



**HAL**  
open science

## Real-time passenger counting in buses using dense stereovision

T Yahiaoui, L Khoudour, C Meurie

► **To cite this version:**

T Yahiaoui, L Khoudour, C Meurie. Real-time passenger counting in buses using dense stereovision. Journal of Electronic Imaging, 2010, vol19, issue 3, 11p. 10.1117/1.3455989 . hal-00855979

**HAL Id: hal-00855979**

**<https://hal.science/hal-00855979v1>**

Submitted on 30 Oct 2013

**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.

# Real-time passenger counting in buses using dense stereovision

**Tarek Yahiaoui**

University Lille 1, Sciences and Technology  
LIFL Laboratory  
FOXMIIRE Team  
IRCICA Building  
Halley Avenue  
Parc scientifique de la Haute-Borne  
F-59650 Villeneuve d'Ascq, France  
E-mail: tarek.yahiaoui@lifl.fr

**Louahdi Khoudour**

French National Institute for Transport and Safety Research  
LEOST Laboratory  
20 rue Elisee Reclus, BP 317  
F-59666, Villeneuve d'Ascq Cedex, France

**Cyril Meurie**

University of Technology of Belfort-Montbeliard  
Systems and Transportation Laboratory  
ICAP Team  
13 rue Ernest-Thierry Mieg  
F-90010 Belfort Cedex, France

**Abstract.** We are interested particularly in the estimation of passenger flows entering or exiting from buses. To achieve this measurement, we propose a counting system based on stereo vision. To extract three-dimensional information in a reliable way, we use a dense stereo-matching procedure in which the winner-takes-all technique minimizes a correlation score. This score is an improved version of the sum of absolute differences, including several similarity criteria determined on pixels or regions to be matched. After calculating disparity maps for each image, morphological operations and a binarization with multiple thresholds are used to localize the heads of people passing under the sensor. The markers describing the heads of the passengers getting on or off the bus are then tracked during the image sequence to reconstitute their trajectories. Finally, people are counted from these reconstituted trajectories. The technique suggested was validated by several realistic experiments. We showed that it is possible to obtain counting accuracy of 99% and 97% on two large realistic data sets of image sequences showing realistic scenarios. © 2010 SPIE and IS&T. [DOI: 10.1117/1.3455989]

## 1 Introduction

The considerable development of passengers traffic in public transportation has made it indispensable to set up specific methods of organization and management. For this

reason, public transport companies are very much concerned with counting passengers,<sup>1</sup> which allows improved diagnosis of fraud, optimization of line management, traffic control and forecast, budgetary distribution between the different lines, and improvements in the quality of service. Therefore, developing a reliable passenger counting system becomes an important issue. Counting objects under controlled conditions, such as in manufacturing, is relatively easy, but counting people is much more difficult, especially under highly variable realistic environmental and operational conditions. Counting should be carried out with good accuracy, i.e., at least  $\pm 3\%$  with a confidence rate of 95%. Accuracy and reliability should be consistently maintained throughout the counting process.

In France, several counting systems have been tested or are currently being tested in buses of the RATP, the Parisian transport operator. According to the results of these tests, the system must either be improved or replaced with a more accurate one. This is particularly necessary where fraud (people using buses without tickets) is concerned. The conclusion is that manual counting is carried out for one week every, on each bus line, in order to have an accurate evaluation of the traffic.

Nonetheless, technological progress has greatly improved systems of counting passengers. For example, the RATP has chosen a system with integrated infrared cells. Two types of cells, developed by ACOREL and ELINAP,

Paper 09128SSR received Jul. 17, 2009; revised manuscript received Dec. 7, 2009; accepted for publication Jan. 25, 2010; published online Dec. xx, xxxx.

75 were initially tested by the RATP. These two solutions were  
76 not considered to provide sufficiently accurate counting.  
77 Thus, in 1996, a third type of cell, developed by BRIME,  
78 was considered to be sufficiently accurate and was installed  
79 in all the new vehicles.

80 Currently, RATP uses two types of automatic counting:  
81 ELINAP cells installed in 1500 vehicles (see <http://www.acorel.com>, for more details) and the BRIME systems  
82 installed in around 1000 vehicles (see <http://www.brime-sud.fr>, for more details). It is clear from this paragraph that  
83 RATP has been looking for automatic passenger counting  
84 systems for many years. The company has tested many of  
85 these without obtaining satisfactory results and now must  
86 carry out manual countings to readjust the automatic ones,  
87 which get less accurate over time. As far as we know, there  
88 are currently no systems in France that allow counting of  
89 passengers with an accuracy of >95% in buses. A study of  
90 the reliability of different systems of counting enables us to  
91 conclude that the two most reliable approaches:

- 94 1. The use of infrared directional sensors
- 95 2. Video sensing and image processing

96 Infrared directional sensors have a number of advan-  
97 tages, which explain their use in several systems of  
98 counting.<sup>2</sup> The major advantages are reduced size and cost,  
99 easy installation, and reliability. However, in crowded situ-  
100 ations, their high sensitivity to noise, to variations in tem-  
101 perature, and to dust and smoke makes them less reliable in  
102 real-life situations. Moreover, they cannot distinguish be-  
103 tween one passenger and a group of passengers, which is a  
104 huge drawback for counting in a bus. Thus, when counting  
105 passengers in a bus, a highly accurate system is necessary,  
106 particularly during rush hours. We believe that video-based  
107 systems are very promising for this task.

108 People counting using video is not a recent approach; we  
109 found in the literature many works dealing with this issue.  
110 The proposed techniques are various; however, based on  
111 their basic principle as a classification criterion, we distin-  
112 guish the following classes:

- 113 1. Motion detection and analysis-based techniques:  
114 These can be described by a succession of two stages.  
115 The first one is to detect moving regions in the scene  
116 corresponding mostly to individuals. The second step  
117 uses the result of detection to rebuild over time, the  
118 trajectories of moving objects. The trajectory analysis  
119 is used to identify and count the people who crossed  
120 a virtual line or a predefined area.<sup>3-5</sup>
- 121 2. Edge analysis-based techniques: As their name sug-  
122 gests, these techniques exploit the extraction of edges  
123 for the detection. The objects of interest, in this case,  
124 correspond to a set of edges with a particular shape  
125 and organization. For example, a head corresponds to  
126 an edge with a circular shape.<sup>6-8</sup>
- 127 3. Model based techniques: These techniques attempt to  
128 find regions in the processed images that match pre-  
129 defined templates.<sup>9,10</sup> These models are either charac-  
130 teristics models or appearance models. The disadvan-  
131 tage of these approaches is either the need of a large  
132 learning database or a problem of model generaliza-  
133 tion.
- 134 4. Spatiotemporal techniques: These involve the selec-

tion of lines of interest in the acquired images and  
build on each line a space-time card by stacking lines  
in time. A second step is to use statistical models  
(templates) to derive the number of persons crossing  
the line and to analyze the discrepancies between the  
space-time maps in order to determine the  
direction.<sup>11,12</sup> These techniques have the advantage of  
being fast and simple to implement; however, works  
based on these techniques have not provided concrete  
solutions to interpret a significant number of cases.  
For example, the “blob” generated by a stationary  
person can be interpreted as that of several people.

Some researchers have been working in the field of  
counting people with monocular vision systems<sup>13,14</sup> and  
some with sets of video cameras scattered in the  
environment.<sup>15,16</sup> In the transport field, a system was devel-  
oped by Mecoci *et al.*<sup>17</sup> to count passengers entering and  
exiting from buses. The authors claim that their system  
reaches a counting accuracy of 98%, but the evaluation  
presented in their paper was performed on a very reduced  
data set. Very few complete systems exploiting optical sen-  
sors and used in operation in transport context exist nowa-  
days. Among these, we can mention the system developed  
by Albiol and Naranjo from Valencia University in Spain,<sup>18</sup>  
which provided interesting results. This system uses a  
single camera installed above the train doors of the RENFE  
railway network. The author announces a counting accu-  
racy of 98% on realistic data sets corresponding to 149  
train stops. The disadvantage of this system is that it mis-  
takes an object and a large person, and the results are ob-  
tained using a correction factor. Given recent advances in  
computer vision and decreasing prices of hardware, the use  
of stereo vision is attractive. This approach is less sensitive  
to illumination changes and could also provide the neces-  
sary information to detect, model, and track objects or  
people. For all these reasons, we have chosen to develop a  
system based on dense stereo vision. However, we will see  
that stereo vision does not solve all the problems related to  
our application. In particular, the stereo matching could be  
very difficult for some cases.

This paper is organized as follows: In Section 2, we  
recall the basic aspects of stereo vision and show the inter-  
est of dense stereo vision for people counting. We also  
describe the hardware part of our system and present the  
overall structure of our image-processing chain. In Section  
3, we present the similarity constraints enhancing the sum  
of absolute differences (SAD) score and compare the pro-  
posed stereo-matching technique with other methods on  
common images of the literature. Section 4 is devoted to  
the description of the other links of the processing chain:  
height map segmentation and feature tracking. In Section 5,  
we present the evaluation of our system on a laboratory  
data set, including various image sequences showing real-  
istic scenarios, and on a real data set. Finally, a conclusion  
and a description of possible future work are provided in  
Section 6.

## 2 Stereovision for Counting Passengers 191

Stereo vision is a well-known method based on the analysis  
of several images (usually two) of the same object taken  
from different angles, along the optical axis of the camera 194

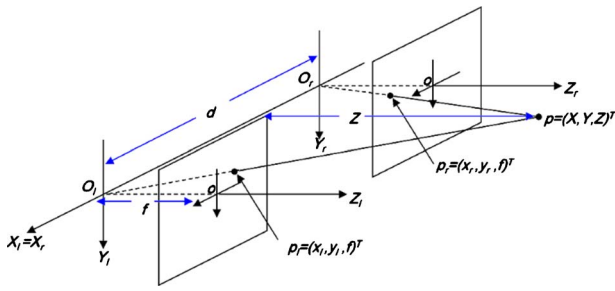


Fig. 1 Geometric modeling of binocular stereoscope.

195 (axial stereo vision), or by moving the acquisition system  
 196 sideways (lateral stereo vision). Passive stereo vision oper-  
 197 ates a set of two (binocular vision) or three (trinocular vi-  
 198 sion) stereoscopic images.<sup>19</sup> It is static when observed ob-  
 199 jects do not move and dynamic where the objects can  
 200 move.

201 In Section 2.1, we present the principle of the adopted  
 202 binocular stereo vision. Then, we describe the hardware  
 203 structure of the people-counting setup.

204 **2.1 Stereovision Vision Principles**

205 Figure 1 shows a typical stereo-vision setup, in which op-  
 206 tical axes of the two cameras are parallel. The distance  $d$   
 207 between these optical axes is called the baseline of the  
 208 stereo-vision setup. It is generally assumed that the two  
 209 cameras have exactly the same focal distance  $f$ . A region of  
 210 the scene exists in which points are visible by both cam-  
 211 eras. In the image-formation process, a point  $P$  of this re-  
 212 gion is projected onto a pixel  $P_1$  of the image sensor of the  
 213 left camera and onto a pixel  $P_r$  of the image sensor of the  
 214 right camera. Pixels  $P_1$  and  $P_r$  are called homologous be-  
 215 cause they correspond to the same point of the scene. The  
 216 disparity is defined as the difference between horizontal  
 217 positions of homologous pixels; the further the point  $P$  is  
 218 from the cameras, the smaller the disparity is. Stereo-vision  
 219 techniques aim at recovering various information about the  
 220 real scene using only the visual data contained in the two  
 221 images. This problem is not trivial since the pairs of ho-  
 222 mologous pixels are not known *a priori*.

223 Usually, stereo-vision techniques include two parts: ste-  
 224 reo matching and 3-D reconstruction. For passenger count-  
 225 ing in buses, because the sensor is very close to persons  
 226 passing under it, it is difficult to extract particular points  
 227 (such as curves) and segments, and to match them. We have  
 228 tested some well-known sparse stereo-vision algorithms on  
 229 our data set,<sup>20-22</sup> without success for features extraction.  
 230 With a dense stereo approach, we will show later that it is  
 231 possible to reconstruct a height map, in which the heads of  
 232 people can be easily located.

233 **2.2 Our People Counting System**

234 The global system is composed of an acquisition part and a  
 235 processing part. The acquisition device is an industrial ste-  
 236 reoscopic sensor called bumblebee (manufactured by the  
 237 PointGrey Company), fixed vertically above the entrance of  
 238 the bus at a height of 235 cm with a baseline of 12 cm. The  
 239 processing chain, which counts people passing under the  
 240 system using the images acquired by the hardware setup, is  
 241 composed of the following links:

1. A stereo-matching block that computes the disparity 242  
 map for each pair of images. This map is then trans- 243  
 formed into a height map for further processing. 244
2. A segmentation block that identifies, in the height 245  
 map, heads of people by detecting round shapes with 246  
 a constant height value. 247
3. Tracking and counting modules that reconstruct the 248  
 trajectories of people’s heads using the round shapes 249  
 marked in successive stereo pairs. A person is 250  
 counted by this module when the trajectory of his/her 251  
 head enters or leaves the stereo field of view. 252

The key point of this processing chain is the computa- 253  
 tion of precise and accurate height maps. The proposed 254  
 dense stereo-matching approach is described in Section 3. 255  
 The other steps of the processing chain (i.e., segmentation 256  
 and marker tracking for trajectory reconstruction) will be 257  
 described later. 258

3 Improved Stereo Matching 259

3.1 Principles of SAD Matching Cost 260

The dissimilarity measure, also called correlation, is one of 261  
 the most widely used techniques for determining all the 262  
 homologous pixels. It consists of defining a neighborhood, 263  
 around each pixel of the right image, and measuring the 264  
 resemblance between it and the same neighborhoods sur- 265  
 rounding pixels of the left image. We calculate for each 266  
 pixel of the left image a dissimilarity curve as a function of 267  
 the shift that defines the minimum and maximum dispari- 268  
 ties allowed by the imaging system. In the case of the SAD 269  
 matching cost [winner-takes-all (WTA) algorithm],<sup>23,24</sup> the 270  
 dissimilarity measurement corresponds to the absolute dif- 271  
 ference defined by Eq. (1). Thus, the shift corresponding to 272  
 the minimum value of the dissimilarity curve marks the 273  
 pixel supposed to be the homologous one of the pixel of the 274  
 left image that we try to match, 275

$$C_{SAD}(x, y, s) = \sum_{ij} |G(x + i + s, y + j) - D(x + i, y + j)|. \quad (1) \quad 276$$

where  $G(x, y)$  is the gray level of the pixel  $(x, y)$  we want to 277  
 match and that belongs to the left image,  $D(x, y)$  is the gray 278  
 level of the pixel  $(x, y)$  in the right image,  $s$  is the shift 279  
 between the two pixels (left and right), and  $d$  is the dispar- 280  
 ity that corresponds to the shift-minimizing  $C_{SAD}$  criterion 281  
 defined in Eq. (1). 282

The advantage of the SAD matching cost (WTA algo- 283  
 rithm) described above is that it is simple to implement, 284  
 robust and fast enough to operate in real time.<sup>25</sup> However, 285  
 some matching errors are caused by this approach, which 286  
 leads to an incorrect disparity value on some given pixels. 287  
 In addition, one of the major drawbacks of this method is to 288  
 systematically yield a matching result even if the area of 289  
 the scene is partially or totally occluded, in which case 290  
 these results are false. Thus, in order to reduce the number 291  
 of matching errors, we propose an approach, based on the 292  
 SAD matching cost (WTA algorithm), in which we impose 293  
 constraints for the selection and better matching of the 294  
 neighborhoods.<sup>26</sup> This improves the matching, taking into 295  
 account various types of areas: hidden, not hidden, and un- 296  
 der the influence of illumination changes. 297

**298 3.2 Improvements Brought to the SAD Matching**  
**299 Cost (WTA Algorithm)**

**300** Four similarity constraints are introduced to improve the  
**301** matching process with the WTA algorithm.

**302 3.2.1 Similarity of the gray levels of pixels to be**  
**303 matched**

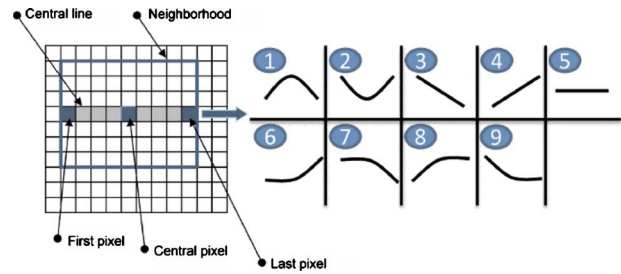
**304** The first similarity criterion between two homologous pix-  
**305** els is the similarity of their gray levels. When using square  
**306** or symmetric rectangular neighborhoods, we consider the  
**307** pixel to match as the center of the first calculation neigh-  
**308** borhood, called fixed, and the candidate pixel as the center  
**309** of the second calculation neighborhood, called sliding. The  
**310** aim of this constraint is to increase the matching accuracy  
**311** by promoting the matching of the most similar pixels. This  
**312** is achieved by promoting a minimum compared to others in  
**313** the case of multiple minima of the dissimilarity curve (for  
**314** example, in the case of repetitive textures). We call  $\alpha$  the  
**315** coefficient assigned to this similarity criterion. This coeffi-  
**316** cient can take only two values, depending on whether the  
**317** constraint is introduced or not. We look for the pixel that  
**318** minimizes the dissimilarity criterion of Eq. (2). Thus, for a  
**319** shift satisfying the constraint, the introduction of the coef-  
**320** ficient  $\alpha$  will further minimize the value of dissimilarity.  
**321** We propose a simple multiplication of the coefficient  $\alpha$  and  
**322** the dissimilarity term of Eq. (2). Let us call this expression  
**323**  $C_1$ . In order to make the overall term lower when the con-  
**324** straint is introduced, it is necessary that the particular value  
**325** that  $\alpha$  takes when the constraint is introduced be  $< 1$ .

**326** 
$$C_1(x, y, s) = \alpha \times \sum_{ij} |G(x + i + s, y + j) - D(x + i, y + j)|, \quad (2)$$

**327** where  $\alpha=1$  if the constraint is not verified and  $\alpha=\alpha_0$   
**328** knowing that  $0 < \alpha_0 < 1$ , if the constraint is introduced. We  
**329** consider that the constraint is introduced if the difference  
**330** between the gray levels does not exceed a given threshold,  
**331** fixed experimentally.

**332 3.2.2 Stereo matching of pixels belonging to**  
**333 identified edges**

**334** We also use an additional similarity criterion to deal with  
**335** the matching of edge pixels. These pixels have a higher  
**336** probability to correspond to regions of hidden areas or  
**337** near-hidden (occluded) regions. Usually, in stereo vision,  
**338** we can reasonably assume that if a pixel corresponds to an  
**339** edge, so does the homologous pixel. On the basis of this  
**340** assumption, we can introduce this constraint to try to im-  
**341** prove the matching of pixels corresponding to these edges.  
**342** Edge pixels are extracted using a classical Laplacian-based  
**343** technique.<sup>27</sup> Because of the difficult application environ-  
**344** ment (occlusion, high illumination variation), good detec-  
**345** tion is hard to achieve. However, even though it is not  
**346** perfect, we use this information. Therefore, there is no need  
**347** to develop a complex approach to obtain it. As with the  
**348** previous constraint, we have associated a weighting factor  
**349** called  $\beta$  to this similarity criterion. Let us call the expres-  
**350** sion linked to this constraint  $C_2$



**Fig. 2** Profiles for the gray levels of the pixels belonging to the central lines of the calculation neighborhoods.

$$C_2(x, y, s) = \beta \times \sum_{ij} |G(x + i + s, y + j) - D(x + i, y + j)|, \quad (3) \quad \mathbf{351}$$

where  $\beta=1$  if the constraint is not introduced and  $\beta=\beta_0$  **352**  
 knowing that  $0 < \beta_0 < 1$ , if the constraint is introduced. **353**

**3.2.3 Similarity of simplified gray-level profiles of** **354**  
**the pixels corresponding to the centerlines of** **355**  
**calculation neighborhoods** **356**

We define an additional similarity criterion in analyzing **357**  
 simplified gray-level profiles of the pixels of the center **358**  
 lines of the two calculation neighborhoods. Figure 2 provides **359**  
 the main simplified gray-level profiles for a given **360**  
 window size. The gray level profiles of the center lines of **361**  
 the two calculation neighborhoods are analyzed and compared. **362**  
 If the two gray-level profiles correspond to homologous **363**  
 pixels, the two-gray-level curves should have the **364**  
 same profile. **365**

We associate to this new constraint the weighting factor **366**  
 $\gamma$ . Let us call the expression linked to this new constraint **367**  
 $C_3$ , **368**

$$C_3(x, y, s) = \gamma \times \sum_{ij} |G(x + i + s, y + j) - D(x + i, y + j)|, \quad (4) \quad \mathbf{369}$$

where  $\gamma=1$  if the constraint is not introduced and  $\gamma=\gamma_0$  **370**  
 knowing that  $0 < \gamma_0 < 1$ , if the constraint is introduced. **371**

**3.2.4 Use of motion** **372**

The motion-detection approach is based on the subtraction **373**  
 of a background image. The motion detection is carried out **374**  
 for both images. Before matching, we classify the pixels of **375**  
 the left and right images into two classes, based on whether **376**  
 or not the pixels belong to regions affected by motion. The **377**  
 basic idea is to introduce, as with the previous similarity **378**  
 constraints, a coefficient called  $\mu$  in the dissimilarity crite- **379**  
 rion (called  $C_4$ ). This coefficient will favor homologous **380**  
 pixels belonging to the same class of regions: moving or **381**  
 static. This also drastically lowers the computation time by **382**  
 matching only pixels belonging to moving areas, **383**

$$C_4(x, y, s) = \mu \times \sum_{ij} |G(x + i + s, y + j) - D(x + i, y + j)|, \quad (5) \quad \mathbf{384}$$

where  $\mu=1$  if the constraint is not introduced and  $\mu=\mu_0$  **385**  
 knowing that  $0 < \mu_0 < 1$ , if the constraint is introduced. **386**

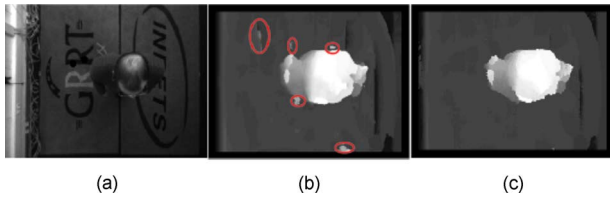


Fig. 3 Example of disparity maps calculated on a pair of images: (a) Left image, (b) SAD, and (c) our method.

### 3.2.5 Associations of constraints

Thus far, we have proposed four similarity constraints to improve the accuracy of pixel matching. Knowing that each of these constraints is of a different nature, it becomes interesting to combine these various similarity criteria to increase the robustness of the matching process and analyze their respective values. In other words, we simultaneously do the following:

1. Compare the similarity or dissimilarity of neighborhoods corresponding to the pixel to match and the candidate pixel
2. Check if their gray levels are similar
3. Test if they belong to edges
4. Verify whether the gray-level profiles of central lines of calculation neighborhoods are similar
5. And, finally, test if they both belong to a region affected by motion

We can find in the literature diverse techniques allowing the association of several criteria in order to optimize a global one. The most used optimization criteria are based on genetic algorithms, fuzzy logic, analysis of variance, decision trees, and derivative approaches. The optimization technique choice should meet a compromise between the complexity of the problem to solve and the optimization result.

In our case, we consider that the similarity criteria are of a different nature and are more or less independent. Thus, we chose to use an additive model for the calculation of dissimilarity, which corresponds to summing the dissimilarity of four criteria,

$$C(x,y,s) = C_1(x,y,s) + C_2(x,y,s) + C_3(x,y,s) + C_4(x,y,s), \quad (6)$$

where  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_4$  match dissimilarity in the order they were presented. The global formulation becomes

$$C(x,y,s) = (\alpha + \beta + \gamma + \mu) \times \sum_{ij} |G(x+i+s,y+j) - D(x+i,y+j)|. \quad (7)$$

Figure 3 provides two disparity maps calculated with the SAD alone and with the four constraints together, on a pair of stereoscopic images. We note that for SAD some matching errors appear (marked with ellipses). This visually shows the improvement brought by the introduction of constraints in the SAD model.

To test the relevance of our algorithm, we compared our approach to classical approaches having the same complex-

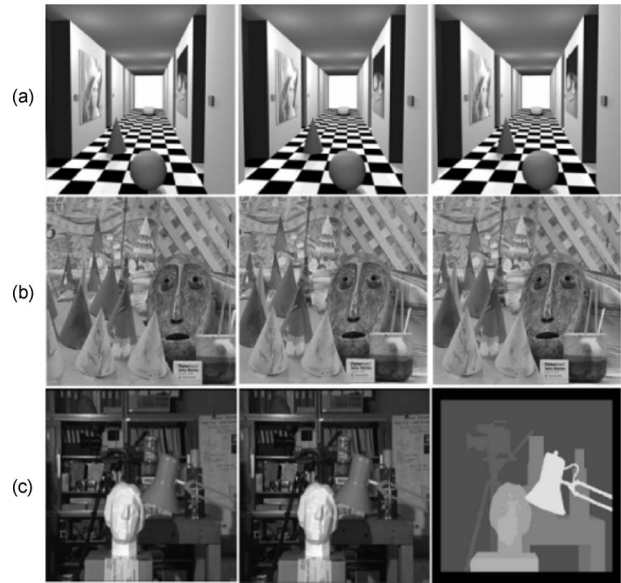


Fig. 4 Pair of stereoscopic images for comparison: (a) Corridor of Lena, (b) cones, and (c) Tsukuba.

ity and calculation time as ours. We retained methods using the following statistical distances: SAD, zero mean SAD, sum of squared differences (SSD), and zero mean SSD. The algorithms with which we conduct a comparison are those proposed by Scharstein and Szeliski. In the framework of this paper, we only provide results on the evaluations of the first three constraints ( $C_1$ ,  $C_2$ , and  $C_3$ ) because we only have single images with ground truth and thus cannot compute motion. Therefore, the  $C_4$  constraint, which requires motion detection, is not used in this comparison. The first stereoscopic images of the test are a couple of synthetic images (Corridor of Lena in Fig. 4). The second stereoscopic pair is relatively difficult to match because of the complex and repetitive textures (Cones in Fig. 4). The third stereoscopic pair of images is a view of a natural scene. The main difficulties of matching pixels of this pair of images is a highly textured background and many occlusions (Tsukuba in Fig. 4). In Fig. 4, for each case, we show left and right images and the disparity map representing the ground truth.

Our algorithm is compared to SAD matching cost (WTA algorithm) and its family following two criteria: with the ground truth, we calculate the number of pixels correctly matched to the total number of candidate pixels. This is achieved separately for occluded and nonoccluded pixels. For each pair of images tested, the best values of the parameters  $\alpha_0=0.85$ ,  $\beta_0=0.85$ ,  $\gamma_0=0.90$ , and  $\mu_0=0.80$  with a neighborhood of  $15 \times 15$  pixels. The coefficients and neighborhood values corresponding to those minimize the matching-error rate curves. The overall results are as follows:

1. Each of the constraints taken independently from the others reduces the matching error rate of mapping.
2. By combining the three constraints, we obtain the best results.
3. By varying the size of the calculation neighborhood from  $3 \times 3$  pixels to  $21 \times 21$  pixels, the matching er-

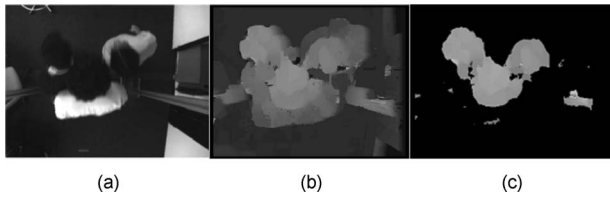


Fig. 5 Artifacts elimination by morphological filtering: (a) Left image, (b) disparity map, and (c) result of smoothing.

ror rate decreases to reach a minimum corresponding to an average calculation neighborhood size (often  $15 \times 15$  pixels), and then it increases. The effect of the three constraints together on the real Cones and Tsukuba images (gain of 3%) are the most important, especially on occluded pixels.

#### 473 4 Segmentation and Tracking

474 In Section 3.2, we described an improved stereo-matching  
475 method that allows the computation of precise and noise-  
476 free height maps. These maps are segmented in order to  
477 detect heads of people, and the marked areas are tracked  
478 across the image sequence.

479 In Fig. 5, we can see the processing carried out and the  
480 results obtained: for a given disparity map in Fig. 5(b), a  
481 threshold is first applied to retain only the parts of the im-  
482 age close to the camera; the result is displayed in Figs. 5(c)  
483 and 6(a). Then, a binarization and size-based artifact re-  
484 moval yields the binary image in Fig. 5(b). One more pro-  
485 cessing step is necessary to highlight the heads of people.  
486 For this, we use binary mathematical morphology. Three  
487 opening operations are applied to the binary images with a  
488 circular structuring element. As with every morphological  
489 filtering, the size of the structuring element is very impor-  
490 tant. The result is shown in Fig. 6(c). We can see in Fig.  
491 6(a) that the majority of the artifacts have disappeared. The  
492 result is satisfactory because we get three different kernels  
493 corresponding exactly to the heads of the persons if we  
494 compare to the original images.

495 For a given stereo configuration, we can define a statis-  
496 tical average size of a head on the image as a function of  
497 the distance that separates the human head from the cam-  
498 eras. This means that we cannot use the same structuring  
499 element for segmenting heads of people having different  
500 heights. To deal with this problem, we define several height  
501 intervals corresponding to different height classes. For each  
502 class, we use a specific structuring element having a size  
503 equivalent to the average size of a head, based on the height

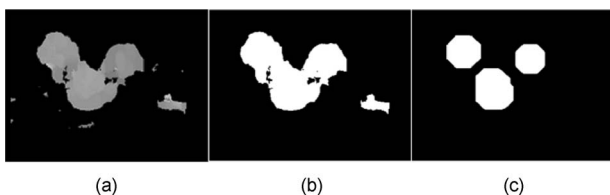


Fig. 6 Use of binary mathematical morphology for the disparity map segmentation: (a) Result of smoothing of the previous step, (b) binary image, and (c) kernels results.

and, therefore, on the distance from the camera. Given the  
variability of people’s heights, defining the number of  
height classes is not easy. This number has a strong influ-  
ence on the quality of the result; thus, it must be chosen  
carefully. It must be large enough to represent the majority  
of people’s height classes and not too large to avoid in-  
creasing the processing time. Experimentally, we found that  
four classes are a good compromise.

These classes are used for thresholding the disparity  
map, and in the same way as shown in Fig. 6, morphologi-  
cal tools are then applied to each thresholding result to  
segment the heads of people. For a given class, the size of  
the kernels resulting from this segmentation step leads to  
differentiate objects larger than the average head size of the  
class. Then, the differentiation between large objects and  
head is carried out by the tracking procedure.

The tracking of the kernels for the final counting is per-  
formed using a Kalman filter.<sup>34</sup> Each kernel resulting from  
the segmentation of the disparity maps is represented by a  
vector of the following seven components:

1. Number of pixels
2. Width of the kernel in pixels
3. Length of the kernel in pixels
4. Average height calculated from the heights of each pixel
5. Average gray level
6. Abscissa in the image
7. Ordinate in the image

The main aim of the tracking algorithm in this case is to  
track the kernels in the processing zone (called also count-  
ing zone) and to analyze the behavior of the kernels (which  
are, in fact, the heads of the persons passing under the  
sensor) in the counting zone. The first step of the tracking  
procedure is the multitarget Kalman filter, which provides  
prediction of kernels positions. We assume that each target  
is represented by a vector  $X$  of two components  $(x, y)$ ,  
where  $x$  and  $y$  are the horizontal and vertical coordinates of  
kernels in the image. The prediction is made based on two  
assumptions: the speed of objects is constant and the mea-  
sures are affected by white noise. The second step corre-  
sponds to the calculation of a probability mapping. In this  
step, the estimation of the probabilities requires the predic-  
tion from Kalman filter, corresponding to horizontal and  
vertical coordinates of the targets, and the five others kernel  
parameters used without prediction. These probability mea-  
sures are also weighted by tracking hypotheses (merging,  
splitting, appearance, disappearance, ...). A similar tracking  
methodology is described in Ref. 34. We introduce, then,  
the notion of trajectory. A valid trajectory corresponds to  
somebody entering and exiting from the counting zone. The  
counting zone has an upper and lower line; the interior is  
called the tracking zone.

The valid trajectories corresponding to an entry in the  
counting zone are the following [Fig. 7(a)]:

1. Appearance of a person at the upper line of the count-  
ing zone and disappearance in the tracking zone (the  
person has entered and stays in the tracking zone:  
they are taken into account)
2. Appearance at the upper line of the counting zone  
and disappearance at the lower line of the counting

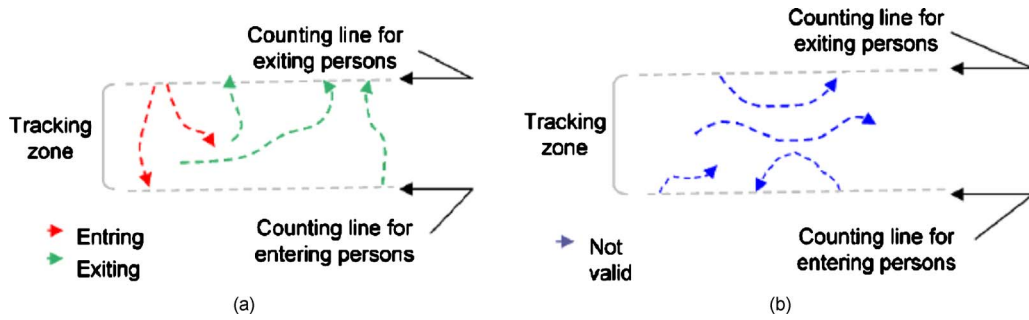


Fig. 7 Examples of (a) valid and (b) nonvalid trajectories.

564 zone (the person entered and crossed the counting  
565 zone: they are counted).

566 The nonvalid trajectories are linked to the following  
567 situations [Fig. 7(b)]:

- 568 1. Appearance at the upper line of the counting zone  
569 and disappearance at the same line (entry followed by  
570 an immediate exit)
- 571 2. Appearance at lower line and disappearance at the  
572 same line
- 573 3. Appearance and disappearance in the counting zone  
574 (wandering under the sensor without intention)
- 575 4. Appearance at lower line and disappearance in the  
576 tracking zone

### 577 5 Evaluation of the Counting System

578 The overall evaluation of the system is carried out follow-  
579 ing two directions. First of all, we are interested in the  
580 performance of the system by comparing globally the re-  
581 sults of the counting system to ground truth determined by  
582 several experts. It is a quantitative evaluation. Then, be-  
583 cause the counting is based on the notion of valid trajecto-  
584 ries, a qualitative evaluation is also carried out in order to  
585 analyze the ability of the system to manage difficult situa-  
586 tions.

### 5.1 Data Sets Used for the Evaluation

587  
588 First of all, let us mention that the counting system was  
589 entirely evaluated on real data sets. The data sets on which  
590 the system was evaluated come from two different data  
591 bases. In the framework of this paper, the data used for the  
592 evaluation includes 30 laboratory scenarios and 96 scen-  
593 arios coming from a bus.

594 Laboratory data respecting specific scenarios was pro-  
595 vided by the RATP, and 30 scenarios were simulated in our  
596 laboratory. They reflect mainly situations where people are  
597 exiting from a bus. The scenarios represent very diverse  
598 situations: high-density groups of people moving in oppo-  
599 site directions; people of different sizes, carrying bags, suit-  
600 cases, or big objects; and people with strollers. One should  
601 note here that the position of the sensor and the choice of  
602 the focal length of the lens were chosen to reproduce ex-  
603 actly the geometrical aspects of the bus. The first 15 scen-  
604 arios were simulated with ambient illumination (artificial  
605 light and daylight coming from the windows), whereas the  
606 must 15 were played with closed windows and artificial  
607 light shut off.

608 Real data coming from a bus during the exploitation  
609 period lasted for one day, on a very crowded line. The  
610 collected data represent various situations: crowd, strollers,  
611 luggage, children, and people with hats; 150 scenarios of  
612 these typical situations were collected. The processing time

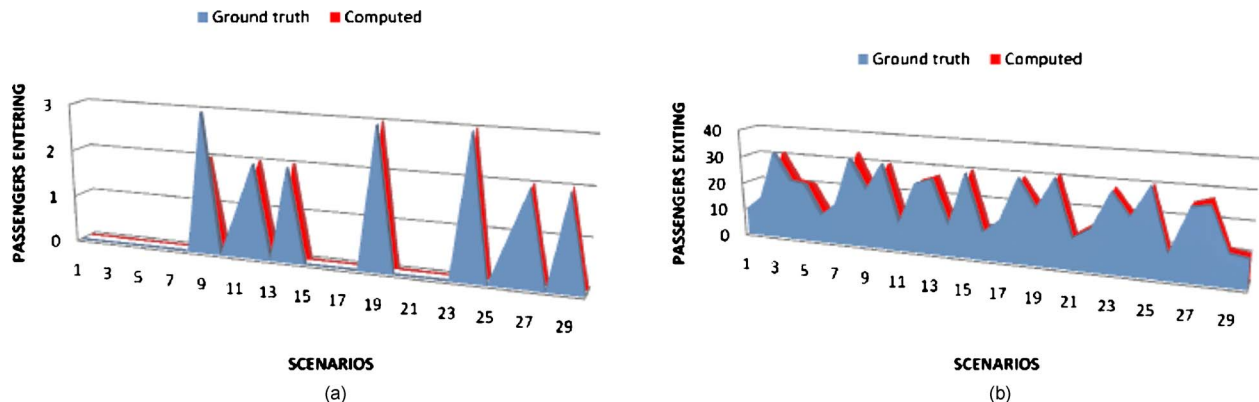


Fig. 8 Counting results for 30 scenarios in laboratory (from top to bottom): (a) entering and (b) exiting by the same door.



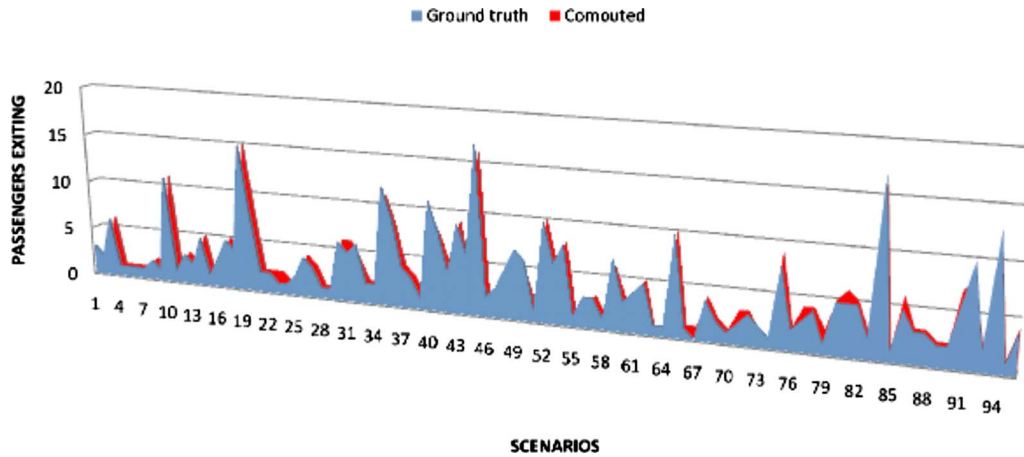


Fig. 9 Counting results for 96 scenarios in a bus.

613 is 30 fps if we consider images whose resolution is 160  
 614 × 120 pixels on a pentium IV 2 GHz. This is compatible  
 615 with our application.

616 **5.2 Quantitative Evaluation**

617 The counting results presented in Fig. 8 indicate the num-  
 618 ber of people entering or exiting for each sequence in the  
 619 laboratory. In Fig. 8, we can see the ground-truth counting  
 620 results versus the counting results computed by our algo-  
 621 rithm. One can note that whatever the difficulty of the sce-  
 622 nario is, the difference between the reference and calculated  
 623 countings is very low. Indeed, these differences are in the  
 624 interval [-1; +1]. This is an encouraging result showing the  
 625 robustness of our algorithm, which is able to cope with  
 626 diverse situations. There are fewer people entering because  
 627 the data set corresponds mainly to people exiting by the  
 628 back door, and there are counting errors because people are  
 629 entering and exiting at the same time by the same door.

630 In order to determine the accuracy of our counting sys-  
 631 tem, globally—that is to say considering all the entering  
 632 and exiting scenarios together—we have defined an error  
 633 rate that is calculated with Eq. (8). In this equation, we  
 634 consider the real counting (the ground truth obtained with  
 635 three different experts) as the basis of comparison and de-  
 636 termine the difference between the counting with the algo-  
 637 rithm. Thus, the error rate is ~1%,

638 
$$\text{Error}_{\text{counting}} = 100 \frac{(\text{Real}_{\text{counting}} - \text{Automatic}_{\text{counting}})}{\text{Real}_{\text{counting}}}. \quad (8)$$

639 The same error rate is obtained with any laboratory sce-  
 640 nario, under any illumination type. This is also encourag-  
 641 ing. For the bus data sets, the results are shown in Fig. 9.  
 642 We can note in Fig. 9 that the ground-truth results are very  
 643 close to the results after computation with our algorithm.  
 644 Even though the scenarios are much more difficult to deal  
 645 with in the bus, the overall counting error is only 3%.  
 646 When analyzing more closely the counting results, we ob-  
 647 serve that when our system differs from the reference  
 648 counting, it systematically underestimates the number of  
 649 people. Several reasons could explain this fact: the diffi-  
 650 culty to detect short people. The fixed size of the structur-  
 651 ing element in the segmentation of the disparity maps could

also be another reason. Finally, the merging of two trajec-  
 652 tories, corresponding to two different people could also be  
 653 an additional reason. Additional explanations could also be  
 654 found with a more intensive evaluation. 655

656 **5.3 Qualitative Evaluation of the Counting System**

657 After the quantitative evaluation of the system, it is inter-  
 658 esting to carry out qualitative evaluation of the algorithm  
 659 on typical image sequences. The main aim of this section is  
 660 to show the behavior of the counting system on different  
 661 trajectories of people passing under the sensor. The objec-  
 662 tive is also to verify the ability of the system to detect  
 663 specific people, to track them, and finally to count them. To  
 664 achieve this goal, we have selected three typical sequences:  
 665 two from laboratory data sets and one from a bus in normal  
 666 operation. For each sequence, we present the following  
 667 conclusions.

668 Sequence 1 represents a crowd exiting from the counting  
 669 zone while at the same time, several other people are en-  
 670 tering one behind the other (Fig. 10). The main interest of  
 671 this sequence is to show the ability of the system to analyze  
 672 the trajectories of people having the same characteristics in  
 673 terms of size and appearance. We have marked people un-  
 674 der analysis, with color ellipses: red for people exiting and  
 675 green for people entering. 675

676 Sequence 2 illustrates two people walking very close to  
 677 each other. One person puts his arm on the shoulders of the  
 678 other. This situation is illustrated in Fig. 11 in four frames.  
 679 As for the previous sequence, the heads are marked with  
 680 red ellipses. The two persons are exiting from the counting  
 681 zone.

682 Sequence 3, which is acquired in the bus, represents a  
 683 crowd getting off the bus. Among this crowd are several  
 684 children, and several other people are standing at the en-  
 685 trance without leaving the bus (typical situation in buses). 685



Fig. 10 Images taken from sequence 1: Evolution in time.

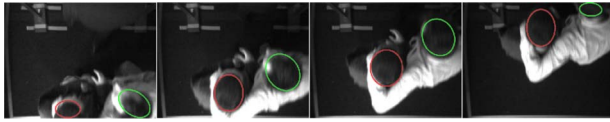


Fig. 11 Images taken from sequence 2: Evolution in time.

686 The main interest of the sequence is to test the ability of the  
 687 system to detect a young child, a stationary person, and a  
 688 person wearing a hat. Figure 12 illustrates this situation.  
 689 The green ellipse indicates the stationary person; the red  
 690 one, the child exiting from the bus; and the blue one, the  
 691 man with the hat who is also exiting from the bus.

692 **5.3.1 Tracking results**

693 The tracking results are illustrated in Figs. 13–15. The col-  
 694 ors used for drawing the trajectories are those used in Figs.  
 695 10–12.

696 In Fig. 13, which corresponds to sequence 1, we have  
 697 represented the trajectory of the person entering in contin-  
 698 uous line and the trajectory of the person exiting in dashed  
 699 line. The abscissa and ordinate in the graph represent the  
 700 spatial position, of the centers of gravity of the heads of the  
 701 passengers, in the counting area, detected during the seg-  
 702 mentation phase. Every kernel is calculated at 30 fps, but  
 703 the center of gravity is plotted only every five frames for  
 704 visual convenience. We note that, in spite of the high prox-  
 705 imity of the two people, the respective trajectories are per-  
 706 fectly identified: one entering and the other exiting. We can  
 707 also note that the trajectory of the person entering is more  
 708 rectilinear than that of the exiting person because the latter  
 709 has diverted his trajectory in order to avoid a collision.

710 In Fig. 14, we can note that the system has perfectly  
 711 dealt with the typical situation where two people are cross-  
 712 ing the counting zone very closely. We can clearly distin-  
 713 guish two parallel trajectories describing their passage.

714 In Fig. 15, we can easily note the trajectory (dashed line)  
 715 of the kid who has rapidly gotten off the bus. The contin-  
 716 uous line corresponds to the man with the hat. For this per-  
 717 son, in spite of the lack of contrast between his clothes and  
 718 the background, the system has detected the trajectory  
 719 properly. The third trajectory is typical of people standing  
 720 at the exit of the bus but moving a little, from time to time,  
 721 to let the other passengers get off the bus. That is why the  
 722 position of the center of gravity of the head moves slightly.  
 723 In Fig. 15, because the child and the man with the hat are  
 724 getting off the bus, one behind the other, the corresponding  
 725 trajectories are almost aligned.

726 **5.4 Real-Time Constraints**

727 The first version of the algorithm was implemented on a PC  
 728 Pentium IV 2 GHz and processed images of size 640  
 729 × 480 pixels. But, with this size, the algorithm was only

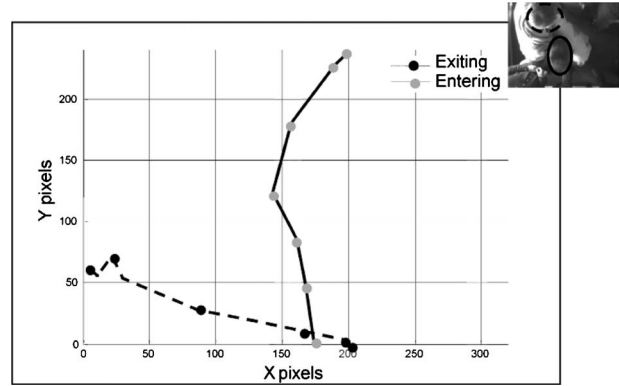


Fig. 13 Trajectories of people marked in sequence 1.

able to process up to 2 fps, and it was impossible to count  
 people moving very quickly. The real-time constraints for  
 this system are the following: Every person must be  
 counted, regardless of their speed of movement. A process-  
 ing time of 2 fps cannot be considered real time.

Therefore, in order to speed up the processing time, we  
 tried to reduce the size of the images while striving to  
 maintain the accuracy. Then, we tested two images sizes:  
 320 × 240 and 160 × 120 pixels. We have concluded that  
 the best compromise, in terms of accuracy and processing  
 time, was achieved by an image size of 160 × 120 pixels. In  
 this case, the accuracy is maintained and the processing  
 speed is 30 fps, which is compatible with a real-time imple-  
 mentation. The accuracy is not affected when we divide the  
 resolution by four moving from 640 × 480 to 160  
 × 120 pixels, which demonstrates the robustness of the al-  
 gorithm proposed.

**6 Conclusion**

In this paper, we have presented a counting system and its  
 evaluation on life-situation data sets. The comparison be-  
 tween ground-truth values and the ones calculated with our  
 algorithm leads to a counting accuracy that is around 99%  
 for laboratory and 97% for bus data sets. These values are  
 obtained on 30 scenarios coming from the laboratory and  
 96 coming from a bus during the exploitation period and  
 representing a total of ~1400 people. This counting accu-  
 racy needs to be confirmed with a more intensive evalua-  
 tion, mainly on the scenarios coming from the bus. We  
 have also conducted a qualitative evaluation in order to test  
 the ability of our algorithm to detect and track persons and  
 their trajectories in a few very difficult situations. We have  
 tested the robustness of the algorithm to deal with very hard  
 cases: very crowded situations where there are people  
 walking in two directions under the sensor.

The results obtained in these cases are very satisfactory  
 and encourage conducting us to continue working in this

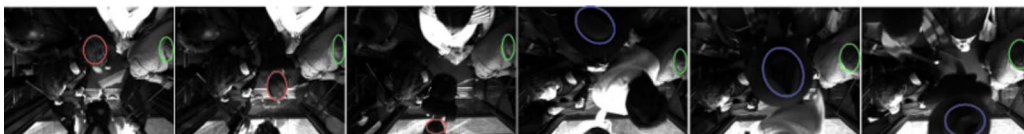


Fig. 12 Images taken from sequence 3: Evolution in time.

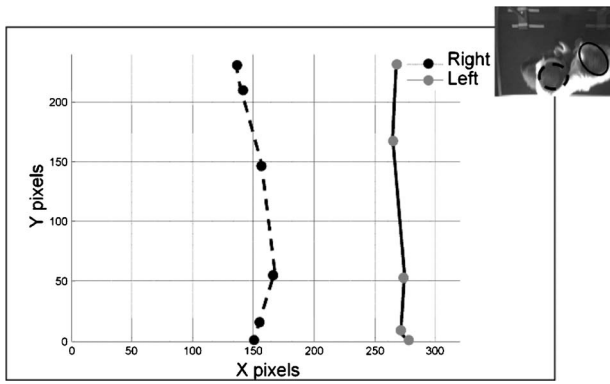


Fig. 14 Trajectories of people marked in sequence 2.

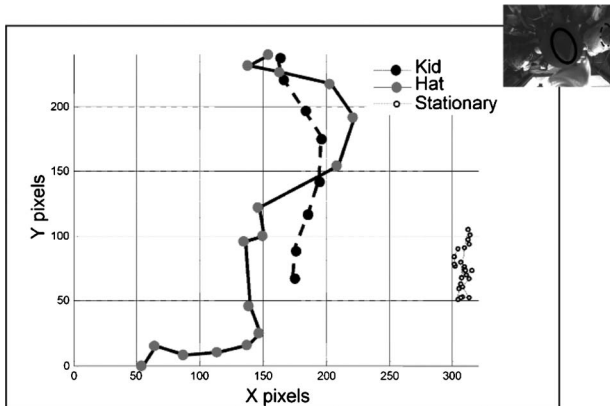


Fig. 15 Trajectories of people marked in sequence 3.

766 direction. That is why numerous perspectives are planned  
 767 in the near future. We plan, for instance, to separate the  
 768 data to assess the results in crowded situations versus non-  
 769 crowded ones. Because we wanted a real-time counting  
 770 system, from the beginning, the use of color images was  
 771 avoided because of the extra processing time they imply.  
 772 However, the use of color would provide improvements in  
 773 the choice of homologous pixels for the stereo-matching  
 774 process because we have more information for neighbor-  
 775 hood comparison. Finally, color information could be used  
 776 to perform pixel clustering of the stereoscopic images in a  
 777 number of classes which could be then exploited. For in-  
 778 stance, we could imagine adding additional constraints de-  
 779 pending on the classification results.

780 Acknowledgments

781 We thank the Paris transport operator (RATP: Regie Auto-  
 782 nome des Transports Parisiens) who funded this research  
 783 carried out by means of a Ph.D. thesis. This collaboration  
 784 between RATP, INRETS, and USTL (LAGIS laboratory,  
 785 University Lille 1, Sciences and Technology) was really  
 786 fruitful. This counting system was patented by the French  
 787 organization CNISF: National Council of Engineers and  
 788 Scientists of France under Grant No. 0953188.

789 References

790 1. D. Beymer, "Person counting using stereo," in *Workshop on Human*  
 791 *Motion*, pp. 127-133, IEEE Computer Society, Washington, DC  
 792 (2000).

2. A. Pincot, "Fiabilisation de la chaine de comptage voyageurs," RATP, Tech. Rep. RATP-MRB (Mar. 2002). 793  
 3. X. Liu, P. Tu, J. Rittscher, A. Perera, and N. Krahnstoeber, "Detecting and counting people in surveillance applications," *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 306-311, IEEE Computer Society, Washington, DC (2005). 794  
 4. E. Zhang and F. Chen, "A fast and robust people counting method in video surveillance," in *CIS '07: Proceedings of 2007 Int. Conf. on Computational Intelligence and Security*, Washington, DC, pp. 339-343, IEEE Computer Society, Washington, DC (2007). 795  
 5. X.-W. Xu, Z.-Y. Wang, Y.-H. Liang, and Y.-Q. Zhang, "A rapid method for passing people counting in monocular video sequences," in *Proc. of 6th Int. Conf. on Machine Learning and Cybernetics*, Hong Kong, pp. 1657-1662, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, WI (2007). 796  
 6. M. Bozzoli and L. Cinque, "A statistical method for people counting in crowded environments," *Proc. 14th International Conference on Image Analysis and Processing (ICIAP 2007)*, pp. 506-511, IEEE Computer Society, Washington, DC (2007). 797  
 7. A. Gardel, I. Bravo, P. Jimenez, J. Lazaro, and A. Torquemada, "Real time head detection for embedded vision modules," in *Proc. of IEEE Int. Symp. on Intelligent Signal Processing (WISP 2007)*, pp. 1-6, IEEE, Washington, DC (2007). 798  
 8. S. Yu, X. Chen, W. Sun, and D. Xie, "A robust method for detecting and counting people," in *Proc. of Int. Conf. on Audio, Language and Image Processing (ICALIP 2008)*, pp. 1545-1549, IEEE, Washington, DC (2008). 799  
 9. O. Sidla, Y. Lypetsky, N. Brandle, and S. Seer, "Pedestrian detection and tracking for counting applications in crowded situations," in *AVSS '06: Proceedings of the IEEE Int. Conf. on Video and Signal Based Surveillance*, Washington, DC, IEEE Computer Society, pp. 70-75 (2006). 800  
 10. G. Garcia-Bunster and M. Torres-Torriti, "Effective pedestrian detection and counting at bus stops," in *Proc. of Robotic Symp., IEEE Latin American*, pp. 158-163, IEEE, Washington, DC (2008). 801  
 11. J. Barandiaran, B. Murguia, and F. Boto, "Real-time people counting using multiple lines," in *Proc. 9th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '08)*, Washington, DC, IEEE Computer Society, pp. 159-162, IEEE Computer Society, IEEE Computer Society, Washington, DC (2008). 802  
 12. Y. Jeon and P. Rybski, "Analysis of a spatio-temporal clustering algorithm for counting people in a meeting," Robotics Institute, Pittsburgh, Tech. Rep. No. CMU-RI-TR-06-04 (Jan. 2006). 803  
 13. G.-P. Adriano, S.-I.-V. Mendoza, F.-N.-J. Montinola, and P.-C. Naval, "Apec: Automated people counting from video," presented at PCSC Conf. Security and Networking (2005). 804  
 14. V. Rabaud and S. Belongie, "Counting crowded moving objects," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 705-711 (2006). 805  
 15. A.-O. Ercan, A. E. Gamal, and L.-J. Guibas, "Object tracking in the presence of occlusions via a camera network," in *Proc. of 6th Int. Conf. on Information Processing in Sensor Networks*, pp. 509-518, ACM, New York (2007). 806  
 16. S. Fleck, C. Vollrath, F. Walter, and W. Straber, "An integrated visualization of a smart camera based distributed surveillance system," in *Proc. of 3rd Conf. on IASTED Int. Conf.: Advances in Computer Science and Technology*, pp. 234-242, ACTA Press, Anaheim, CA (2007). 807  
 17. A. Mecoci, F. Bartolini, and V. Cappellini, "Image sequence analysis for counting in real time people getting in and out of a bus," *Revue Signal Process.* 35, pp. 105-116 (1994). 808  
 18. A. Albiol, V. Naranjo, and I. Mora, "Real-time high density people counter using morphological tools," *IEEE Trans. Intell. Transp. Syst.* 3, 204-217 (2001). 809  
 19. S. Bahrodi, L. Iocchi, G. Leone, D. Nardi, and L. Scozafava, "Real-time people localization and tracking through fixed stereo vision," *Rev. Appl. Intell.* 26(2), 83-97 (2007). 810  
 20. Y. Zhang and C. Kambhamatteu, "Stereo matching with segmentation-based cooperation," in *Proc. of 7th European Conf. on Computer Vision*, Vol. 2, pp. 556-571, In Lecture Notes in Computer Science, vol. 2351/2002, pp. 521-522, Springer, Heidelberg (2002). 811  
 21. Y. Ruichek and J.-G. Postaire, "A new neural real-time implementation for obstacle detection using linear stereo vision," *Real-Time Imaging J.* 5, 141-153 (1999). 812  
 22. Y. Ruichek, H. Issa, and J.-G. Postaire, "Genetic approach for obstacle detection using linear stereo vision," in *Proc. of IEEE Intelligent Vehicles Symp.*, pp. 261-266 (2000). 813  
 23. S. Haris, W. Vandermark, and D.-M. Cavriila, "A comparative study of fast dense stereo vision algorithms," in *Proc. of IEEE Intelligent Vehicles Symp.*, pp. 319-324 (2004). 814  
 24. S. Wong, S. Vasiliadis, and S. Ctofana, "A sum of absolute differences implementation in FPGA hardware," in *Proc. of 28th EURO-MICRO Conf.*, pp. 183-188, IEEE, Washington, DC (2002). 815  
 25. T. Yahiaoui, F. Cabestaing, L. Khoudour, and P.-H. Leny, "Le comptage de passagers entrant et sortant d'un autobus par stereovision dense," presented at Int. Workshop: Logistique & Transport 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841  
 842  
 843  
 844  
 845  
 846  
 847  
 848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863  
 864  
 865  
 866  
 867  
 868  
 869  
 870  
 871  
 872  
 873  
 874  
 875  
 876  
 877  
 878

879 (2006).  
 880 26. T. Yahiaoui, "Une approche de stéréovision dense intégrant des  
 881 contraintes de similarité. application au comptage de passagers entrant et  
 882 sortant d'un autobus," Ph.D. dissertation, University of Lille (2007).  
 883 27. H. Maitre, Détection de contour dans les images, ([http://](http://www.tsi.enst.fr/~bloch/TDI/poly_contours.pdf)  
 884 [www.tsi.enst.fr/~bloch/TDI/poly\\_contours.pdf](http://www.tsi.enst.fr/~bloch/TDI/poly_contours.pdf) (accessed July 2,  
 885 2010).  
 886 28. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Ma-*  
 887 *chine Learning*, Kluwer, Dordrecht (1989).  
 888 29. L.-A. Zadeh, G.-J. Klir, and B. Yuan, *Fuzzy Sets, Fuzzy Logic, and*  
 889 *Fuzzy Systems*, World Scientific, 1996.  
 890 30. R.-W. Pike, Optimization for engineering systems, (2001, [http://](http://www.mpri.lsu.edu/bookindex.html)  
 891 [www.mpri.lsu.edu/bookindex.html](http://www.mpri.lsu.edu/bookindex.html)), (accessed July 2, 2010).  
 892 31. N.-J. Nilsson, Decision trees, Chapter 6 of introduction to machine  
 893 learning, (1996, <http://ai.stanford.edu/people/nilsson/mlbook.html>)  
 894 (accessed July 2, 2010).  
 895 32. P. Parpas, B. Rustem, and E.-N. Pistikopoulos, "Linearly constrained  
 896 global optimization and stochastic differential equations," *J. Global*  
 897 *Optim.* **36**(2), 191–217 (1996).  
 898 33. Scharstein and Szeliski, "High-Accuracy Stereo Depth Maps Using  
 899 Structured Light," *IEEE Computer Society Conference on Computer*  
 900 *Vision and Pattern Recognition (CVPR03)*, Madison, USA, vol. 1,  
 901 pp. 195–202. IEEE, Washington, DC (June 2003).  
 902 34. D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans.*  
 903 *Autom. Control* **24**(6), 843–854 (1979).



**Tarek Yahiaoui** electronic engineer received his PhD degree from the University of Lille, France in 2007 in the field of computer science. He is a researcher in the field of image processing applied to safety and security in public transport. He has several publications mainly about automated people counting and stereo matching. He is currently working in the field of image processing at LIFL Laboratory. University of Lille, France.



**Louahdi Khoudour** received a degree in applied mathematics from the University of Toulouse in 1992 and a Master Degree in Computer Science from the University of Toulouse in 1993. He then obtained a PhD in Control and Computer Engineering from the University of Lille in 1996. In 2006, he obtained the Leading research degree (Director of research) in Physical Sciences from the University of Paris. He is currently a researcher at INRETS (French National Institute on Transport and Safety Research). Since 1997, he has

supervised 6 Phd students (3 completed and 3 ongoing) in the field of computer vision applied to safety and security in public transport. He has been in charge of various European projects, such as Cromatica, Prismatica, Boss, Selcat, Securemetro, PANSafer dealing with safety and security aspects in guided transport systems. His main competencies are video surveillance and image processing applied to safety in public transport. He is author or co-author of around 20 papers in journals, several chapters in books, 50 international conference papers and several grants.



**Cyril Meurie** received his PhD in Computer Science, from University of Caen Basse-Normandie (Caen) France in 2005. From 2006 to 2008, he was post-doctoral researcher with the Electronic, Waves and Signal Processing Research Laboratory for Transport (LEOST) of the French National Institute for Transport and Safety Research (INRETS). He participated to the European project BOSS (On Board Wireless Secured Video Surveillance) and actually to the French project PANSafer. Since 2008, he is an associate professor with the Systems and Transportation Laboratory (University of Technology of Belfort-Montbéliard). His research interests focus on image segmentation and classification techniques for color and textured images (multi-scale and morphological methods), stereovision approaches, localization and autonomous navigation for intelligent vehicles.