



HAL
open science

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.

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)

¹Present 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

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

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

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

24 INRETS-GRETIA is participating in the TRACKSS project which ad-
25 dresses the potential of video-sensors in such matters. As part of the Infor-
26 mation Society policies of the European Commission, the TRACKSS project
27 - Technologies for Road Advanced Cooperative Knowledge Sharing Sen-
28 sors - aims to develop new systems for cooperative sensing and predic-
29 tive flow, infrastructure and environmental conditions surrounding traffic,
30 with a view to improving the safety and efficiency of road transport opera-
31 tions (Trackss, 2008). As part of this project INRETS-GRETIA is working
32 with the TRACKSS partner Citilog on the potential of using existing non-
33 dedicated video sensors for online detection of pedestrian crossing events; at
34 the same time Citilog and the ITACA Institute are working on bus detection
35 and tracking through cooperation between magnetic loop and video-sensor.

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

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

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

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

51 The data fusion process concerns both inter-sensor and intra-sensor fu-
52 sion: inter-sensor fusion takes advantage of two sensors observing one cross-
53 walk from different angles, while intra-sensor fusion takes advantage of the
54 spatiotemporal characteristic of spatial occupancy. Both fusion processes are
55 defined within the transferable belief model framework.

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

61 **2. System architecture**

62 Over the last few years INRETS-GRETIA has equipped a real inter-
63 section in the close suburbs of Paris with a multi-camera system for road
64 traffic management research projects (Midenet et al., 2004; Boillot et al.,
65 2006). Figure 1(a) depicts one view of this experimental site, which shows
66 one double-lane outbound link. The crosswalk that goes over this link has
67 been chosen for the experiments reported here.

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

78 place Fig. 1 about here

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

86 PCE and other occupancy-inducing events are differentiated using a pat-
87 tern recognition module on the basis of the occupancy state spatiotemporal
88 patterns. Those that are consistent with the evolution of pedestrian move-
89 ment on the crosswalk are detected as PCE based on an analysis of local
90 occupancy dynamics. Occupancy state patterns are more appropriate than
91 occupancy rate patterns as they do not depend on the number or apparent
92 size of pedestrians.

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

97 - Being based on movement detection, the spatial occupancy rate over
98 the region of the image constitutes a coarse but robust basic measure-
99 ment that can be exploited under a large variety of weather and lighting
100 conditions.

101 - Pedestrian movement is better translated into occupancy rate when the
102 crosswalk is positioned horizontally in the image - like in Figure 1(a) -,
103 since pedestrian appearance remains comparable from one side to the
104 other.

105 - A single pedestrian may not produce sufficient apparent movement and
106 may not be detected over some ROI. This is because movement detec-
107 tion is intentionally thresholded to avoid noise.

108 - Other events than pedestrian movement also induce positive occupancy
109 rates, perpendicular vehicle flow for instance, since movement is de-
110 tected without pattern recognition.

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

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

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

131 place Fig. 2 about here

132 **3. Data Fusion**

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

139 *3.1. The TBM framework*

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

$$m : 2^\Omega \rightarrow [0, 1] \quad (1)$$

$$A \rightarrow m(A)$$

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

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

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

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

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

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

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

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

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

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

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

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

183 3.2. Occupancy rate processing within the TBM framework

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

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

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

$$\tilde{m}_{t,k}^i = \begin{bmatrix} 0 \\ \rho(r_{t,k}^i)\alpha^i \\ (1 - \rho(r_{t,k}^i))\alpha^i \\ 1 - \alpha^i \end{bmatrix} \quad (7)$$

196 using the vector notation $m = \left[m(\emptyset) \quad m(E) \quad m(O) \quad m(\Omega) \right]^T$.

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

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

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

204 place Fig. 3 about here

205 3.3. Intra-sensor data fusion

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

218 The idea is to identify situation changes and to anticipate temporal con-
219 flict. The model is meant (i) to integrate spatiotemporal information that
220 increases the degree of belief in the current state and (ii) to adapt the re-
221 action time to a situation change thanks to an evolution model. As we are
222 interested in favoring occupancy detection, the evolution models are chosen
223 in order (i) to increase rapidly the degree of belief in the state O when there
224 is evidence of the state transition from E to O, and (ii) to decrease slowly the
225 degree of belief in the state O when there is evidence of the reverse transition.

226 The intra-sensor fusion model is composed of three main steps (Figure
227 4):

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

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

234 2. **Update fusion:** the past bbms $m_{t-1,k}^i$ are updated by fusing them
 235 with the evolution bba. It gives updated bbms $mu_{t-1,k}^i$.

236 3. **Temporal fusion:** The new bbms $m_{t,k}^i$ are provided by the temporal
 237 fusion between the instantaneous bbms $\tilde{m}_{t,k}^i$ and the updated bbms
 238 $mu_{t-1,k}^i$.

239 place Fig. 4 about here

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

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

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

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

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

264 3.4. Inter-sensor data fusion

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

274 place Fig. 5 about here

275 3.5. Input-dependant discounting of sources

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

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

284 The input-dependant discount process $\alpha(r_{t,k}^i)$ of the bba (see Eq. 7) is set
 285 to a value α^i when the measurement $r_{t,k}^i$ is higher than σ and is reduced to
 286 $\alpha^i - \gamma$ otherwise. Regarding the multi-sensor fusion step, the discount process
 287 is applied when only one sensor detects movement on a ROI. If the sensor
 288 a is the one that does not detect movement, the bba $m_{t,k}^a$ taken as input of
 289 the multi-sensor fusion is discounted by a factor $(1 - \alpha^a) = 0.3 + 0.2m_{t,k}^a(\mathbf{E})$.
 290 Thus, the discounting factor is higher when the hypothesis \mathbf{E} is more strongly
 291 supported.

292 4. Pattern recognition

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

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

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

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

320 *4.1. Occupancy state coding: fuzzy occupancy duration states*

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

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

332 place Fig. 6 about here

333 4.2. Local pattern recognition

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

338 Bayesian inference is a simple and effective way to address this recognition
 339 problem. The conditional probability of observing a pair of active FOD states
 340 (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
 341 is computed from a learning set following the frequentist approach; posterior
 342 probabilities are computed by applying Bayes' theorem which reverses the
 343 conditional probabilities (9).

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

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

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

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

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

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

358 place Fig. 7 about here

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

368 4.3. Global pattern recognition

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

371 the local inferences computed for all the inference nodes over several sec-
372 onds in order to determine whether the spatiotemporal occupancy pattern is
373 consistent with a pedestrian crossing or not.

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

386 place Fig. 8 about here

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

393 5. Experiments

394 5.1. Experimental data

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

403 The two internal ROI have been used to label the events and determine
404 their limits: a PCE begins when a pedestrian steps onto the pavement and
405 ends when the last pedestrian steps onto the second sidewalk. Two events
406 belonging to the same class are distinct when there is a break longer than
407 2 seconds between them. Events of different classes are not exclusive, for
408 instance a PCE is in conjunction with a VFE when a pedestrian is on the
409 pavement while a vehicle is still on the other lane.

410 Two 40-minute sequences have been recorded at two different dates. The
411 learning set has been generated based on a 15-minute sequence: it is com-
412 posed of 15 PCEs (1'37) and 24 VFEs (6'16"). The test set is composed of
413 87 PCEs (9'37), 260 VFEs (21'03") and 17 OLFEs (1'18).

414 5.2. Illustration of the data processing

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

417 pedestrian crossings. The pedestrian patterns show a typical occupancy
 418 propagation from one sidewalk to the other one. Then, there are several
 419 vehicle flow events detected on the pavement (the two inner ROI). Figure 9
 420 shows the bbms obtained ($m_{t,k}^{12}$) on the states E and O when the data fusion
 421 process² is applied on these data. At the beginning of a new occupancy in an
 422 ROI, the bbm on O increases gradually when only one sensor detects some
 423 movement and increases very quickly when both do. The bbm on O decreases
 424 to a low value after two seconds without movement detection. The bbm on
 425 Ω , defined as $(1 - m_{t,k}^{12}(\text{O}) - m_{t,k}^{12}(\text{E}))$, is high mainly at state transitions.

426 place Fig. 9 about here

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

435 place Fig. 10 about here

436 5.3. Evaluation protocol and metrics

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

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

440 process proposed.

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

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

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

452 place Table 1 about here

453 In order to assess the quality of PCE detection, we use another criteria
454 that measures how well the real PCE time intervals are matched by the de-
455 tected PCE time intervals. We define the time percentage of PCE detections
456 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,
457 and $dur(e)$ computes the duration of an event e . This criteria is computed
458 on all the real events and is given as cumulative distribution.

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

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

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

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

470 5.4. Evaluation results

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

472 PCE detection results are given in Table 2 according to the test configu-
 473 ration. These results are quite satisfactory when the primary sensor is used
 474 with and without data fusion ($\mathbf{H}_1, \mathbf{S}_1$): the detection rate of real PCEs is
 475 high (81%) even if the FARs are quite high. Note that the PCEs represent
 476 only one fourth of real events in the test set. The application of the intra-
 477 sensor fusion (\mathbf{H}_1) enables a 10% drop in the FAR: this process extends the
 478 occupancy a few seconds and fills the gaps between very close occupancy
 479 sequences. The FAR is reduced because fewer occupancy state patterns are
 480 consistent with pedestrian crossing patterns.

481 place Table 2 about here

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

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

491 Figure 11 depicts the cumulative distribution of PCE detection time per-
492 centage (TP) for the configurations that use data fusion. It makes possible
493 to compare the PCE detection quality of the different configurations. For in-
494 stance, 40% (resp. 45%, 16%, 46%) of real PCEs are detected during at least
495 80% of their duration for \mathbf{F}_{12} (resp. $\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_{12}$). The best configuration
496 is the one whose graph is at top left, which is \mathbf{G}_{12} or \mathbf{H}_1 . This figure shows
497 that the benefit in false alarms obtained by \mathbf{F}_{12} (see Table 2) is not at the
498 expense of the detection quality.

499 place Fig. 11 about here

500 Table 3 gives the detection rates obtained on real PCEs according to
501 type: whether or not the pedestrian is alone, whether or not the crossing is
502 isolated from other events. A crossing is isolated from other events if there is
503 a gap longer than 2 seconds between the crossing and the previous and next
504 events. The detection rate of pedestrian groups and isolated crossings are
505 very high: OS patterns induced by these crossings are complete and disjointed
506 from vehicle flow-inducing OS patterns. The intra-sensor fusion improves the
507 detection of single pedestrians and isolated crossings. The poor performance
508 of the systems using the secondary sensor ($\mathbf{S}_2, \mathbf{H}_2$) comes from their failure
509 in detecting single pedestrians.

510 place Table 3 about here

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

513 12 shows the errors made on the beginning and end of the detected events,
514 whereas Figure 13 shows the detection delays. Figure 12 shows that the PCE
515 decisions are good when they relate to a real PCE: their beginning time and
516 end time are accurate (± 2 seconds) for around 70% of them. The error on
517 the beginning time is centered at zero. The detected events last mostly one
518 second longer than the real ones.

519 place Fig. 12 and 13 about here

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

525 **6. Conclusion**

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

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

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

550 **Acknowledgment**

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

553 **References**

554 Aubert, D., Bouzar S., Lenoir, F., Blosseville J.M., 1996. Automatic vehicle
555 queue measurement at intersections using image-processing. Proceedings of
556 the 8th International Conference on Road Traffic Monitoring and Control,
557 422, London, pp. 100-104.

558 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. *Uncertainty 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.

603 **List of Figures**

604 1 An example of the two views with the underlined regions of
605 interest on the same crosswalk. 31

606 2 System architecture with a primary sensor and an optional
607 secondary sensor. 32

608 3 An example of sensor measurements on a 4-ROI crosswalk; the
609 time scale is in seconds. 33

610 4 Intra-sensor data fusion process. 34

611 5 Data fusion processes: the mono-sensor and the multi-sensor
612 cases. 35

613 6 Conversion of an Occupancy State sequence into Fuzzy Oc-
614 cupancy Duration states where R: Recently, L: Lengthily, E:
615 Empty and O: Occupied. 36

616 7 Bayesian inference of local occupancy sources based on FOD
617 states of pairs of adjacent ROI, where R: Recently, L: Lengthily,
618 E: Empty and O: Occupied. 37

619 8 Illustration of the global pattern recognition process based on
620 temporal inferences in the 3 local occupancy source nodes.
621 The inner nodes (only one when $n = 4$) are linked to ROI on
622 the pavement. 38

623 9 An example of sensor measurement and the results of the data
624 fusion process; the time scale is in seconds. 39

625	10	Pignistic decisions on occupancy state on the 4-ROI crosswalk	
626		- the state occupied is represented by $s_{t,k} = 1$ and the state	
627		empty by $s_{t,k} = 0$ - (graphs 1 to 4), local and global pattern	
628		recognition results (graphs 5 to 6), and real events (graph 7). .	40
629	11	Cumulative distribution of time percentage of PCE detection.	41
630	12	Distribution of errors between real PCEs and PCE decisions	
631		for the configuration \mathbf{F}_{12} : errors on the beginning time (left)	
632		and end time (right).	42
633	13	Distribution of delays of PCE decisions according to the real	
634		PCEs for the configuration \mathbf{F}_{12}	43



(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.

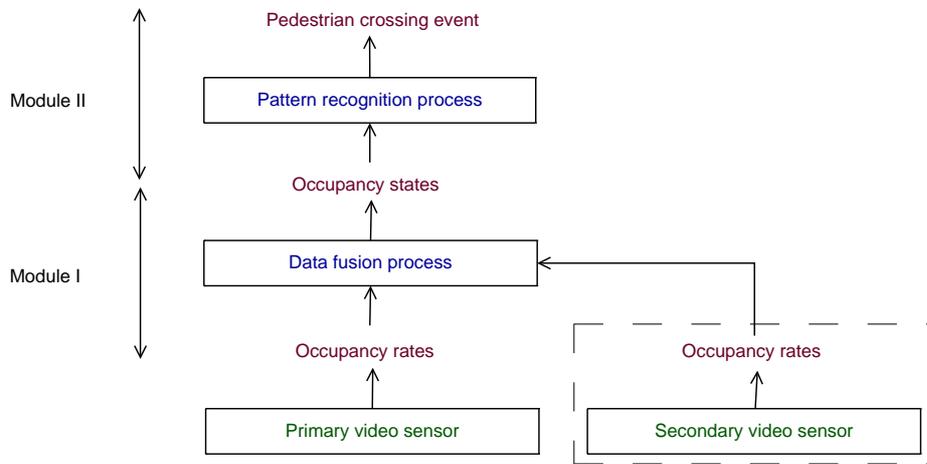


Figure 2: System architecture with a primary sensor and an optional secondary sensor.

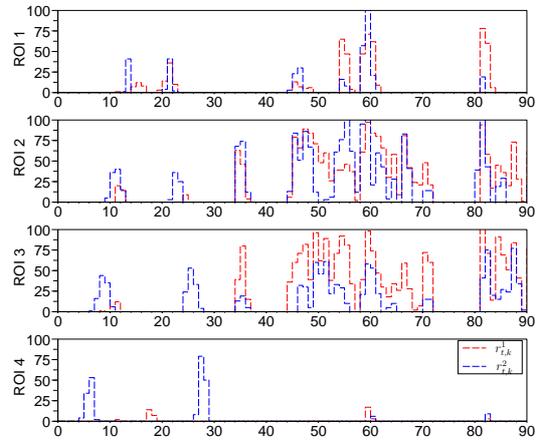


Figure 3: An example of sensor measurements on a 4-ROI crosswalk; the time scale is in seconds.

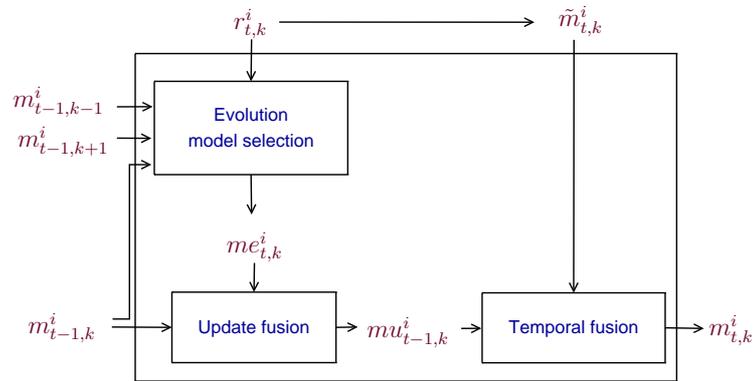


Figure 4: Intra-sensor data fusion process.

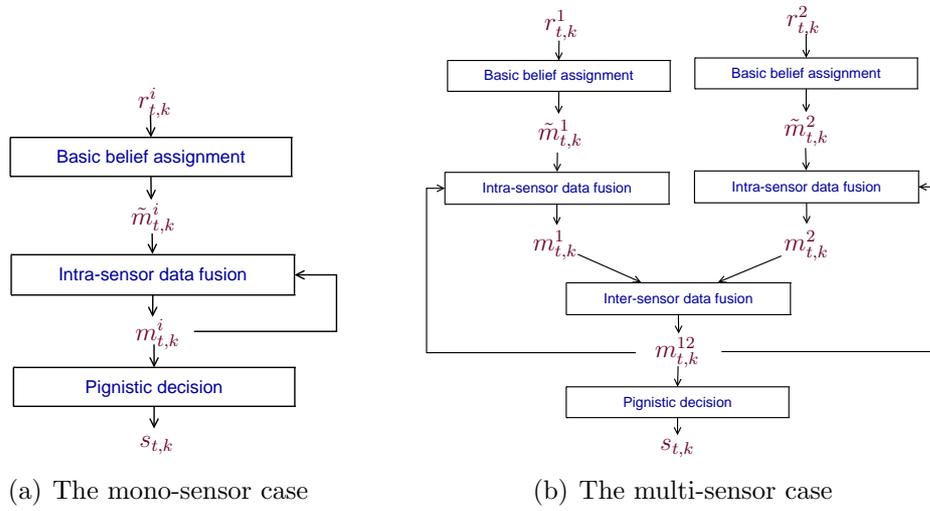


Figure 5: Data fusion processes: the mono-sensor and the multi-sensor cases.

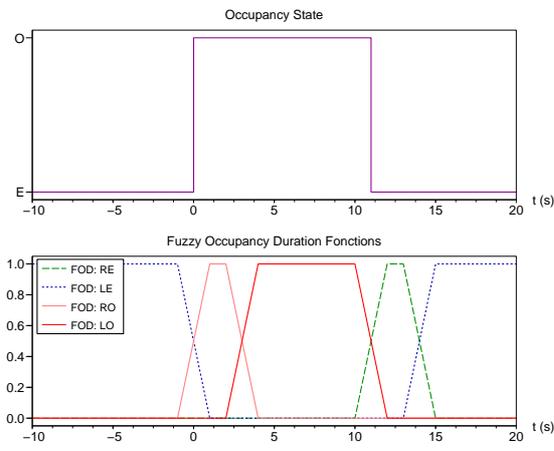
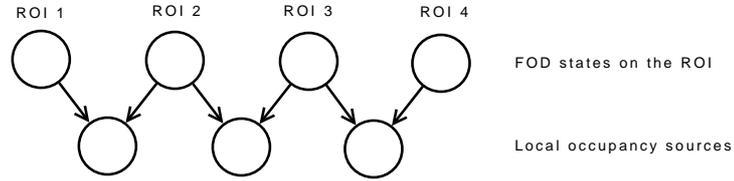
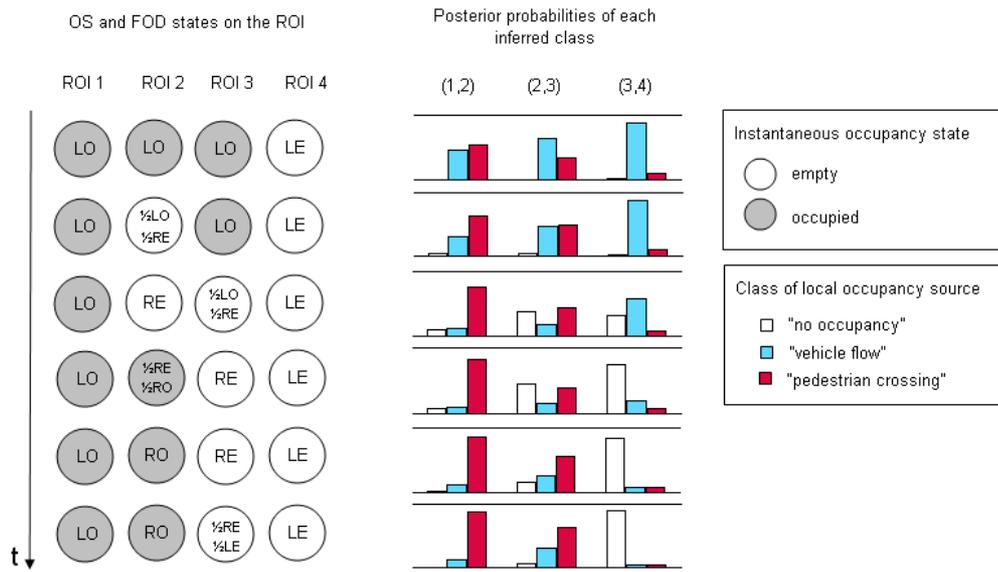


Figure 6: Conversion of an Occupancy State sequence into Fuzzy Occupancy Duration states where R: Recently, L: Lengthily, E: Empty and O: Occupied.



(a) Bayesian architecture



(b) Bayesian inference on a sequence with state transitions

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

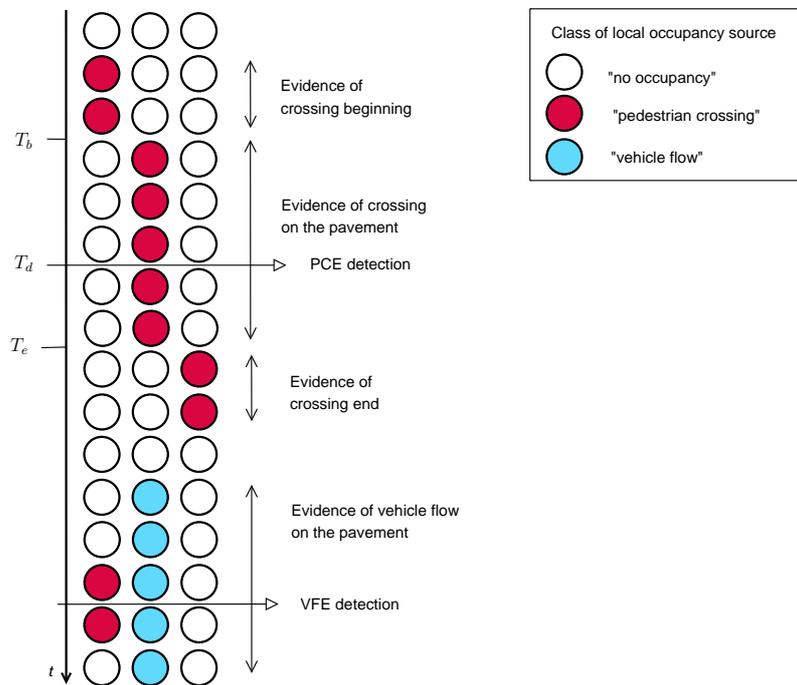


Figure 8: Illustration of the global pattern recognition process based on temporal inferences in the 3 local occupancy source nodes. The inner nodes (only one when $n = 4$) are linked to ROI on the pavement.

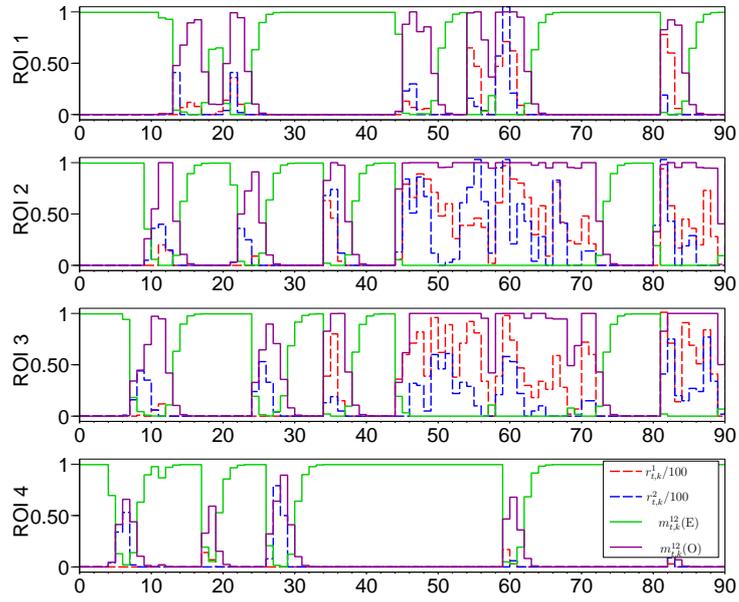


Figure 9: An example of sensor measurement and the results of the data fusion process; the time scale is in seconds.

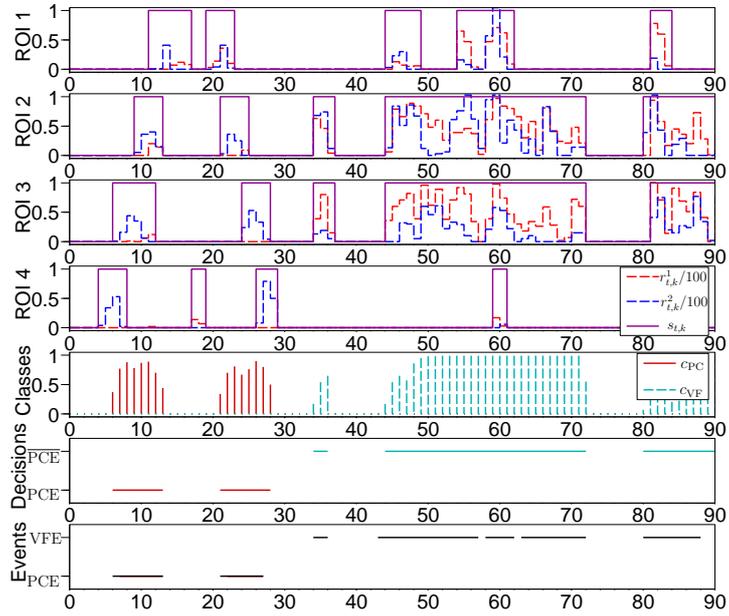


Figure 10: Pignistic decisions on occupancy state on the 4-ROI crosswalk - the state occupied is represented by $s_{t,k} = 1$ and the state empty by $s_{t,k} = 0$ - (graphs 1 to 4), local and global pattern recognition results (graphs 5 to 6), and real events (graph 7).

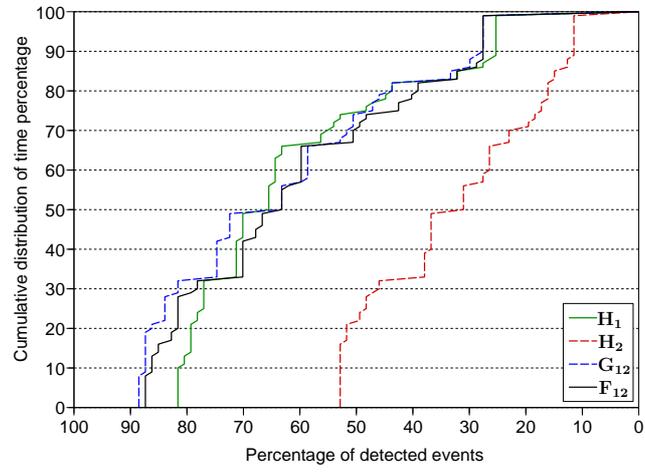


Figure 11: Cumulative distribution of time percentage of PCE detection.

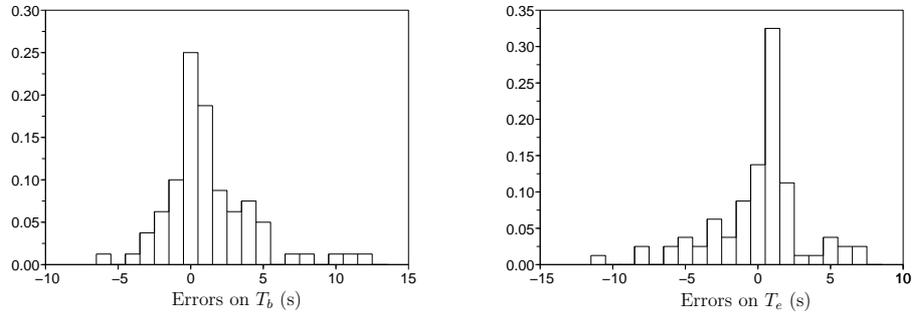


Figure 12: Distribution of errors between real PCEs and PCE decisions for the configuration \mathbf{F}_{12} : errors on the beginning time (left) and end time (right).

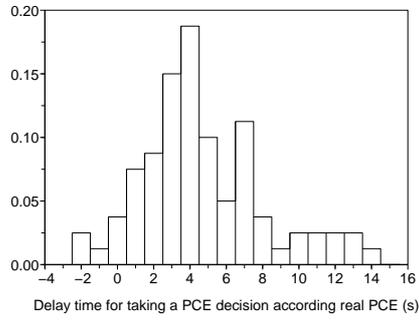


Figure 13: Distribution of delays of PCE decisions according to the real PCEs for the configuration \mathbf{F}_{12} .

635 **List of Tables**

636	1	Data fusion processes used according to the configurations . . .	45
637	2	PCE detection results according to test configurations (with	
638		the number of examples)	46
639	3	PCE detection rate obtained according to type of crossing and	
640		test configuration (with the number of examples)	47

Table 1: Data fusion processes used according to the configurations

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

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

Test configuration	S₁	S₂	H₁	H₂	G₁₂	F₁₂	#
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	S₁	S₂	H₁	H₂	G₁₂	F₁₂	#
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