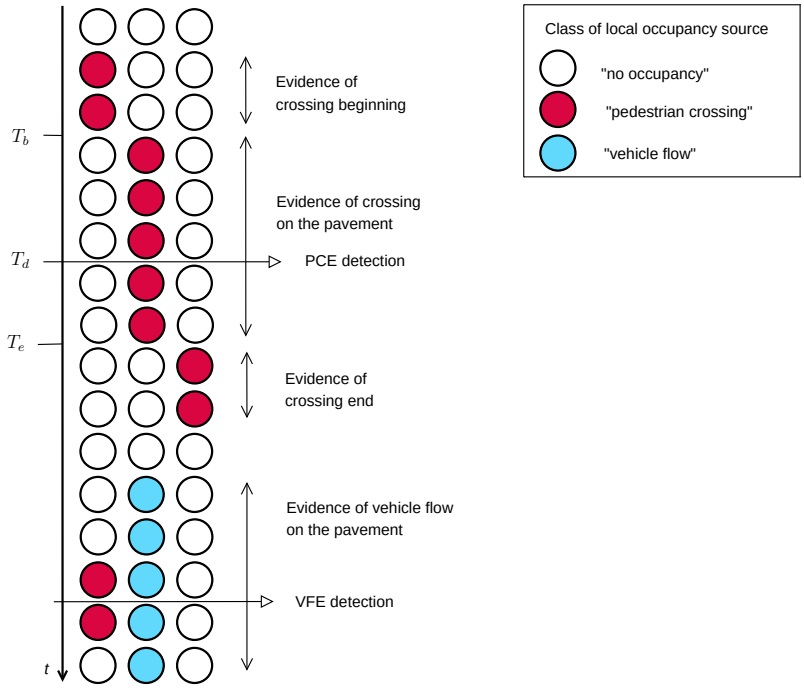


Figure 8: Illustration of the global pattern recognition process based on temporal inferences in the 3 local occupancy source nodes. The inner nodes (only one when $n = 4$) are linked to ROI on the pavement.

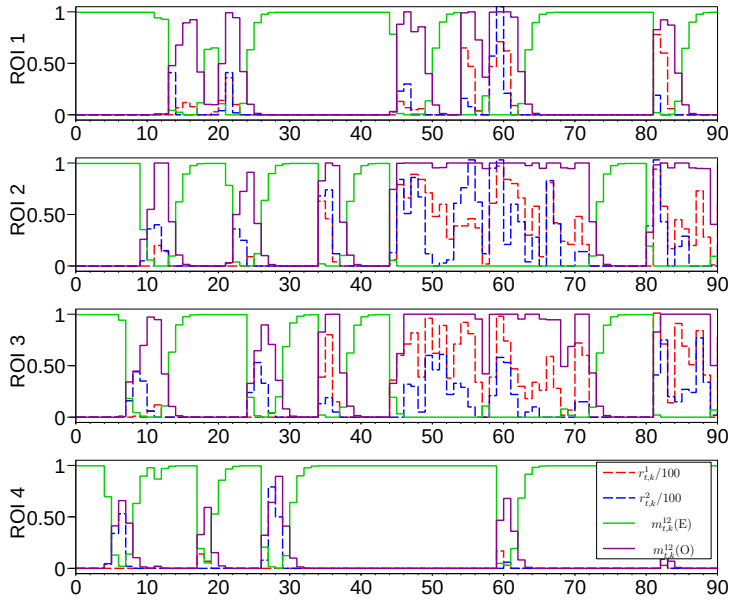


Figure 9: An example of sensor measurement and the results of the data fusion process; the time scale is in seconds.

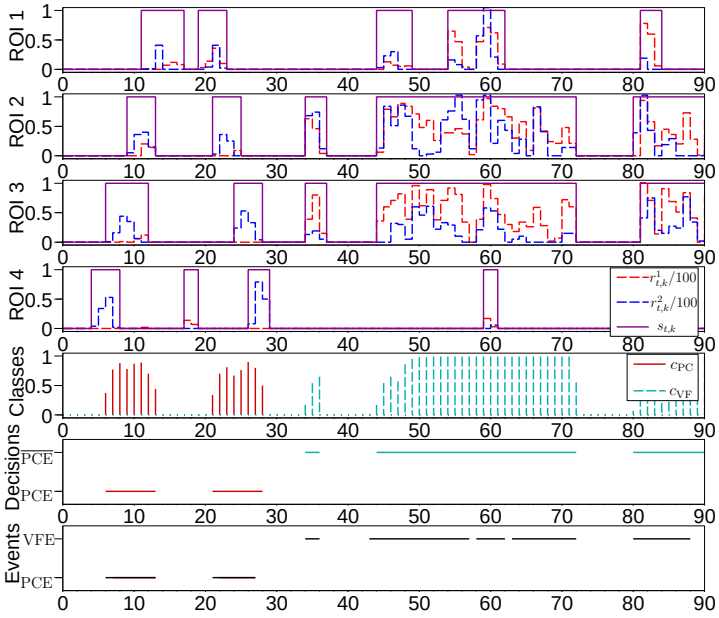


Figure 10: Pignistic decisions on occupancy state on the 4-ROI crosswalk - the state occupied is represented by $s_{t,k} = 1$ and the state empty by $s_{t,k} = 0$ - (graphs 1 to 4), local and global pattern recognition results (graphs 5 to 6), and real events (graph 7).

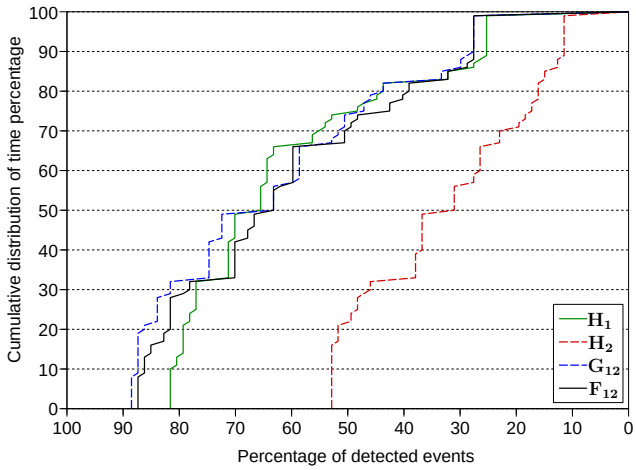


Figure 11: Cumulative distribution of time percentage of PCE detection.

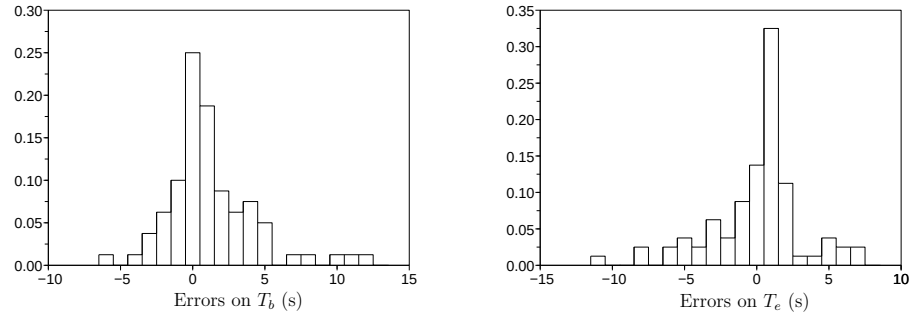


Figure 12: Distribution of errors between real PCEs and PCE decisions for the configuration $\mathbf{F_{12}}$: errors on the beginning time (left) and end time (right).

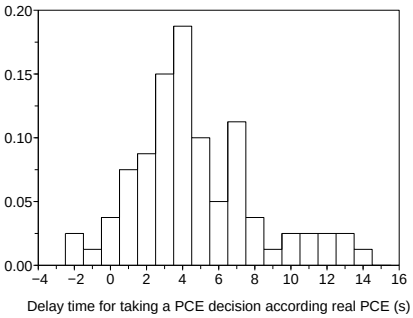


Figure 13: Distribution of delays of PCE decisions according to the real PCEs for the configuration $\mathbf{F_{12}}$.

List of Tables

Table 1: Data fusion processes used according to the configurations

Test configuration	**Measurement used**	**Intra-sensor fusion**	**Inter-sensor fusion**
$\mathbf{S_1}$	$r^1_{t,k}$	-	-
$\mathbf{S_2}$	$r^2_{t,k}$	-	-
$\mathbf{H_1}$	$r^1_{t,k}$	✓	-
$\mathbf{H_2}$	$r^2_{t,k}$	✓	-
$\mathbf{G_{12}}$	$r^1_{t,k}$; $r^2_{t,k}$	-	✓
$\mathbf{F_{12}}$	$r^1_{t,k}$; $r^2_{t,k}$	✓	✓

Table 2: PCE detection results according to test configurations (with the number of examples)

Test configuration	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{H_1}$	$\mathbf{H_2}$	$\mathbf{G_{12}}$	$\mathbf{F_{12}}$	#
PCE detection rate	81.6%	48.3%	81.6%	52.9%	88.5%	87.4%	87
PCE false alarm rate	38.4%	32.4%	28.4%	33%	34.7%	21.6%	[61, 125]

Table 3: PCE detection rate obtained according to type of crossing and test configuration (with the number of examples)

Test configuration	$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{H_1}$	$\mathbf{H_2}$	$\mathbf{G_{12}}$	$\mathbf{F_{12}}$	#
PCE detection rate on single pedestrian	76.7%	33.3%	78.3%	36.7%	86.7%	86.7%	60
PCE detection rate on pedestrian groups	92.6%	81.5%	88.9%	88.9%	92.6%	88.9%	27
PCE detection rate on isolated crossings	85.4%	50.0%	87.5%	54.2%	93.8%	93.8%	48
PCE detection rate on non-isolated crossings	76.5%	47.1%	73.5%	52.9%	82.4%	76.5%	34