Abstract  Road safety, whatever the considered environment, relies heavily on the ability to detect and track moving objects from a moving point of view. In order to achieve such a detection, the vehicle’s ego-motion must first be estimated and compensated. This issue is crucial to complete a fully autonomous vehicle; this is why several approaches have already been proposed. This study presents a method, based solely on visual information that implements such a process. Information from stereo-vision and motion is derived to extract the vehicle’s ego-motion. Ego-motion extraction algorithm is thoroughly evaluated in terms of precision and uncertainty. Given those statistical attributes, a method for dynamic objects detection is presented. This method relies on 3D image registration and residual displacement field evaluation. This method is then evaluated on several real and synthetic data sequences. It will be shown that it allows a reliable and early detection, even in hard cases (e.g. occlusions,...). Given a few additional factors (detectable motion range), overall performances can be derived from visual odometry performances.

Keywords  Stereo-vision · Visual odometry · Motion detection · Intelligent vehicles

1 Introduction

To be efficient, driver assistance systems must have a good perception of their environment. One way to build a representation of this environment can be to rely on communications. However, this raises several issues. First, it requires important infrastructure investments and/or a critical proportion of road actors to be equipped. On the other hand, one can rely on embedded sensors to provide an intelligent vehicle with autonomous perception. Many systems are exploited in prototypes, research vehicles, or even high-end commercial vehicles: LIDAR, RADAR, inertial measurement unit (IMU), dGPS, wheel encoder, and of course cameras.

Most of these sensors are very specialized and provide limited information. For instance, an IMU can only be used to provide linear and angular accelerations, and it cannot serve any other purpose. On the other hand, vision can be used for scene structure analysis (through stereo-vision), for motion analysis, for recognition, and so on. A vision-enabled system is much more future proof than a specific one, and, given the life cycle of a typical product in the automotive industry, that should be taken into account. For a more extensive overview of the development of vision, applied to intelligent vehicles, one can refer to [1].

Object detection, and more importantly mobile object detection, has been and still is an important research area. Some authors, like [2], detect potentially dangerous objects without any motion information.

Others segment directly a global motion field, computed through classical optical flow method [3]. This segmentation can be based on geometrical concerns [4], or different methods, such as tensor voting [5, 6]. These methods are completely monocular; hence, they are unable to absolutely estimate the motion and/or position of obstacles.
Though, most proposed methods rely on prior knowledge or estimation of the vehicle ego-motion, some authors assume a known and highly constrained ego-motion. For instance [7] considers only 1D translational motion and uses the fact that the translational component of the image motion is linked to real-world depth to detect image points that violate this constraint. Using the same constrained dynamic, authors in [8] detect mobile objects but are also able to derive 3D scene structure from optical flow.

Finally, most authors tend to merge information from stereo-vision and motion to achieve a detection of dynamic objects. Indeed, assuming a perfectly known ego-motion, whether by estimating it through visual odometry or by the mean of inertial sensor, one can use the knowledge of both the motion and the scene structure in to compensate or estimate the displacement due to the ego-motion.

For instance, authors in [9] check every pixel to assess their movement (or the lack thereof). However, their segmenting is prone to merging different targets and more generally to errors, because some key factors have not been sufficiently investigated. An interesting middle-level approach has been unveiled in [10]. However this method relies heavily on the availability of a dense stereo map and a dense optical flow field. As of now, most authors use some flavor of the semi-global matching procedure exposed in [11]. However, recent real-time implementation of semi-global procedures relies on FPGA or GPU [12], when a block-matching approach could be much more efficient [13].

However, as far as the authors of the present paper know, several key factors have never been investigated before. First, the impact on detection of performances and accuracy of the visual odometry procedure is never investigated. Then, most authors do not consider properly image acquisition frequency, although, this study will show that its impact is crucial, and that the faster is not necessarily the better.

This paper describes an embeddable mobile objects detection system. It relies on a cooperation between stereo-vision and motion analysis. This cooperation is exploited to estimate the sensor ego-motion through a visual odometry stage and then to identify objects as dynamic or static. Contrary to previous works, particular attention was brought to the analysis of the precision of the different stages. Notably, the impact of the precision of visual odometry on the detection procedure is studied, in terms of both detection threshold and resolving power. Also, time resolution and its impact on visual odometry, dynamic object detection and a complete embedded system are investigated.

**Paper layout** This paper will be organized as follows.

First, visual odometry will be discussed. Several existing methods will be presented. A modular method will then be presented and evaluated with respect to its different parts. Overall performances will then be evaluated.

A detection procedure will then be exposed. The proposed method consists in compensating the ego-motion and in extracting a residual displacement field. A constraint on the bounds of this displacement field is exhibited. Additionally, a procedure for integrating and refining the results over time is presented.

Then, given the performances of visual odometry, the detection procedure will be properly refined. The impact of visual odometry on several performance indices will be studied.

Additionally, the effect of time resolution on the whole system will be overlooked.

Finally, test results, obtained on real data, through in-car testing will be presented and discussed, before taking a glimpse at possible future studies.

## 2 Ego-motion extraction

As visual odometry is a very active field of research, the beginning of this section will be devoted to a brief overview of existing technics. The second part will focus on the world model and notations, and then the method will be detailed and thoroughly tested.

### 2.1 Visual odometry: a state of the art

It has already been mentioned that IMU and other odometry sensors are either too expensive or not reliable enough. Thus visual odometry has known a very active development in the last few years. Monocular methods, due to their inability to estimate absolute depths, must rely on additional constraints. For instance, authors in [14] assume a priori detection of an image plane. Such additional constraints can also be a reduction of the motion space; the study presented by [15] illustrates such a reduction.

On the other hand, stereo-vision allows working in a less constrained world. Many methods have been developed, some using rectified stereo-rigs and optical flow [16] or robust feature points [17]. The latter using a formalism developed in [18] providing a satisfactory model for motion in the disparity space, thus enabling Gaussian isotropic and homogeneous noise hypothesis. Others choose to estimate simultaneously the 6-degrees-of-freedom motion and the extrinsic parameters of the camera rig, through a tri-focal [19] or quadri-focal [20] approach. These methods are all based roughly on the same principle: for every point, the image motion can be expressed as a linear function of the 6D world motion.

Some point-to-point correspondences are extracted through optical flow or feature points matching. The resulting set of equations can then be solved using any robust evaluation technique, such as RANSAC [21], iterated least
squares and so on. Eventually, some time integration, usually through an extended or unscented Kalman filter, is performed \cite{19, 20} to improve the overall robustness. Some completely different methods, such as \cite{22} have been presented but, although promising, they do not compete with more traditional approaches. A more comprehensive review of existing visual odometry can be found in \cite{23}.

2.2 World model

The world is considered to be composed of static and dynamic objects, either rigid (e.g. a car) or not (e.g. a pedestrian). No further assumption is made. The world is described by a set of two frames of reference. The first one is labelled \( R_a \) and is absolute. The second one is labelled \( R_r \) and is relative, bound to a mobile vehicle equipped with a rectified stereo-sensor. The origin of \( R_r \) is coincident with the optical center of the vision sensor. At the initial point in time, both frames of reference are coincident. In the following, unless further notice, only two successive instants will be considered. The two frames of reference are represented in Fig. 1.

The four images to be used (left and right channels for the two points in time) will be noted: \( I_{R}^{1} \), \( I_{L}^{1} \), \( I_{R}^{0} \) and \( I_{L}^{0} \) where \( R \) (resp. \( L \)) stands for right (resp. left). By extension, and given that the disparity maps are computed with respect to the right-hand side images, the notation \( I_{R}^{0} \) will refer to the right image at \( t_0 \) and its associated disparity map.

The stereo-sensor consists of two identical vision sensors, modelled by the pinhole camera model. The baseline is noted as \( b_s \), the focal length \( f \) and the pixel size \( t_p \). The term “world point” will refer to a physical point located in the object space, whereas the term “image point” will refer to a vector constituted by a point’s coordinates in the image frame of reference and its disparity. The resulting 3D space will later on be referred as image space. In the same manner, the terms “world vector”, “image vector”, “world motion” and “image motion” will be used to differentiate physical displacement vector from their image-space counterpart.

A given world point \( M = \begin{bmatrix} X_M(t) \\ Y_M(t) \\ Z_M(t) \end{bmatrix} \) is mapped to the right-hand side disparity value.

\[
\begin{align*}
\text{image point } m &= \begin{bmatrix} x_M(t) = f \frac{X_M(t) - b_s/2}{Z_M(t)} \\ y_M(t) = f \frac{Y_M(t)}{Z_M(t)} \\ d_M(t) = f \frac{Z_M(t)}{Z_M(t)} \end{bmatrix} \quad \text{where } d \text{ is the}
\end{align*}
\]

Between \( t = t_0 \) and \( t = t_1 \), the stereo-sensor undergoes an unconstrained motion \( \vec{\Theta} = \begin{bmatrix} TX \\ TY \\ TZ \end{bmatrix} \) where \( \vec{\Theta} = \omega \begin{bmatrix} \alphaX \\ \omegaY \\ \omegaZ \end{bmatrix} \) is the rotational component of the motion and \( \vec{\Omega} = \begin{bmatrix} aX \\ aY \\ aZ \end{bmatrix} \) is the translational component of the motion. It will be assumed until further notice that all components of \( \vec{\Omega} \) are small enough to linearize trigonometric lines. The image motion due to this world motion can be expressed as:

\[
\begin{align*}
\mu(\vec{\Theta}, \vec{\Omega}) &= \begin{bmatrix} x \overline{X} \omegaX - \left( f + \frac{x^2}{f} \right) \omegaY + \omegaZ \frac{df}{bs} + \frac{dx}{bs} \\ \left( f + \frac{x^2}{f} \right) \omegaY - \omegaX \frac{df}{bs} - \frac{dy}{bs} \\ \frac{dz}{bs} \end{bmatrix} \\
\nu(\vec{\Theta}, \vec{\Omega}) &= \begin{bmatrix} x \overline{Y} \omegaY - \omegaX \frac{df}{bs} + \frac{dx}{bs} \\ \left( f + \frac{x^2}{f} \right) \omegaX + \omegaZ \frac{df}{bs} - \frac{dy}{bs} \\ \frac{dz}{bs} \end{bmatrix} \\
\xi(\vec{\Theta}, \vec{\Omega}) &= \begin{bmatrix} \frac{dx}{bs} \\ \frac{dy}{bs} \end{bmatrix}
\end{align*}
\]

(1)

The image point \( m' \) is then defined as:

\[
\begin{align*}
m' &= \begin{bmatrix} x_m' = x_m + \mu(\vec{\Theta}, \vec{\Omega}) \\ y_m' = x_m + \nu(\vec{\Theta}, \vec{\Omega}) \\ d_m' = d_m + \xi(\vec{\Theta}, \vec{\Omega}) \end{bmatrix} = P(\vec{\Theta}, \vec{\Omega})(m)
\end{align*}
\]

(2)

This displacement field will be noted as:

\[
\begin{bmatrix} \mu \\ \nu \\ \xi \end{bmatrix} = \Pi(\vec{\Omega}, \vec{\Theta}) \begin{bmatrix} x \\ y \\ d \end{bmatrix}
\]

2.3 Algorithm description

An overview of the following algorithm can be found in Fig. 2.

Two inputs are needed to efficiently and absolutely determine the ego-motion of the sensor: disparity information and motion information. To obtain the former, two methods will be investigated: block-matching and semi-global matching. To obtain the latter, a set of point-to-point correspondences
will be considered. Such correspondences can derive from an optical flow method, a KLT tracker [24] or some robust feature points matching [25]. These different methods will be evaluated in Sect. 2.4. This set of point-to-point correspondences will be noted:

$$S = \{m_i, i \in [1; N]\} \rightarrow \{m'_i, i \in [1; N]\}$$

From Eq. 1, a set of linear equations in \((\overrightarrow{T}, \overrightarrow{\Omega})\) can be derived. This system is over-constrained. There is no regular solution and the problem is equivalent to minimize the following energy:

$$\epsilon = \sum_S \text{dist}\left(m'_i, P(\hat{\overrightarrow{\Omega}}, \hat{\overrightarrow{T}})(m_i)\right)$$

where \text{dist} is a convex metric. Practically, a square metric is used and the problem then becomes a least squares solving, for which a classical singular value decomposition (SVD) approach can be used.

Outlier rejection Yet, as scenes can contain both static and dynamic objects, some correspondences can be outliers with respect to the global ego-motion. Point-to-point correspondences can also result in mismatches. Thus, an outlier rejection scheme is needed, to ensure robustness. The well-known RANSAC scheme [21] is chosen as it allows a great proportion of outliers (as long as a relative majority consists in inliers) and its ease of tuning. Therefore, RANSAC parameters are set assuming a minimum proportion of inliers of one-third and a probability of false rejection of 5% [26].

Time integration: filtering Despite the robustness ensured by the least square solving and the outlier rejection, it is still possible that our method yields to bad results. Those bad results can derive from several factors. First, a lack of useful feature points (i.e. images of static world points) can happen, especially in poorly textured environments, or in heavily cluttered environments. Moreover, hardware failure, such as missed frames, can obviously damage the results.

To cope with these issues, a temporal filtering, implemented as an extended Kalman filter (EKF), is to be used. The estimation of the covariance matrices is part of the estimation process that will be described later on. However, the EKF estimate is not to be used routinely. Instead, the final estimate of the retroprojection error \(\epsilon\) is used as an indicator. If this final error is superior to 5 pixels, the Kalman prediction is used instead of the measurement.

2.4 Evaluation

To evaluate the performances of the visual odometry algorithm described in 2.3, the Sivic simulator [27] has been used. It allows the precise simulation of vision sensors and car dynamics and provides an absolute ground truth for every relevant variable.

This section will be organized as follows. First, the relative performances of the visual odometry algorithm, depending on the chosen stereo algorithm, block-matching stereo algorithm (BM) or semi-global, sub-pixellic algorithm (SG), will be measured and compared to performances obtained when using an absolute depth map retrieved from the simulator. This will be done using robust SURF points, as a priori those are the most precise and robust ones. Then, the relative performance of SURF points\(^1\) and KLT tracker will be measured. This will be done, using the ground truth disparity maps from the simulator’s Z-buffer. Those tests will be conducted on two-frame data, without any time integration, but with different motion values. Full path reconstructions will also be provided, but only as more tangible results and will not be used later on.

2.4.1 Stereo algorithm impact

To evaluate the impact of the stereo algorithm, the following procedure is followed. A sequence of 600 stereo-pairs is considered. It is a pseudo-realistic sequence, in an urban environment, with moderate traffic and presenting several problematic elements (roundabout, highly repetitive patterns) (Fig. 3).

The sequence presents a wide variety of possible motions. For each pair, a disparity map is computed with a BM method, an SG method or directly from the Z-buffer of the simulator (see Figs. 4, 5 and 6 for examples). For each consecutive pair of disparity maps, a set of SURF feature points are extracted and matched. The motion of the sensor between the two points in time is estimated with a RANSAC process. This last step is re-iterated 10,000 times to estimate the standard deviation of the process. Indeed, both SURF

---

\(^1\) Different descriptor sizes were investigated in the case of SURF feature points, but little to no difference was perceived, hence these different versions of SURF will not be discriminated.
extraction and disparity map computation are deterministic, whereas RANSAC process is inherently random. Standard deviation of the estimated motion is given by the well-known estimator.

Bias will not be studied. Indeed, the two tested stereo algorithms being unbiased, the odometry results do not present a bias due to the stereo algorithm.

The overall results show a sensible impact of the used algorithm. If standard error in the case of a perfect disparity is considered to be unitary, results with a BM disparity map present a standard error of 1.5 and results with an SG disparity map present a standard error of 1.1. This difference is due, for a major part, to the difference in density between the two disparity maps, rather than to the sub-pixellic precision of the SG method. Indeed, additional tests have been conducted to evaluate the impact of both the density and sub-pixellic estimation. When an SG disparity map, filtered to fit the density of a BM one, is used, results are slightly better than the one obtained with the BM disparity map and present a normalized standard error of 1.4. On the other hand, a full SG disparity map, computed with a pixellic accuracy leads to a precision comparable to the one attained with a sub-pixellic SG disparity map.

The density of the disparity directly affects the detection capabilities of the system. As it will be shown later, a detection can only be achieved for points that present a disparity value. A denser disparity map leads to better chances of detecting an object (especially if the object presents a low contrast). For this reason, the SG method will be used.

However, as of now, real-time SG methods rely on expensive or highly power-consuming hardware and, for that reason, are not well fitted for embedded applications. Results obtained with the two methods will be compared in Sect. 5.
2.4.2 Correspondences impact

The next factor to investigate is the relative performance of the visual odometry algorithm, with respect to the feature points used. The same evaluation procedure as in 2.4.1 will be used. However, the results will be interpreted in terms of repeatability and in term of bias.

Dispersion and repeatability The first result that can be derived from these tests is that, if the standard deviation of SURF-based algorithm is unitary, the standard deviation of a KLT-based algorithm is around 0.5. This is due to the fact that SURF feature extraction produces much less points than a KLT tracker. Besides, the matching procedure in use (a correlation-based two-way winner-takes-all approach) reduces again the number of useful features.

This hypothesis has been verified by considering only the first points extracted by the KLT tracker, where \( n \) is the number of SURF points extracted for each image pair. In such conditions, both KLT Tracker and SURF feature points lead to the same results.

Bias and measurement precision This is perhaps the most important drawback of KLT or optical flow based technics. It performs as well and as accurately as SURF points under small motion assumption. However, if this assumption is not valid, the motion estimation will be highly biased.

In standard conditions (normal speeds, 30 Hz acquisition) this is of little concern. However, robustness to large motions can be an important asset. Indeed, depending on the objects of interest, 30 Hz might not be the best frame rate in terms of sensitivity. This will be dealt in details in Sect. 4.4.

2.4.3 Overall evaluation

Finally, an overall evaluation allows to evaluate, for every motion component, the maximum possible relative error of the estimation. For instance, on the test sequence presented as Fig. 7, the average positioning error after 600 frames was around 2%, and the maximum relative error was around 10%.

Figure 7 also shows the loss of precision of the KLT tracker as the acquisition frequency lowers. As this frequency lowers, the drift due to the fact that KLT tracker can cope with great displacement increases. On the other hand, SURF points are remarkably robust.

Given those results, the visual odometry system will consist of SURF points and an SG disparity map. Indeed, the versatility of SURF points with respect to the acquisition frequency might be needed (this point will be explained in 4.4). On the other hand, block-matching stereo algorithms are faster than semi-global ones, but the loss of density can be a substantial issue.

3 Detecting dynamic points

The following section will describe how dynamic points can be detected and discriminated in the sources images, given
the estimate of the ego-motion of the sensor, see Fig. 8 for an overview of this system. The various causes for mis-detection (false positives and false negatives), along with the precision and resolving power of the system, will be dealt with in Sect. 4.

3.1 Compensating the ego-motion

Given the estimate of the ego-motion \((\hat{\Omega}, \hat{T})\), one can compute, using Eq. 1, the 3D image-based displacement field due to this ego-motion. The first image, \(I^0_R\), can then be warped toward the second one, \(I^1_R\). This warping does not only concern image data, but also disparity maps.

Three issues can occur. First, when computing coordinates, rounding errors can deteriorate the results. Those errors are dealt with by using a bi-linear interpolation. The sole purpose of this interpolation is to avoid the creation of artificial edges in the image, so that a very good precision is not necessary.

Then, it is possible that, given two distinct points, their apparent motions lead to an occlusion. In that case, disparity is used to take into account only the closest of the two points.

Finally, given the fact that image-based displacement is computed from the disparity of a given point, the density of the warped image is the same as the density of the disparity map. Since dense disparity maps are used, this is not a major issue. However, some empty region can still occur, such as occultation regions. Those empty regions can lead to numerical issues; for that reason, they are filled with data from \(I^1_R\).

The final warped image is noted \(I^{0'}_R\).

3.2 Detecting the outliers

At this point, dynamic points can be considered to be the points that had been “moving” between \(I^{0'}_R\) and \(I^1_R\). Several methods can be considered to detect those outlying points. First, an absolute difference image can be computed but, albeit faster, this method should be avoided because it is very sensitive to illumination changes, it performs poorly in underexposed conditions (e.g. night, fog) and it implicitly assumes that the extracted ego-motion allows a sub-pixellic accurate warping. Hence, to ensure robust results, a more sophisticated approach is to be used.

The chosen solution is to compute a correspondence field. Once again, several methods can be envisaged. First, as shown in [17], a block-matching approach can be used. Several metrics can be used to perform such a task: zero-mean sum of absolute differences (ZSAD), zero-mean sum of squared differences (ZSSD) or the popular census-transform based metric [28].

Since the object is to identify a field of point-to-point correspondences, optical flow technics can also be used. A recent study on efficient and accurate optical flow technics has been focused on global methods, derived from [29], for instance [30].

Due to the recent and promising work in this field, this last solution will be preferred and a simple implementation of the Horn and Schunk method will be used. However, those present some limitations. Firstly, they are unable to handle large image displacements without the use of a multi-resolution scheme. Secondly, sudden variations of the illumination violate the constant brightness hypothesis and thus can perturb the results.

Ultimately, for every point \(m\) in \(I^{0'}_R\), for which a disparity value is available, it is possible to define a vector:

\[
\vec{A}(m) = \begin{bmatrix} \mu_A \\ \nu_A \\ \xi_A \end{bmatrix} = \begin{bmatrix} x_m - x_{m'} \\ y_m - y_{m'} \\ d_m - d_{m'} \end{bmatrix}
\]

where \(m'\) is the correspondent of \(m\) in \(I^1_R\), found through a block matching, or by an optical flow technique. Dynamic points can then be detected. They are the ones that verify:

\[
\| \vec{A}(m) \| > Th
\]

where \(Th\) is a given threshold. If the warping can be considered to be perfect, this threshold would be 0. However, as it will be shown in Sect. 4, this threshold must be carefully chosen.

3.3 Constraining the search space

Whatever the chosen method, given the fact that there is no global motion between the two images, the search for the correspondent of a particular point can be constrained, based on the expected dynamics of the potential objects. To that end, the 3D interval \(R^m\) is defined as:
\( \mathcal{R}^m = [x - f(x, y, d); x + f(x, y, d)] \times [y - g(x, y, d); y + g(x, y, d)] \times [d - h(x, y, d); d + g(x, y, d)] \tag{5} \)

where:

\[
\begin{align*}
    f(x, y, d) &= \max \left\{ \mu(\overrightarrow{\xi}, \overrightarrow{\Omega}) \mid \left( \overrightarrow{\xi}, \overrightarrow{\Omega} \right) \in \mathcal{M} \right\} \\
    g(x, y, d) &= \max \left\{ v(\overrightarrow{\xi}, \overrightarrow{\Omega}) \mid \left( \overrightarrow{\xi}, \overrightarrow{\Omega} \right) \in \mathcal{M} \right\} \\
    h(x, y, d) &= \max \left\{ \xi(\overrightarrow{\xi}, \overrightarrow{\Omega}) \mid \left( \overrightarrow{\xi}, \overrightarrow{\Omega} \right) \in \mathcal{M} \right\}
\end{align*}
\]

where \( \mathcal{M} \) is the space of all possible motions. Those possible motions are determined according to the maximum allowed speed. In the presented case, test images where taken in urban areas where speed is limited to 50 km h\(^{-1}\).

For every image point, a search space can be computed, independently of ego-motion results or scene structure. Thus, this can be done only once, making this process computationally efficient, while allowing the most economic search for every possible point.

### 3.4 Time integration

Integration of detection results over time can improve detection in several ways. First, it can reduce false negatives.

Indeed, a slow-moving object, or an object that falls in the cases studied in 4.2, can be detected intermittently. Thus, a measurement integration can enhance the confidence of such a detection.

It can also reduce the number of false positives. Two major causes can produce false positives. Repetitive patterns or (to a lesser extent) noise can be a source of wrong matches in the detection procedure, and thus false positives. Excessive error in the ego-motion extraction can also be a source of false positives.

On the other hand, it should be noted that, although allowing for more robust detection, time integration also delays an actual detection. This is of little concern for slow-moving objects. By definition, an early detection is not necessary in those cases. However, an early detection is indispensable for rapid objects. This particular point should be taken into account when designing the integration system.

Two ways of performing such an integration can be envisaged. A target-based tracking can be performed. It presents the advantage of reducing the number of objects to track, while achieving a higher level detection and representation. However, it relies on, possibly error prone, additional target extraction algorithm and tracking methods. Furthermore, issues such as occultation or partial detection can arise and interfere with a tracking algorithm. Especially, partial detection is particularly common, due to the density of the disparity map, and thus the density of valid points in the images.

On the other hand, pixel-based integration offers much more flexibility. The used method is described hereafter.

First, the following symmetrical logistic function is defined:

\[
f_{T(h, \lambda)}(x) = \frac{1}{1 + e^{-\lambda(x|T_h)}} \tag{7}\]

where the threshold \( T_h \) will be defined in Sect. 4.1, and the growth rate \( \lambda \) will be used to tune the system, to get fast detection for fast objects and yet robust detection for slow objects.

An instantaneous confidence image CI is then defined as:

\[
ICI(m) = \max \left\{ f_{T(h, \lambda)}(i_A(m)) \mid i \in \{ \mu, v, \xi \} \right\} \tag{8}\]

An integral confidence image CI can then be defined as:

\[
CI(m) = \alpha \cdot ICI(m) + (1 - \alpha)CI(m) \tag{9}\]

The \( \alpha \) factor directly impacts the delay introduced by the time integration. The chosen value is 0.3. In that way, any decision will be based upon the three latest images.

At every iteration, the image CI undergoes two registrations. The first one is the warping due to the ego-motion. The second one is the registration due to the residual displacement field \( \overrightarrow{A} \). The first registration ensures that the current confidence image will be consistent with the current pose of the sensor. The second one ensures that moving objects will not leave "trails" and that their associated confidence will cumulate over time.

At every step, this image is used to filter the output of the detection process:

\[
\overrightarrow{A}(m) = \begin{cases} 
\overrightarrow{A}(m) & \text{if } CI(m) \geq 0.5 \\
0 & \text{or } CI(m) < 0.5
\end{cases} \tag{10}\]

That way, false positives that lead to a punctual confidence value of 1.0 will be discarded. On the other hand, a mobile object will be detected after two frames. Detection of slow-paced objects will also be improved.

However, some false positives can still occur. Particularly when they originate from a repetitive pattern that covers itself across several frames.

### 4 Accuracy evaluation

In this section, the accuracy of the different stages of the algorithm will be evaluated. First, the warping will be considered. Then the detection will be investigated in terms of detection sensitivity and resolving power. Finally, the influence of the time resolution will be described.
4.1 Warping accuracy

In the previous section, it has been implicitly assumed that the warping is absolutely accurate. It is not true and inaccuracy should be taken into account. Indeed, the visual odometry algorithm is not perfect and its performances directly impact on the image warping. This section will describe the influence of visual odometry and warping on the detection capacities of the system. Firstly through the propagation of uncertainty in Eq. 1, the uncertainty of the displacement field can be estimated:

\[
\begin{align*}
\partial \mu &= \frac{\partial}{\partial \omega_X} \partial \omega_X + (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \partial \omega_Y + \frac{\partial}{\partial \omega_Z} \partial \omega_Z - \frac{\partial}{\partial T_X} \partial T_X + \frac{\partial}{\partial T_Y} \partial T_Y + \frac{\partial}{\partial T_Z} \partial T_Z \\
\partial \nu &= (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \partial \omega_X - \partial \omega_Y - \frac{\partial}{\partial T_X} \partial T_X + \frac{\partial}{\partial T_Y} \partial T_Y + \frac{\partial}{\partial T_Z} \partial T_Z \\
\partial \xi &= \frac{\partial}{\partial \omega_X} (\frac{\partial \omega_X}{\partial \omega_Y} - \frac{\partial \omega_X}{\partial \omega_Z} + \frac{1}{T_Z} \partial T_Z)
\end{align*}
\]

(11)

where \( a = \omega_X - x \omega_Y + \frac{T_Z}{b_2} \).

Partial derivative in Eq. 11 can be replaced either by the upper bounds of the estimates, as found in Sect. 2.4, or by a 3\(\sigma \) value, since the perturbation can be estimated to be Gaussian [18]. Assuming the latter:

\[
\begin{align*}
\Delta \mu &= \frac{\partial}{\partial \omega_X} \Delta \omega_X + (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \Delta \omega_Y + \frac{\partial}{\partial \omega_Z} \Delta \omega_Z - \frac{\partial}{\partial T_X} \Delta T_X + \frac{\partial}{\partial T_Y} \Delta T_Y + \frac{\partial}{\partial T_Z} \Delta T_Z \\
\Delta \nu &= (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \Delta \omega_X - \Delta \omega_Y - \frac{\partial}{\partial T_X} \Delta T_X + \frac{\partial}{\partial T_Y} \Delta T_Y + \frac{\partial}{\partial T_Z} \Delta T_Z \\
\Delta \xi &= \frac{\partial}{\partial \omega_X} (\frac{\partial \omega_X}{\partial \omega_Y} - \frac{\partial \omega_X}{\partial \omega_Z} + \frac{1}{T_Z} \Delta T_Z)
\end{align*}
\]

(12)

Therefore, it is possible to define a 3D interval of \( I^m_R \) for every point \( m \) in \( I^m_R \) by:

\[
\mathcal{V}^m_{\Delta \mu,s}(\hat{\omega}, \hat{T}) = \{x + \mu - \Delta \mu; x + \mu + \Delta \mu\} \\
\times \{y + \nu - \Delta \nu; y + \nu + \Delta \nu\} \\
\times \{d + \xi - \Delta \xi; d + \xi + \Delta \xi\}
\]

(13)

If \( m \) is the image of a static world point, there is a 99% probability that its correspondent in \( I^m_R \) lies within the boundaries of \( \mathcal{V}^m_{\Delta \mu,s}(\hat{\omega}, \hat{T}) \). The detection condition exposed in Eq. 4 can now be expressed as:

\[
| \mu_A(m) | > \Delta \mu
\]

or

\[
| \nu_A(m) | > \Delta \nu
\]

or

\[
| \xi_A(m) | > \Delta \xi
\]

(14)

It also implies that if \( m \) is the image of a dynamic world point, but its correspondent lies within \( \mathcal{V}^m_{\Delta \mu,s}(\hat{\omega}, \hat{T}) \), it cannot be detected as dynamic.

It also allows the determination of the threshold \( Th \) in Eq. 7:

\[
Th = \Delta \nu \forall i \in \{\mu, \nu, \xi\}
\]

(15)

4.2 Detection sensitivity

In the previous section, a limitation due to the accuracy of the ego-motion extraction has been unveiled. This limitation leads to the specification of an image-based threshold below which an image-based displacement cannot be considered to be due to independent motion.

The purpose of this section is to evaluate how this limitation impacts on the ability to detect a particular object. This will be done by determining under what circumstances (position, motion) a dynamic world point can not be detected as dynamic.

Let \( M(t_0) = \begin{pmatrix} X_M(t_0) \\ Y_M(t_0) \\ Z_M(t_0) \end{pmatrix} \) be a dynamic world point. Its image counterpart is \( m = \begin{pmatrix} x_m \\ y_m \\ d_m \end{pmatrix} \) according to the projection equations. Its motion is considered to be:

\[
\overrightarrow{T} = \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix}
\]

As the object is considered punctual, it is not relevant to consider a rotational component.

Its motion, relative to the sensor, is then:

\[
M(t_1) = \begin{pmatrix} X_M(t_1) \\ Y_M(t_1) \\ Z_M(t_1) \end{pmatrix} = \begin{pmatrix} X_M(t_0) \\ Y_M(t_0) \\ Z_M(t_0) \end{pmatrix} + \overrightarrow{T} + \overrightarrow{T} + \overrightarrow{T}
\]

Hence, its image displacement will be

\[
\begin{align*}
\mu_m &= \frac{\partial}{\partial \omega_X} \omega_X - (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \omega_Y + \frac{\partial}{\partial \omega_Z} \omega_Z - \frac{\partial}{\partial T_X} T_X + \frac{\partial}{\partial T_Y} T_Y + \frac{\partial}{\partial T_Z} T_Z \\
\nu_m &= (\frac{\partial^2}{\partial \omega_Y} + \frac{\partial^2}{\partial \omega_Z}) \omega_X - \omega_Y - \frac{\partial}{\partial T_X} T_X + \frac{\partial}{\partial T_Y} T_Y + \frac{\partial}{\partial T_Z} T_Z \\
\xi_m &= \frac{\partial}{\partial \omega_X} (\frac{\partial \omega_X}{\partial \omega_Y} - \frac{\partial \omega_X}{\partial \omega_Z} + \frac{1}{T_Z} T_Z)
\end{align*}
\]

(17)

The \( M \) point should be mistakenly considered as static if it satisfies the following conditions:

\[
\begin{align*}
| \mu_m - \mu | &\leq \Delta \mu \\
| \nu_m - \nu | &\leq \Delta \nu \\
| \xi_m - \xi | &\leq \Delta \xi
\end{align*}
\]

(18)

where \( a = \omega_X - \omega_Y + \frac{T_Z}{b_2} + 1 \).
Equation 18 introduces a systems of three inequalities, for six unknowns: \(x, y, d, T^M_X, T^M_Y, T^M_Z\). Even if we consider that, most of the time, \(T^M_Y\) is highly constrained, and this systems is still underdetermined.

There are several immediate solutions to this system. First, if an object is dynamic, but its motion is inferior to the precision of the ego-motion extraction algorithm, it cannot be detected. An object located at the infinite \((d = 0)\) will also be undetectable, no matter what its motion is.

Apart from these, other solutions can exist. For instance, let us consider the case of a purely translational sensor motion, then Eq. 18 becomes:

\[
\begin{align*}
|\mu_m - \mu| & \leq \Delta \mu \\
|\nu_m - \nu| & \leq \Delta \nu \\
|\xi_m - \xi| & \leq \Delta \xi
\end{align*}
\]

\(T^M_Z\) that satisfies the following can be found:

\[
\left| \frac{dT^M_Z}{b_s(1 - \frac{T_Z}{b_s})(1 - \frac{T_Z}{b_s} - \frac{T^M_Z}{b_s})} \right| \leq \Delta \xi
\]

Those solutions are illustrated in Figs. 9 and 10.

In particular, Fig. 9 highlights the importance of the choice of the baseline. Indeed, when \(T_Z \approx b_s\) the possibilities for undetectable \(T^M_Z\) are negligible. The practical system used in this work presents a baseline of 60 cm, that is optimal for speeds around 35 km h\(^{-1}\) and an acquisition frequency of 15 Hz. Indeed, if a baseline optimal for a speed of 50 km h\(^{-1}\) were chosen, at low speeds, error would have been higher. Choosing a baseline optimal for speeds around 35 km h\(^{-1}\) ensures that, at every possible speed between 0 and 50 km h\(^{-1}\), the minimum detectable longitudinal displacement will be 0.03 m, that is a speed of 1.5 km h\(^{-1}\).

Given a particular \(T^M_Z\), solutions for

\[
\begin{align*}
|d(xT^M_Z - fT^M_X)| & \leq b_s \Delta \mu \\
|d(yT^M_Z - fT^M_Y)| & \leq b_s \Delta \nu
\end{align*}
\]

can be found. The existence of these solutions depend on two factors: the distance between the sensor and the object, \(Z \propto \frac{b_s}{T_Z}\), and the ratio between the total size of the active matrix of the sensor and the focal length. Very wide angle optics should then be used carefully as those can be prone to wider undetectable motion ranges.

The system used for experimentation does not present this particular problem, as its focal length is 8 mm and the sensor size is below 5 mm. Although the fact that a purely translational motion has been considered as an example is not a limitation, indeed, for realistic values of \(\omega_Y\) and \(\omega_X\):

\[
|x \omega_Y - y \omega_X| \ll 1 - \frac{T_Z}{b_s}
\]

The previous conclusions are not affected.

4.3 Resolving power

Different objects, with different motions, can happen to be very close to one another (e.g. a pedestrian crossing a street in front of a car). In such types of scenario, it is important to be able to distinguish between the two objects, as one can be an immediate danger.

In the following, it will be assumed that two world points, \(M\) and \(N\), with two distinct independent motions, \(\overrightarrow{T^M}\) and \(\overrightarrow{T^N}\), are correctly imaged. The ability to differentiate between the two objects directly depends on the difference
between their residual displacements:

\[ \vec{A}(m) - \vec{A}(n) \]  

In the following, the worst case will be assumed: that is, \( m \) and \( n \) will be supposed to be coincident for \( t = t_0 \). This worst case scenario is purely hypothetical, but the point of this section is to highlight the resolving power of a system based solely on independent motion analysis.

Since both points are affected the same way, visual odometry precision has no effect on the following equation for the resolving power.

Hence, by rewriting Eq. 21:

\[ d \begin{align*} f \frac{T_{m}^{Z} - T_{n}^{Z}}{b_{z}} \\ f \frac{T_{m}^{Y} - T_{n}^{Y}}{b_{y}} \\ f \frac{T_{m}^{X} - T_{n}^{X}}{b_{x}} \end{align*} \]

with \( a = b_{x}(x\omega_{Y} - y\omega_{X} + 1) - T_{Z} \).

It appears that the ability to separate two objects that differ only by their motion is directly linked to their distance from the observer: \( Z \propto \frac{1}{f} \).

The only factor usable to improve the resolving power of the system is the difference between the motions. This can be done through the acquisition frequency of the image sensors.

4.4 Time resolution influence

Framerate is rarely properly discussed. In fact, most authors settle for the fastest available, using it as a misleading performance measurement. Even though this is of little importance for static approaches, motion, and especially independent motion, can benefit from a fine analysis of the impact of the framerate.

The motion of the sensor between two points in time can be rewritten as:

\[ \begin{align*} \vec{T} &= \begin{bmatrix} T_{X} \\ T_{Y} \\ T_{Z} \end{bmatrix} \\ \frac{1}{f_{acq}} \vec{V} &= \begin{bmatrix} V_{X} \\ V_{Y} \\ V_{Z} \end{bmatrix} \\ \frac{1}{f_{acq}} \vec{\Omega} &= \begin{bmatrix} \omega_{X} \\ \omega_{Y} \\ \omega_{Z} \end{bmatrix} \\ \frac{1}{f_{acq}} \vec{R} &= \begin{bmatrix} R_{X} \\ R_{Y} \\ R_{Z} \end{bmatrix} \end{align*} \]  

where \( f_{acq} \) stands for acquisition frequency.

By replacing the previous equation accordingly in 22, it appears that the lower the framerate is, the greater the resolving power. However, lowering the framerate presents four drawbacks.

First, as the apparent motion increases, the search domain \( \mathcal{R}_{m}^{m}(\vec{\omega}, \vec{T}) \) also grows, thus requiring more computational power. Moreover, an increase in the size of the search space can also deny the use of optical flow methods.

Second, decreasing the acquisition frequency implicitly means increasing the detection time of the complete system.

Third, as shown in 2.4.2, some odometry methods can perform poorly at low acquisition frequencies.

Finally, the visual odometry algorithm can be disturbed. Indeed, at some point, the assumption that rotations are small enough to linearize trigonometric lines is no longer true.

In usual road conditions, the most important rotation is the one around the vertical axis. The two other rotations can exist, but the first one to violate the linearization assumption...
will be the yaw. Furthermore, given the second-order components of sin and cos, error relative to the approximation of sin can be omitted. Translational components will also be omitted in the expression of $\mu$ and $\nu$ for clarity purposes, and it can be clearly shown that those have no effect on the conclusion. When the yaw angle is not small enough to linearize trigonometric lines, Eq. 1 becomes:

$$\begin{align*}
\mu (\vec{1}) &= \frac{x y}{f_{\text{acq}}} R_X \cos \frac{R_Y}{f_{\text{acq}}} - \left( f + \frac{x^2}{f^2} \right) R_Y + \frac{y}{f_{\text{acq}}} R_Z \cos \frac{R_Y}{f_{\text{acq}}} \\
\nu (\vec{1}) &= \frac{R_X}{f_{\text{acq}}} \left( f + \frac{x^2}{f^2} \right) \cos \frac{R_Y}{f_{\text{acq}}} - \frac{x y}{f_{\text{acq}}} R_Y \cos \frac{R_Y}{f_{\text{acq}}} - \frac{x R_Z}{f_{\text{acq}}} \\
\xi (\vec{1}, \vec{T}) &= d \frac{y R_X \cos \frac{R_Y}{f_{\text{acq}}}}{x R_Y - y R_X \cos \frac{R_Y}{f_{\text{acq}}} + 1}
\end{align*}$$

(24)

The error committed when erroneously considering that all angles allow a linearization is then:

$$\begin{align*}
\text{err}_\mu &= (1 - \cos \frac{R_Y}{f_{\text{acq}}}) \times \left[ \frac{x y}{f_{\text{acq}}} R_X + \frac{y}{f_{\text{acq}}} R_Z \right] \\
\text{err}_\nu &= (1 - \cos \frac{R_Y}{f_{\text{acq}}}) \times \left[ \frac{R_X}{f_{\text{acq}}} \left( f + \frac{x^2}{f^2} \right) - \frac{x y}{f_{\text{acq}}} R_Y \right] \\
\text{err}_\xi &= d \frac{y R_X \cos \frac{R_Y}{f_{\text{acq}}}}{(1 - \cos \frac{R_Y}{f_{\text{acq}}}) (1 - \frac{y}{f_{\text{acq}}} + \frac{x R_Y}{f_{\text{acq}}} + \frac{x R_X}{f_{\text{acq}}} \cos \frac{R_Y}{f_{\text{acq}}})}
\end{align*}$$

(25)

This error is represented in Fig. 11.

All three components exhibit a dramatic error increase for acquisition speeds under 10 Hz. Ideally, an acquisition frequency of 11 Hz would lead to an acceptable error margin (below 1 pixel), while increasing the resolving power of the system. Moreover, it would still allow for a fast detection (under a tenth of a second). However, hardware used for the implementation did not allow such a frequency and should be defaulted to 15 Hz. At this speed, the impact of time resolution is negligible (under a half of a pixel for every component).

**Conclusion** Setting up the acquisition frequency of the system should result from a trade-off between resolving power on one hand, and visual odometry stability and detection time on the other hand. Acquisition frequency is too often considered to be a consequence of a particular algorithm. However, its impact on detection performances and on the ability to take proper actions (brake, avoid) are important and should not be neglected.

5 Results

Used stereo sequences are issued by the french LOVe (LogicIEL d’Observation des Vulnérables) project. Those sequences were taken in urban environment, with many different scenarios. In these presented results, the vector $\vec{A} (m)$ will be color-coded, with a saturation and value proportional to the norm of the vector, and a hue equal to its angle with respect to the horizontal. The $\xi$ component will not be represented for clarity purposes, and it should be noted that it is the smaller one in most cases.

5.1 Time integration influence

As it has been shown in 3.4, integration of confidence over time can serve two purpose. The first one is to improve the detection, especially when a target presents a small image motion. On the other hand, the elimination of false positives is also a goal of time integration.

---

Fig. 11 Error as a function of acquisition frequency, assuming a worst case scenario for the motion values and the coordinates

 Springer
Dynamic objects detection through visual odometry and stereo-vision

5.1.1 Detection improvement

The detection improvement is illustrated by Fig. 12. Without time integration, detection is lost in frames 3 and 4. Apparent motion is much smaller and is filtered out. On the other hand, with the time integration, the apparent motion in frame 3 is no longer filtered. Detection is based roughly on the current frame and the two last; thus, frame 4 presents only a partial detection, but a detection nonetheless.

Moreover, overall detection is more localized around the pedestrian. Typically, confidence values around the pedestrian are low enough to filter out those small false positives.

5.1.2 False positives elimination

Given the fact that the residual displacement field is computed with a traditional optical flow method, sudden variations in illumination can heavily disturb the results.

Figure 13 presents results for a particular frame in which a sudden illumination change occurs.

As expected, without any time integration, the used Horn and Schunk implementation shows a lot of mismatches and errors. However, with the time integration procedure, the results are almost perfect, with only the pedestrian in the foreground and in the background detected.

5.2 Time resolution influence

As explained in 4.4, lowering the acquisition frequency has two major effects: it will lead to a growth of the search interval \( \mathcal{R}_{\Omega} \), and it will enhance the resolving power. The former will not be taken into account, as this issue can be easily overcome through a multi-resolution approach, for example.

The resolving power improvement is illustrated by Fig. 14. The slow-moving pedestrian in the foreground is not well detected at 30 Hz; on the other hand, at 15 Hz, the apparent motion is increased and detection is much easier.

The pedestrian of Fig. 12 is also a good example of the interest of carefully choosing the acquisition frequency.
Fig. 14 Comparison of the results obtained at a 15 and a 30 Hz acquisition frequency. *Left* 15 Hz, *right* 30 Hz. Images are cropped and magnified round the area of interest.

Fig. 15 Output example. *Top* source image, *bottom* processed image. The two crossing pedestrians are merged together, but in the source image, they are poorly contrasted and merged as well. The global image presents no noise or false positives.

Indeed, with a frequency of 30 Hz, she would not have been detected, while a 15 Hz processing allows an early and reliable detection.

5.3 Miscellaneous results

Figures 15, 16, 17, 18, 19 and 20 present additional results. Those images illustrate a variety of scenarios. Different targets, with different independent motions, are correctly imaged. These detections are early enough to plan an emergency brake or a trajectory modification. The presented system achieved a detection rate of 98%, i.e. 2% of the objects of interest were not detected. Moreover, 90% of the detection were true positives. This quantitative evaluation was conducted on 5,000 stereo pairs. All observed false positives were due to repetitive patterns that disturbed the optical flow method used.

6 Conclusions and future works

6.1 Conclusions

The presented method allows accurate and early detection of mobile objects, from a mobile stereo-sensor. This detection is based upon a visual odometry algorithm and a residual...
Dynamic objects detection through visual odometry and stereo-vision

Fig. 16 Output example. The extracted car is initiating a turn on its left, and its shadow is extracted as well. The pedestrians on the left side of the image are standing still.

Fig. 17 Output example. The car in the far background is initiating a turn on its left, while the other vehicles are waiting. Regarding the pedestrian in the foreground, the apparent motion of her leg is different from the apparent motion of the rest of her body, which is coherent. One might note that the shadow of the pedestrian is detected as well.

displacement field measurement. Indeed, the vast majority of detectable mobile objects were detected within three frames, or 0.3 s. The causes for undetectability have been investigated and means to optimize the detectable motion range have been proposed.

Contrary to existing methods, several key factors are exploited to improve the overall performances. First, the visual odometry accuracy is evaluated. This evaluation allows a proper definition of search spaces for the estimation of the residual displacement field. The limitations of this detection system were investigated, and recommendations for possible future systems were derived from these conclusions. The importance of time resolution on detection performances, but also on visual odometry performances, has been studied to establish criteria allowing system designers to avoid some pitfalls.

Fig. 18 Output example from the same sequence as Figs. 13 and 12. The pedestrian in the foreground just turned her head toward the camera. The two groups of pedestrians in the background are moving toward the right side of the image.

This whole system relies solely on visual information, whereas others must have recourse to additional sensors, such as IMU. At this moment, implementations that do not rely on expensive or power-consuming hardware are possible and actually embeddable.

Detection can be achieved, even in hard cases (partial occultation, slow-paced objects), thanks to a simple, but efficient, time integration procedure.

Furthermore, the presented method can process 640×480 images in real time (15 Hz), on a Intel T9400-powered laptop, granted that disparity maps are available. However, the authors do not consider this as a limitation, as several methods are currently investigated to implement stereo algorithm on GPU [12].

6.2 Future works

Despite the good results this framework can achieve, several improvements and further developments can be envisaged. First, the detection of shadows as dynamic objects could be an issue. Such a detection could lead to a false estimation of the free space in front of the vehicle. This could be easily avoided by using V-disparity [31] to eliminate road points from the detection.

Total occultation of the objects of interest can also be an issue. As of now, the authors of this study do not handle this issue. However, a high-level tracking procedure, based on a model of the motion of the object (e.g. a KALMAN filtering), could be used to track the occulted object and improve its detection.

Then, it has been shown in 2.4.2 that, depending on the used feature points, different acquisition frequencies can be used. This has been used to maximize the resolving power of the dynamic object detection system. However, the
optimization was done assuming an urban environment and does not reinforce road safety in highways, where accidents can be much more severe. An improved system should be able to adapt its acquisition frequency and parameters to work in several different road conditions. For instance, in an urban area, obstacles are relatively slow, and so a low frequency should be necessary. On the other hand, on highways, objects are much faster, rotational motions are more constrained, and thus a high frequency and KLT-based feature points could be used instead. Such an adaptation can be realized using additional sensors (like a GPS coupled with an accurate map) or by decimating the input video feed to emulate different acquisition frequencies.

Finally, additional efforts should be put into computation and optimization to achieve more flexibility.

Acknowledgments The authors would like to thank the research cluster for science and technology Digiteo for its funding and its support to their project.

References


Author Biographies

Adrien Bak was born in France on July 25, 1985. He graduated from the Institut d’Optique Graduate School in 2007, where he received an engineer’s diploma of optics and photonics. He received a Ph.D. degree in Image Processing for his work at Université Paris XI and LIVIC. His work, funded by the cluster for scientific research “Digiteo” was focused on the development of Computer Vision systems for autonomous vehicles, using a cooperation between stereo-vision and motion analysis. During this period, his research interests included 3D reconstruction, visual odometry and motion detection. He now works as an Image Science Engineer, for the privately owned French company DxO Labs. He has authored and coauthored several scientific papers, and is credited for the redaction of a book chapter.

Samia Bouchafa was born in Algiers, Algeria, on July 16, 1971. She graduated from Computer Science National Institute (INI, Algiers) in 1993. She received the M.S. degree in Robotics from University Paris VI in 1994. She received, in 1998, the Ph.D. degree (highest honours) in electrical and computer engineering from University Paris VI and Institut National de Recherche sur les Transports et leur Sécurité (INRETS). Her thesis work was undertaken in the framework of the European project CROMATICA (DG XIII), dealing with crowd monitoring by image processing in the subway. Since 1999, she has been an associate professor at University Paris XI. Her research interests include robust primitive extraction and decision strategy for image matching applied to motion analysis, registration or stereovision. Currently she focuses on cumulative and voting decision strategies for motion or stereo-based applications in the context of automatic driver assistance systems. She has authored more than 30 publications.

Didier Aubert received the M.S. and Ph.D. degrees, respectively, in 1985 and 1989 from the National Polytechnical Institute of Grenoble (INPG). From 1989 to 1990, he worked as a research scientist on the development of an automatic road following system for the NAVLAB at Carnegie Mellon University. From 1990 to 1994, he worked in the research department of a private company (ITMI). During this period he was the project leader of several projects dealing with computer vision. In 1995 he joined INRETS (French National Institute for Transportation and Safety Research). From 2002 to 2009 he was the manager of the LIVIC perception team. He is currently a senior researcher and director of LEPIS. He works on both automated highway and driving assistance systems. He taught and teaches in several universities (Paris VI, Paris XI, Paris XII, Evry, Versailles) and engineering schools (ENPC, ENST, ENSMP). He is the author and co-author of several scientific papers and has participated in the redaction of several books.