RAWSEEDS
Robotics Advancement through Web-publishing
of Sensorial and Elaborated Extensive Data Sets

Raw seeds can be consumed as they are...
or be the start for the growth of new results.

RAWSEEDS
WorkPackage 1

Deliverable D1.2
Roadmap of Outdoor Activity

date due: end of month 12 (October 31st, 2007)
author: Domenico G. Sorrenti (WP-1 Leader)
internal reviewer: C. Stachniss (ALU-FR), Neira J. and Tardós J. D. (UNIZAR)
external reviewer:

contributors:

Caccia Claudio, UNIMIB

Fontana Giulio, POLIMI

Matteucci Matteo, POLIMI

Sorrenti Domenico G, UNIMIB

Table of Contents

1. About this document

This document has the purpose of defining the framework for the technical work of the RAWSEEDS project, for what concerns the outdoor scenarios (please see the following section for a definition of *scenario*). It collects high-level information only, without going into the implementation details, unless required to clarify specific points.
The overall structure of this document mirrors that of Deliverable D1.1 (Roadmap of Indoor Activity). Moreover, the content of D1.2 will frequently make references to that of D1.1. This is due to the fact that both D1.1 and D1.2. are the result of the work of WorkPackage 1 (WP-1), and that the Partners made a deliberate effort to maintain - whenever this was valid - a similar hardware and software framework and a common methodological stance for indoor and outdoor operations. This design choice was aimed at insuring maximum usefulness and consistency for the results of the RAWSEEDS project.

1.1 RAWSEEDS terminology

Note: this subsection is a copy of the corresponding section of Deliverable D1.1. It has been reproduced here to be easily available as a reference.

The aim of the RAWSEEDS project is to produce and make easily available through the Internet a **benchmarking toolkit** (or simply **toolkit**) comprising all the elements needed to test algorithms designed for the problems of mapping, self-localization or SLAM (Simultaneous Localization And Mapping). RAWSEEDS is specifically oriented towards robotic systems, although its toolkit could be useful also in different contexts (e.g. surveillance); the use of the toolkit should greatly reduce the time and effort needed for the successful development of innovative algorithms or products, by eliminating the need to set up costly data acquisition campaigns.
The RAWSEEDS toolkit will be based on real-world data exclusively (i.e. there will not be simulated data), and will include instruments (called **Benchmark Problems** and **Benchmark Solutions**) to test, rate and compare different algorithms. Along with these, the toolkit will include readily usable examples of state-of-the-art algorithms, to be used as examples in the design of new ones.

The base of all the work of RAWSEEDS are the data sets, also called *datasets.* In the rest of this document, the term **dataset** will be used to identify the set of synchronized data streams obtained by recording the output of the sensors mounted on a robot when the robot explores an environment. A single instance of this exploration procedure will be called a (data-gathering) **session**. A session can be actually performed by splitting it into the exploration of different (but strictly related, e.g. adjacent in space or time) environments, thus generating multiple datasets; in this case the single explorations will be called **subsessions**.
Alternative but "real" datasets (i.e. comprising actual sensor measurements,

not calculated or simulated data) can be obtained from a given one by discarding part of the data: for example by omitting the data generated by one or more sensors or by performing undersampling of the data. This can be useful to test the performance of algorithms which use different sensor sets or to simulate the effect of sensors with lower performance than the ones actually used. Such datasets will be called **derived datasets**.
The complete set of conditions defining a single data-gathering session will be called **scenario**. A scenario, then, will be defined by information such as: hardware setup, physical location of the experiment, presence or absence of people, lighting conditions, kind of terrain, and so on. Please note that for the same location and hardware setup, different scenarios can be defined.

Project RAWSEEDS will gather two types of datasets: *indoor* datasets and *outdoor* datasets. The former have the objective of covering the typical environments encountered by robots operating in locations where surrounding walls and roof are present: e.g. homes, industrial plants, offices, warehouses. In this kind of environment artificial lighting is usually present, possibly along with sunlight entering through windows or other openings, and the terrain is generally smooth (though not necessarily flat everywhere: ramps or stairs are common). Currently (as in the past history of robotics) most research or commercial robots are designed to operate in indoor environments. Therefore indoor datasets are the most important and used; on the other hand this means that several indoor datasets (albeit usually with a much lower quality compared to the ones that RAWSEEDS will make available) are already available to the community. On the contrary, the second type of datasets (i.e. outdoor datasets) is extremely rare to find: partly because outdoor robotic applications are still rare, and partly because setting up a session of outdoor data-gathering with mobile robots is time-consuming, difficult and costly. Thus the datasets provided by RAWSEEDS will address a serious stumbling block to the development of outdoor robotic applications.

It is important to note, at this point, that the data upon which the RAWSEEDS toolkit will be based will all be verified and *validated*, i.e. their quality and correspondence to requirements will be explicitly certified by the RAWSEEDS Consortium with reference to specific, published standards. Moreover, along with each of the datasets RAWSEEDS will provide the associated **ground truth**. This is a set of information describing in the most accurate way the real environments explored by the robots and the trajectory followed by the robots through them. Ground truth is used as a reference against which the results obtained by applying algorithms (e.g. for mapping) to the datasets can be evaluated. None of the real-world datasets currently available to the robotics community has been validated in the way described above, nor provides a ground truth.

This document describes the activities related to the outdoor datasets only: both for the generation of the datasets (which requires specific hardware and software architectures) and for the generation of the parts of the RAWSEEDS toolkit which are based on those datasets.

2. Project overview

The work of the RAWSEEDS project can be split into different *aspects*, each of which requires a specific design phase preceding the implementation. In the context of this section, the word "aspect" is used as a generalization of the concept of "task".

Below is a table of all the aspects of RAWSEEDS' work concerning the outdoor activity. To each aspect we associated a brief note, describing its advancement status. The status ranges from "open", for aspects where everything except a basic description is absent, to "closed", for aspects where every detail has been settled. Of course even "closed" aspects could be re-evaluated and possibly modified if such a need emerges, during the subsequent activities of RAWSEEDS.

Please note that the following table includes two different categories of aspects: those that are an integral part of the work of WP-1 and those that lie *outside* of WP-1. The latter are aspects which pertain to WorkPackages that at the moment are not completed or not yet started, but whose activities have to be defined and planned during WP-1. As a consequence, the meaning of the status column is different for the two categories: for the first category, it reflects the actual state of accomplishment of the set of tasks concerning an aspect; for the second one, it describes the advancement in the planning for that aspect.

In the following sections of this document each aspect outlined in the table will be described in detail. Please note that each element of the following table is associated to an item in the Table of Contents of the document, to facilitate consultation.

Aspects of the outdoor activity of RAWSEEDS	**Advancement status**
Hardware and software setup	
robot platform(*)	closed
sensor systems	closed
complete data-acquisition robot(*)	closed
outdoor scenarios	
locations	closed
scenarios	closed
data-acquisition methods	closed
data-gathering sessions	mostly open
Data validation	

evaluation criteria	closed
acceptability thresholds	mostly open
evaluation instruments	closed
Ground truth	
ground truth for localization	closed
ground truth for mapping	closed
Benchmark Problems	
problems	closed
data representation and file formats	ongoing work
evaluation methodologies for the solutions	closed
Benchmark Solutions	
solution algorithms	mostly closed
web-publishing policy	closed
Documentation and manuals	almost closed

(*) In the following sections it will be explained that a second platform, and a second data-acquisition robot based on that platform, are currently in preparation to generate additional datasets.

3. Hardware and software setup

3.1 Robot platforms

In comparison to indoor activities, outdoor operation puts harsher constraints on the choice of the robot platform. Outdoor environments are extremely variable in their characteristics and requirements, and it is impossible to cover such a wide spectrum with a single robotic platform. Moreover, most rural or wild outdoor environments require the use of special propulsion systems (e.g. track-based ones); systems that are, unfortunately, both very costly and unsuitable for solid ground.
However, most outdoor robotic applications that can be envisaged at present are aimed at urban (or *urban-like*, e.g. roads, malls, parking lots) settings. Moreover, the design and construction of a robot platform able to operate on uneven and/or unpaved ground is out of the scope of RAWSEEDS. For these reasons, RAWSEEDS will restrict its operations to *urban* outdoor environments.

The approach of RAWSEEDS towards the gathering of data from urban environments is twofold.
First of all, we will use the same platform (i.e. Robocom) used for indoor activities also for the exploration of outdoor environments. Robocom is, in fact, fully capable of urban outdoor use: its only limitation in this respect is the maximum speed it can reach, which is important in long-range operations. Robocom has the advantage of being sufficiently small to navigate through spaces designed for human use (e.g. doorways, corridors, halls), and it therefore makes the use of *mixed* scenarios (partly indoor, partly outdoor) feasible. The use of Robocom for outdoor data gathering will lead to outdoor datasets which are absolutely consistent with the indoor ones, as the hardware platform and most of the sensor set will be exactly the same. This, along with the availability of *mixed* datasets, is important because it allows the use of the same software algorithms for indoor and outdoor data, thus highlighting any limitation or shortcoming that an algorithm can show when confronted with real-world (i.e. not simulated) outdoor data. As algorithms for robotics are typically developed and tested in indoor environments, the availability of such datasets should give a significant impulse to the development of algorithms yielding good performance both indoor and outdoor.
In addition to Robocom, RAWSEEDS plans to set up a second, very different, platform for outdoor data gathering. We think that the application of robotics to automotive products will experience a fast growth in the next years, therefore we will do our best to provide the scientific and industrial actors (present and prospective) in this field with high-quality data that can be used to test their algorithms. To this aim RAWSEEDS has acquired a car-like (i.e. four wheels, Ackerman-type steering) vehicle, and is augmenting it with suitable odometry systems, in order to use it as a robot-like sensorized mobile platform. The datasets gathered by this vehicle will be very close to those that

a car (equipped with sensors) would produce. The possibility of using a car (or a *cybercar*, such as those from Robosoft) as a sensor platform has been evaluated and ruled out for its higher cost, unfortunately not compatible with the budget of RAWSEEDS.

The following sections will give a detailed description of the chosen outdoor data-gathering platforms. They will mostly focus on the four-wheel platform, as Robocom has already been described in Deliverable D1.1.

3.1.1 Robocom platform

The first outdoor platform that RAWSEEDS will use is the mobile robot *Robocom*, shown in Figure 1 and already presented in D1.1: for this reason it will not be described in depth here.

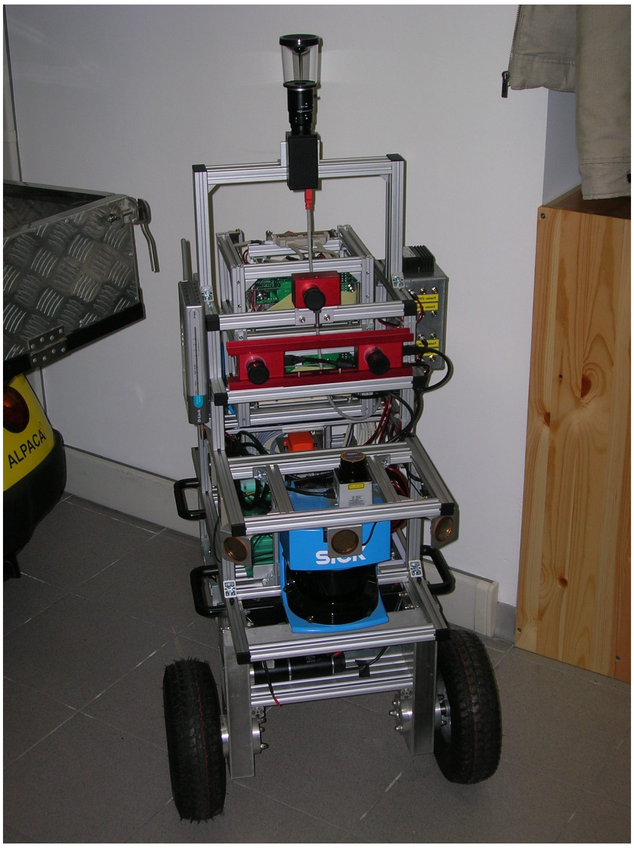

Figure 1: Robocom robot (equipped with RAWSEEDS' sensor frame*)*.

In this context it is sufficient to repeat that it is a differential drive autonomous robot, governed by an Apple *Mac mini* computer with the Linux operating system. For RAWSEEDS it will be guided manually by a human operator, as autonomy is neither required nor useful. Robocom is capable of limited outdoor operation, i.e. it can overcome modest slopes and slightly rough terrain; its small footprint and high payload/volume ratio are useful when movement through narrow passages is needed. For these reasons it will be also used for *mixed* scenarios, i.e. scenarios where indoor and outdoor environments are both present.

3.1.2 Alpaca platform

In additon to Robocom, RAWSEEDS is setting up a second platform. It is not a robot, but an electric four-wheel vehicle for the transportation of people: specifically, it is a *golf cart*. The chosen model of cart, shown in Figure 2, is called Alpaca BE448 and is manufactured by an Italian company called Ecology Runner.

Figure 2: Alpaca BE448 golf cart.

The Alpaca BE448 is driven by a 4kW electric motor, powered by a 48V DC battery pack composed of four automotive lead-acid batteries. It can overcome slopes up to 28% and has a maximum speed of 33km/h, with an autonomy of 80km (full payload). Minimum steering radius is a conveniently small (for RAWSEEDS) 3m. Overall dimensions of the vehicle are 2700mm (L) x 1180mm (W) x 1750mm (H): comparable with those of a small car. Width is limited, though, helping with narrow passages. The steering system is of Ackerman type, i.e. of the type used by cars. This is one of the main reasons for the choice of the Alpaca for RAWSEEDS: in fact, as previously stated, the data gathered by the Alpaca will be a realistic example of the kind of data that a (suitably equipped) car would generate.
A key difference between the Alpaca and a car lies in the fact that the Alpaca is not authorized to circulate on city roads (the authorization procedure is still under way). For RAWSEEDS this is not an issue, because (as will be explained in the following sections) data acquisition will be performed on roads that are under the jurisdiction of Politecnico di Milano.

RAWSEEDS' Alpaca has been modified to have a large platform instead of the

rear-facing seats shown in Figure 2. This platform is used for the installation of the data-processing hardware. In addition to that, work is ongoing by the Partners to modify the Alpaca in order to mount an odometry system on it; the data from that system will be, in fact, an important complement to the data produced by the sensors on the cart.
Please note that, for the purpose of gathering datasets for RAWSEEDS, the golf cart is completely equivalent to a robotic platform such as Robocom, given that in both cases the platform is driven by a human operator: the only difference being the fact that in the case of the golf cart the operator sits on the platform instead of using a remote controller.

3.1.3 Fitting of the odometry system to the Alpaca

The problem of fitting an odometry system to a vehicle not designed for it is not trivial. This section will describe how that problem has been studied, and what solution has been selected, in the case of the Alpaca. Currently the actual modification work on the Alpaca's parts is being performed.

Cart odometry sensor

The vehicle that will be used to make outdoor acquisitions needs to be equipped with suitable sensors providing useful information for dead-reckoning operations. The kinematic model describing the vehicle is the well-known Ackerman steering system, represented in Figure 3:

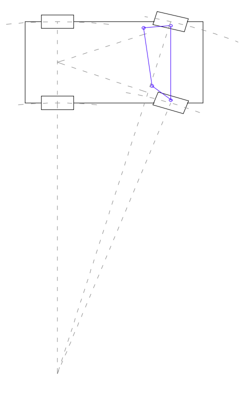

Figure 3: kinematic model.

The vehicle has two driving wheels (rear wheels) and two steering wheels (front wheels). Ackerman steering is designed to ensure that the inside front wheel is rotated to a slightly sharper angle than the outside wheel when turning, thereby eliminating geometrically induced tire slippage. As seen in Figure 3, the extended axes for the two front wheels intersect in a common point that lies on the extended axis of the rear axle. The locus of points traced along the ground by the center of each tire is thus a set of concentric arcs about the center point of rotation, and all instantaneous velocity vectors will subsequently be tangential to these arcs. Such a steering geometry (named “neutral Ackerman steering”) is said to satisfy the Ackerman equation:

$$\cot(\phi_i)-\cot(\phi_e)=\frac{d}{l} \tag{1}$$

where:

- ϕ_i steering angle of the internal wheel
- ϕ_e steering angle of the external wheel
- *d* lateral wheel separation
- *l* longitudinal wheel separation

For the sake of convenience, the vehicle steering angle ϕ_{SA} can be thought of as the angle (relative to vehicle heading) associated with an imaginary center wheel located between the two front wheels. ϕ_{SA} can be expressed in terms of either the inside or outside steering angles as follows:

$$\cot(\phi_{SA})=\frac{d}{2\cdot l}+\cot(\phi_i)=\cot(\phi_e)-\frac{d}{2\cdot l} \tag{2}$$

The Ackerman and differential steering systems is non-holonomic, meaning that it is impossible to specify one system parameter without affecting another. The kinematics of the system is represented by the following equations:

$$\begin{cases} \dot{x}=v\cdot\cos(\vartheta) \\ \dot{y}=v\cdot\sin(\vartheta) \\ \dot{\vartheta}=v\dfrac{\tan(\phi_{SA})}{l} \end{cases} \tag{3}$$

where *v* is the vehicle’s forward velocity (measured at the center axle of the rear wheels), ϕ_{SA} is the steering angle, the point (*x*,*y*) refers to the center of the rear axle, and ϑ is the vehicle’s orientation.

The pose of the vehicle is determined if two variables, for example v and $\dot{\vartheta}$, are known: so the minimum number of sensors needed for pose estimation is also two; any further data can be used for redundancy.
The easiest way to make dead-reckoning with two sensors considers the rear wheels as the ones to be equipped with encoders. In fact the two relations that link sensor readings and the pose are:

$$\begin{cases} \Delta dist = \dfrac{(enc_{rL} + enc_{rR})}{2} \\ \Delta \vartheta = \dfrac{(enc_{rL} - enc_{rR})}{d} \end{cases} \tag{4}$$

where:

- $\Delta dist$ distance between two readings
- $\Delta \vartheta$ rotation between two readings
- enc_{rL} distance read by the encoder on the rear left wheel between two readings
- enc_{rR} distance read by the encoder on the rear right wheel between two readings

Equation 4 is valid in conditions of no slippage.

Pose estimation

The operations needed to estimate the robot pose are the following: first the instantaneous speed and angular velocity are estimated:

$$\begin{cases} \hat{v}(t) = \left(\dfrac{\Delta tick_{rL}(t) + \Delta tick_{rL}(t)}{2 \cdot \Delta t} \right) \cdot \dfrac{1}{resol} \\ \hat{\vartheta}(t) = \left(\dfrac{\Delta tick_{rL}(t) - \Delta tick_{rL}(t)}{d \cdot \Delta t} \right) \cdot \dfrac{1}{resol} \end{cases} \tag{5}$$

A simple integration of the equations in 3 with values obtained in 5 allows to estimate the pose of the vehicle:

$$\begin{cases} x(t) = x(t-\Delta t) + \hat{v}(t) \cdot \cos(\vartheta(t-\Delta t)) \cdot \Delta t \\ y(t) = y(t-\Delta t) + \hat{v}(t) \cdot \sin(\vartheta(t-\Delta t)) \cdot \Delta t \\ \vartheta(t) = \vartheta(t-\Delta t) + \hat{\vartheta}(t-\Delta t) \cdot \Delta t \end{cases} \tag{6}$$

where:

- Δt is the sampling interval
- $\Delta tick$ are the encoder counts in Δt
- *resol* is the sensor resolution (counts per turn)
- x, y, ϑ is the pose of the vehicle

Equations 5 and 6 show that the only variables to be decided, given the vehicle's geometry, are the the sensor resolution and the sampling time.
To obtain a feasible and effective sensor, simulations have been made to analyze the way the sensor properties affect the pose estimation.

Simulations

The target consists in simulating the course of a vehicle with a neutral Ackerman steering system (the system described above) and analyzing the results obtained by odometry sensors at different conditions. The simulations allow to define the requirements of the sensors to be used.
Given the vehicle's geometry (real data of the vehicle):

- Wheel radius: 230 mm
- Lateral distance between wheels: 750 mm
- Longitudinal distance between wheels: 1800 mm

The tests have taken into examination the following parameters:

- vehicle's speed during the path (supposed t be constant through all the path)
  - $0.1 \div 5\, m/s$
- number of counts per turn (sensor resolution)
  - $18 \div 180$ counts per turn
- sampling frequency
  - $1 \div 100\, Hz$

Given all the parameters, the performance of the dead-reckoning operations depend only on the initial state of each sensor and on the quantization due to resolution. As the model considers no wheel slippage, the simulations results can be considered a lower bound of the error.

Initialization

The initial state of each sensor may greatly influence the final result of the estimation, because, in our case, the discretization is quite coarse[1]. The combination of initial state, resolution, speed and sampling frequency define the capability of the sensor to estimate the pose of the vehicle.
In our simulations each sensor has been initialized to a uniformly distributed random position between 0 and $\frac{C_{wheel}}{n_{divis}}$, where C_{wheel} is the wheel circumference and n_{divis} is the number of the sensor's counts per round.
As the relation between all the variables (initial state, speed, resolution, sampling frequency) is not deterministically known, different experiments have been carried out: for every combination of speed, resolution and sampling frequency, 1000 simulations with different initializations have been made and the results have been analyzed.
Also the path of the vehicle may influence the estimation: so, two different paths have been considered.

1 Please note that common resolutions for car ABS sensors are 48 counts per turn. As we will show, with our vehicle this figure is not sufficient for accurate odometry.

Type of paths

Two kinds of path have been used in the simulations: a straight route and the one represented in Figure 4.

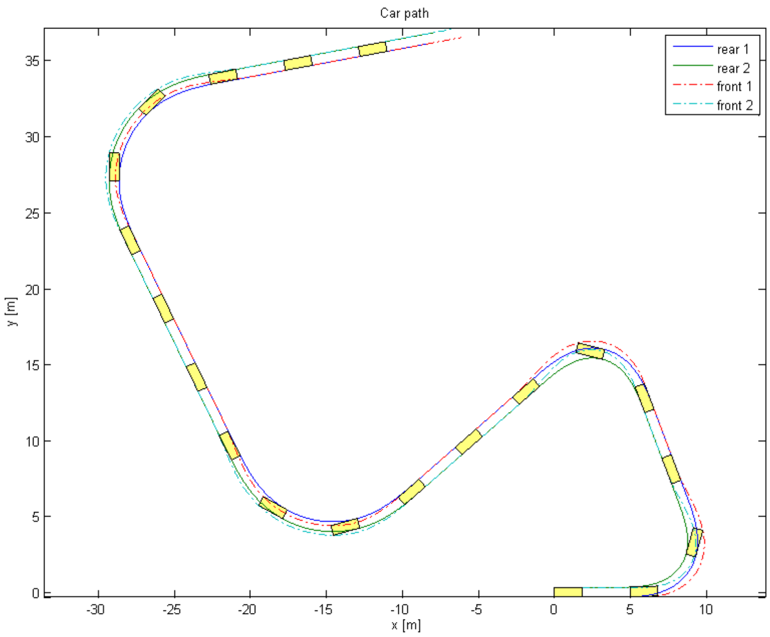

Figure 4: vehicle's path.

In each case, the path length is 100m. The final pose is:

- $x_{fin} = 100\ m$, $y_{fin} = 0\ m$, $\theta_{fin} = 0\ rad$ for the straight route and
- $x_{fin} = -7.92\ m$ $y_{fin} = 36.45m$, $\theta_{fin} = 0.184\ rad$ for the path in Figure 4

The reference point (*x*,*y*) is the center of the rear axle.

The curves are at minimum radius $(R_{min}=3\text{m})$ or at 200% of R_{min}. The curves are connected to the straight paths by means of clothoids.

Simulation results

Every simulation has brought an output similar to the one presented in Figure 5.

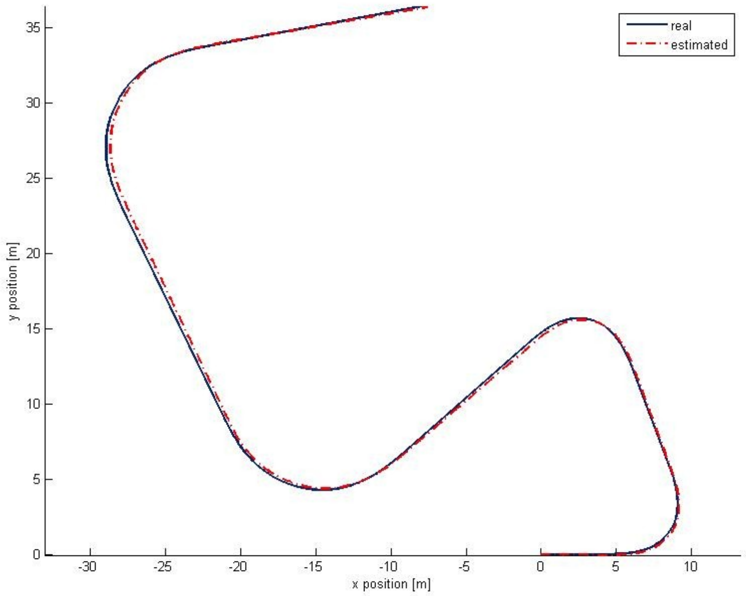

Figure 5: example of path estimation.

The analysis of the performance of the odometry system has been made on the final pose, after the 100 meters of course. All the following analyses consider the estimation of the final pose, as obtained by integrating equations 5 and 6, and compared to the real final pose.
The following figures show the distribution of the final pose after different simulations. The real final pose is represented in red, while the mean estimated pose is in green. Besides, the covariance ellipsis with 95% of confidence interval is indicated.

Example of results at 20Hz

The following figures show some results of the simulations of the final pose estimation at different speeds and different resolution with sampling frequency of 20Hz:

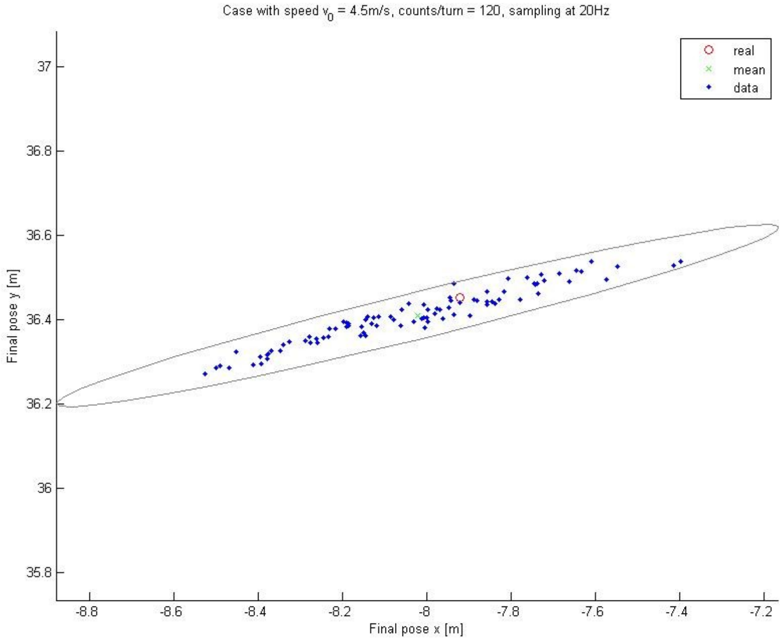

Figure 7: example results for curved path.

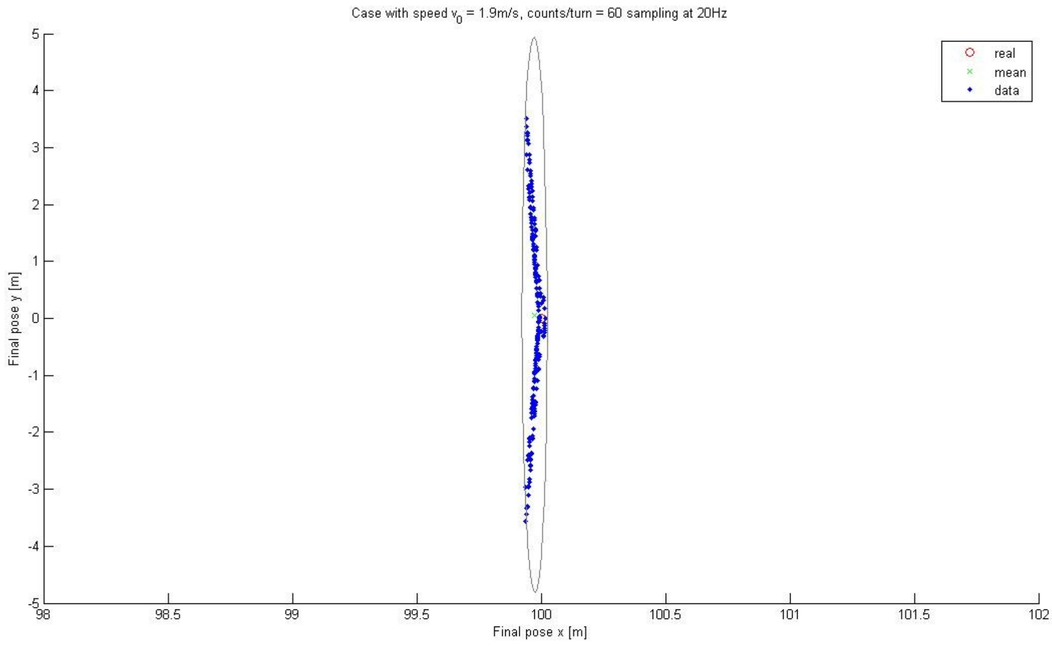

Figure 6: example results for straight path.

To better understand the results of the simulations, mean values and variances of the final pose error are considered. The error is calculated as the euclidean distance between the estimated coordinates and the real ones.

Mean values and variances

The following figures show how the number of counts per turn affect the final pose estimation at different vehicle speeds.

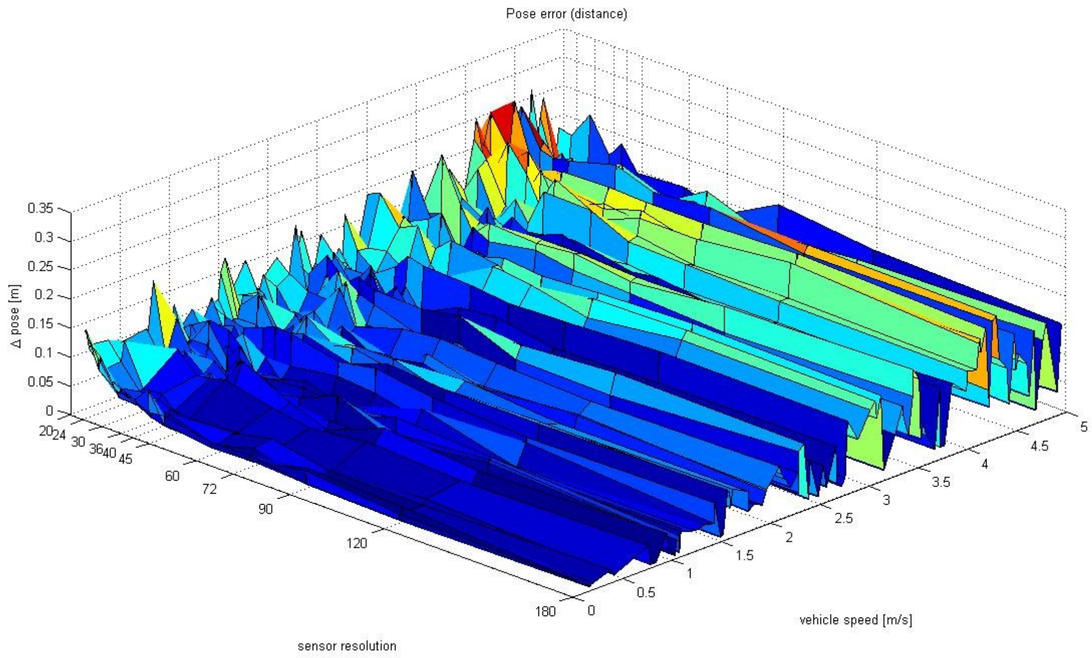

Figure 8: mean final pose error (curved path).

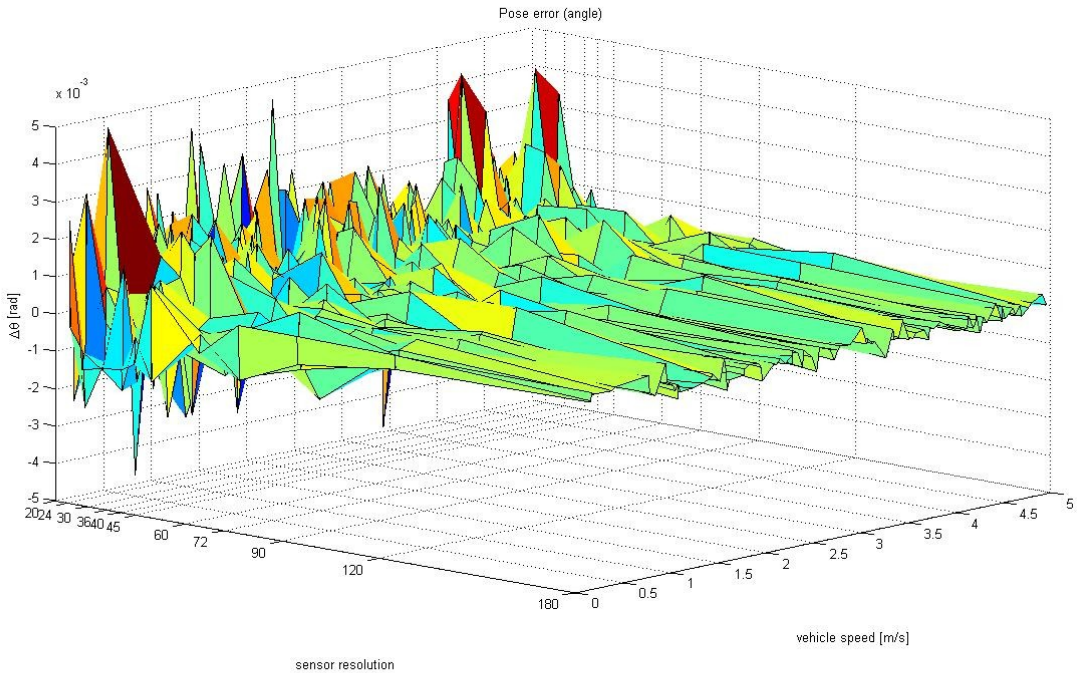

Figure 9: mean final angle error (curved path).

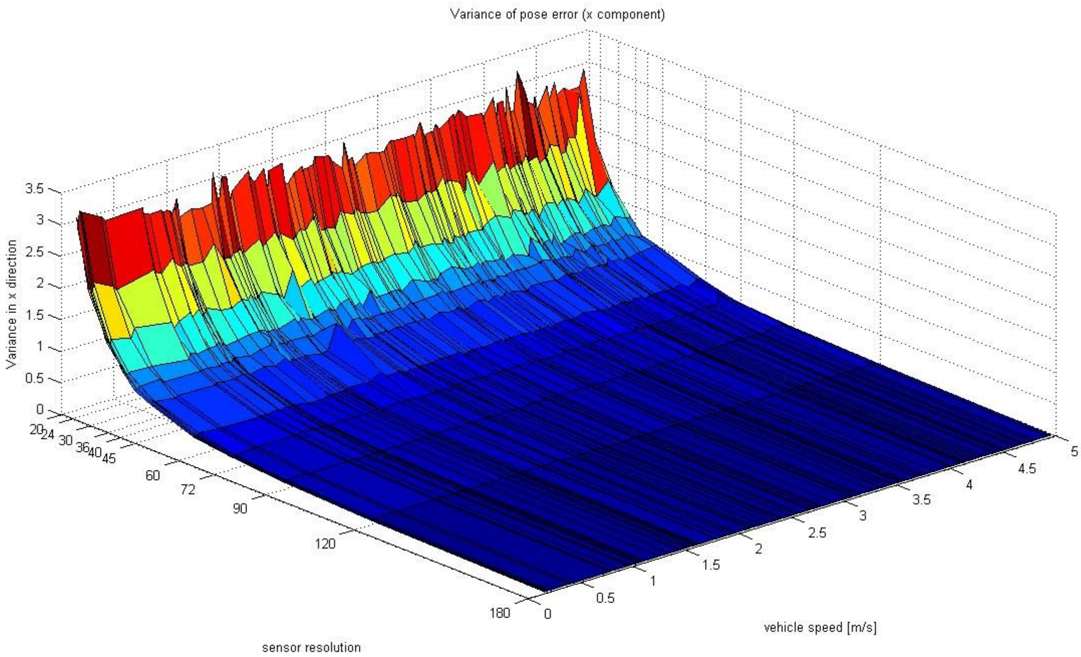

Figure 10: variance of the final pose - x component.

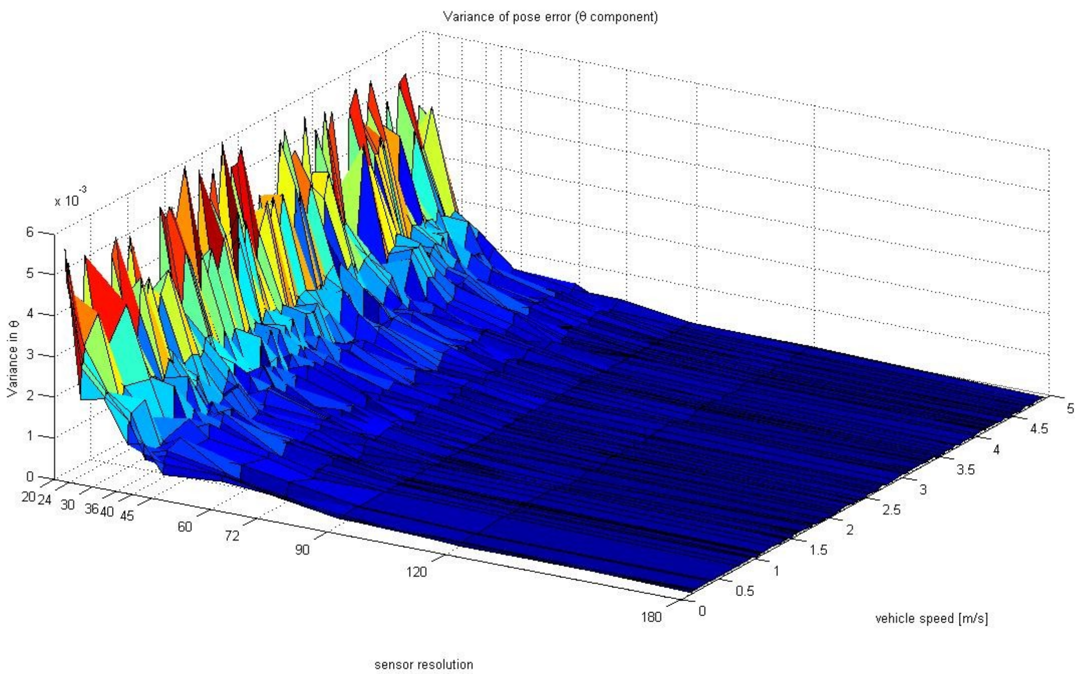

Figure 11: variance of the final pose - angle.

While the mean error seems to be slightly affected by the number of counts of the sensor, the variance of the error decreases considerably, becoming low and stable after about 90 counts per turn. The same kind of results appear with the two different paths, showing that they can be confidently considered significant in general.

The data collected allow to make some evaluations about the relation between the sensor resolution and the sampling frequency.

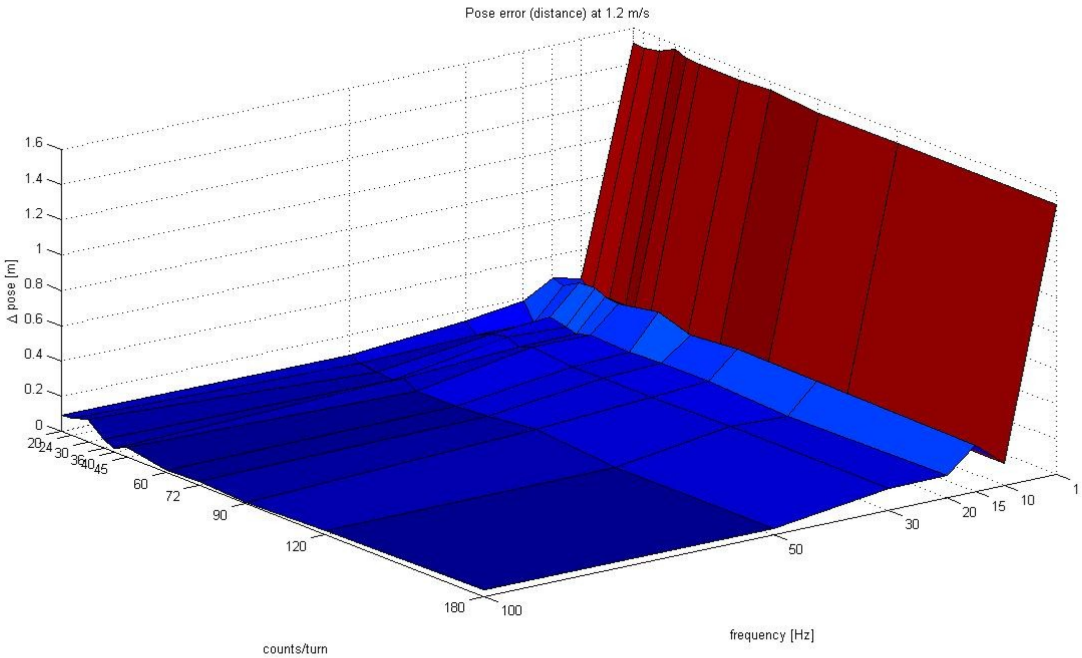

Figure 12: mean pose error w.r.t. sampling frequency and sensor resolution.

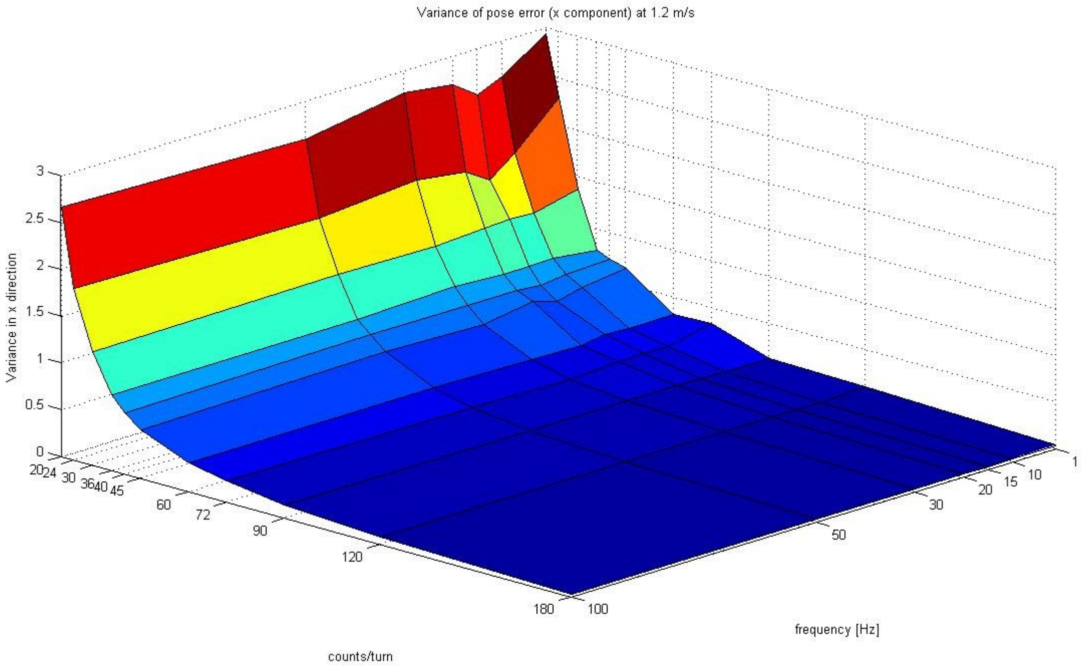

Figure 13: variance of pose error w.r.t. sampling frequency and sensor resolution.

The figures above show that the mean pose error is affected by the sampling frequency but not by the sensor resolution, while the variance of the pose error is only affected by the sensor resolution.

The Sensor

The analysis made on the variables affecting the pose estimation allow to identify the various properties of the odometry sensor to be built:

- Number of counts per turn $n_c \geq 120$
- Sampling frequency $f_s \geq 20\,Hz$

These parameters guarantee a reliable pose estimation of the vehicle.

We decided to use *reflective sensors* to set up the apparatus because the acquisition system remains almost the same as the one used for indoor acquisitions. The acquisition system allows to use quadrature encoders and to recognize motion direction by means of two 90° degrees out of phase sensors and one HCTL- 2032 IC that performs the quadrature decoder, counter, and bus interface function. The sensors considered for the application are:

- Honeywell HOA 0708/0709 reflective sensor
- Honeywell HOA 1405 reflective sensor

These sensors consist of an infrared emitting diode and a silicon phototransistor encased side by side on converging optical axes. The detector responds to radiation when a reflective object passes within its field of view. Two reflective sensors for each wheel will be mounted with a disk composed of alternating reflective and opaque sectors.

Sensor placement

The vehicle used for outdoor acquisitions is equipped with two front disc brakes and two rear drum brakes.

Figure 14: vehicle's rear drum brake.

Figure 15: vehicle's front disk brake.

The encoders will be placed into the rear drum brakes.

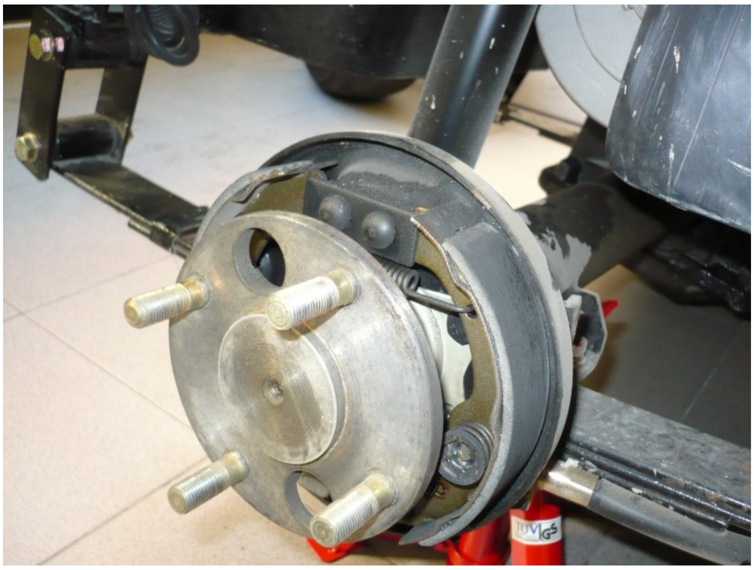

Figure 16: rear drum brake (open).

The encoder disk is placed in the inner part of the drum as shown in the following figures:

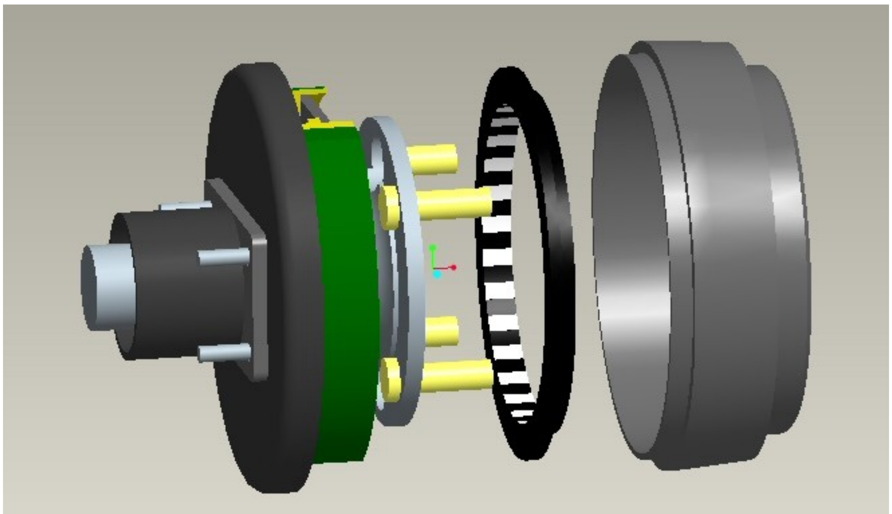

Figure 17: disk position (exploded).

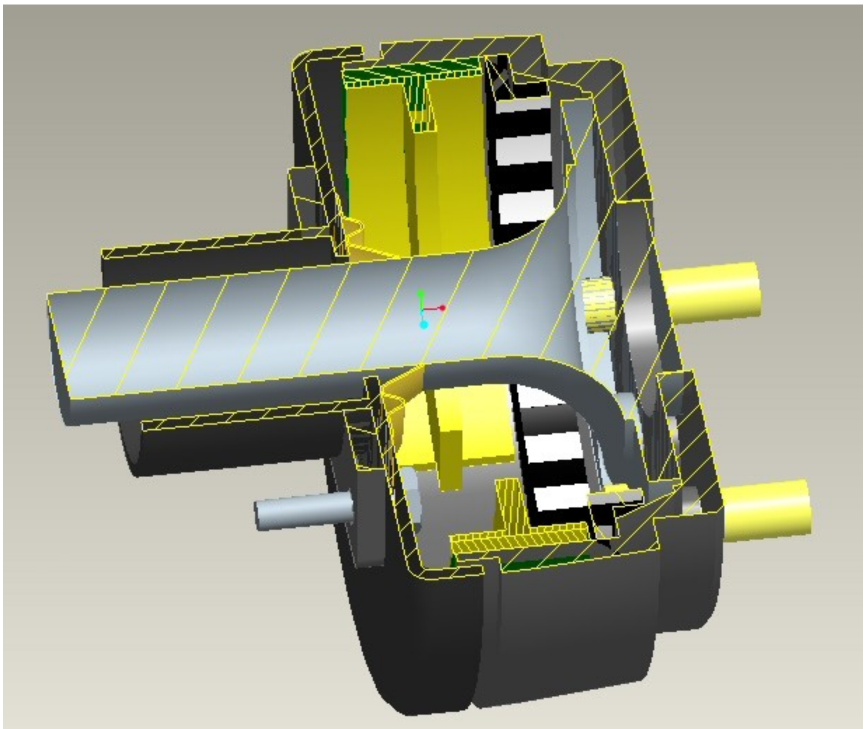

Figure 18: disk position (section).

The two sensors for each disk will be placed at the required reading distance from the disk surface and their signals are 90° off phase from each other. A suitable frame will connect the sensors to the fixed part of the brake. The frame is omitted in the figures for the sake of sensors and disk visibility.

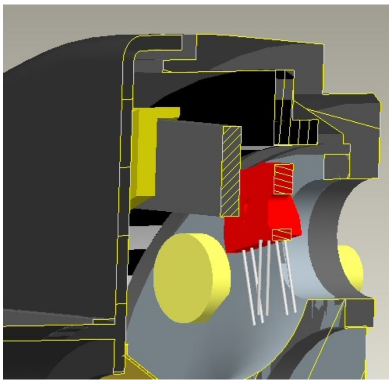

Figure 19: sensors.

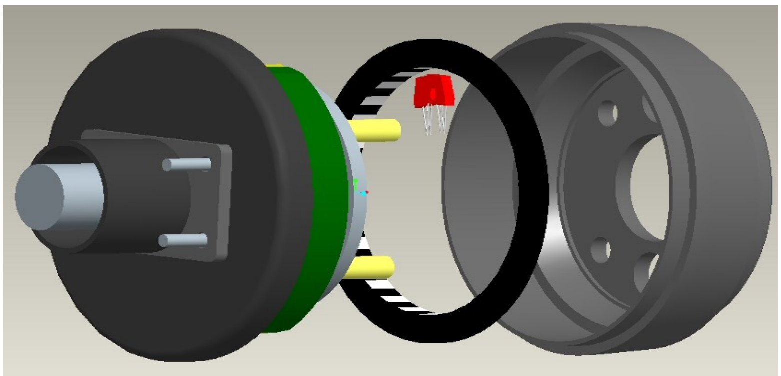

Figure 20: sensors (exploded).

Acquisition Error

The simulations described allowed to determine the main features of the odometry sensors. The error sources considered are only due to sensor geometry and to sampling frequency; so the error estimation is simply a lower bound of the real error. Like any other odometry system, many other aspects affect the total error:

- Geometric issues: the real geometry might be slightly different from the measures;
- Kinematic issues: the Ackerman system might be not perfect, so that there is no single center of rotation. In this case some wheels will slip; furthermore a finite tire width forces the tire itself to slip;
- Different wheel radius: if the radius is different form the estimate, a systematic error is added to the measures.

All the error components listed here make the real error larger. The first acquisitions with the odometry system will aim to estimate and (if possible) to reduce as many as possible error components. Anyway, real odometric systems are heavily affected by such errors, which reduce significantly their accuracy. As RAWSEEDS aims to deliver data that is consistent with real-world situations, we could choose to make available a lower-precision version of the odometry data (e.g. one evaluated by considering industry-standard encoder data, with low number of counts per turn) along with the high-precision one.

3.2 Sensor systems

The selection criteria used when choosing the **sensing suite** for outdoor data acquisition, i.e. the set of sensors which will be mounted on the robot platforms, were the same used for indoor operations: comprehensiveness, *state-of-the-art* data quality, presence of both low- and high-end devices, (when possible) commercial availability, and finally diffusion throughout the robotic community. Less importance was given to mass, bulk and dimensions, because the robots will not need to navigate through very cramped and/or narrow environments.

As outdoor autonomous robotics is a much less beaten research path, compared to indoor robotics, outdoor sensor solutions are less standardized. Therefore our choices for the outdoor sensor suite were subject to fewer *a priori* constraints; however, we strongly wanted to maximize consistency between the indoor and outdoor datasets, so our starting point in choosing the outdoor sensor suite has been the already established indoor suite.

The possibility to re-use for all outdoor operations the same *sensor frame* built for indoor data gathering (and described in Deliverable D1.1) was ruled out at the start. The first reason for that lies in the fact that some of the sensors we used for the indoor setup are almost useless in wide-open outdoor environments: this is the case of short-range sensor systems such as ultrasonic sensors or the Hokuyo URG-04LX laser range scanners. The second reason lies in the fact that the mounting framework, built for Robocom, is not suitable for use aboard the much different Alpaca platform. The *sensor frame* will be used, then, only for the scenarios where the Robocom platform has been chosen. For the Alpaca a completely new mounting system is currently being developed: however, the set of sensors that will be mounted on it has been kept - as we will see in the following section - as similar as possible to that of the *sensor frame*.
It is interesting to note, here, that the Robocom platform and the *sensor frame* are especially suitable for the *mixed* (i.e. comprising both indoor and outdoor environments) location we chose. Such location, described in the following sections, comprise a significant presence of wall- and roof-bounded tracts, where indoor-oriented sensors are usable and a compact platform is required. Conversely, for wide-open outdoor data gathering the Alpaca platform is preferrable, along with a modified sensor suite.

The RAWSEEDS outdoor sensing suite that will be used with the Robocom platform is the same described in Deliverable D1.1. It is then composed of:

1. robot odometry;
2. binocular and trinocular black-and-white (B/W) vision;
3. normal perspective, color and B/W cameras;
4. omnidirectional color vision with hyperbolic mirror;
5. short-range (<4m range, shorter at low reflectivity) cheap Laser Range Finders (LRF);
6. medium- and long- range (respectively <30m and <100m range, at 100% reflectivity) high performance LRFs;
7. sonar belt with multiple ultrasonic sensors;
8. Inertial Measurement Unit (IMU) providing 3-axis angular orientation, acceleration, rate-of-turn and Earth magnetic field data.

We are currently evaluating the addition to these sensors of an upward-looking camera, whose data is usable by algorithms using roof images for localization. Of course, such a camera would be useful only in indoor or *mixed* scenarios.
The sensing suite that will be mounted on the Alpaca platform for outdoor operations is composed of the following types of sensors:

- robot odometry;
- binocular and trinocular black-and-white (B/W) vision;
- normal perspective, color and B/W cameras;
- omnidirectional color vision with hyperbolic mirror;
- medium- and long- range (respectively <30m and <100m range, at 100% reflectivity) high performance Laser Range Finders (LRF);
- Inertial Measurement Unit (IMU) providing 3-axis angular orientation, acceleration, rate-of-turn and Earth magnetic field data.

Comparation of this list with the preceding one reveals, as expected, the omission of the short-range LRF and ultrasonic sensors. In both lists the binocular and b/w monocular systems are, in practice, realized with subsets of the trinocular system, to avoid adding unnecessary devices to an already very populated sensor suite.
The possibility of adding a GPS positioning sensor to one or both the above sensing suites is currently under scrutiny of RAWSEEDS' Partners. The kind of device we are evaluating is a low-end GPS sensor, similar to those used in automotive applications, and the justification for its use would be the will to add a data stream that could be useful for automotive applications.

The specific devices used for both outdoor sensing suites (items tagged with an asterisk are not present in both suites) are:

1. Odometry, made available by the computer-based odometry systems fitted to Robocom and Alpaca.
2. Binocular vision system composed of a two-camera Videre Design STH-DCSG-VAR system (two FireWire, B/W, 640x480 pixel cameras mounted on a common mechanical frame that allows for an adjustable baseline). Trinocular vision system is realized combining the binocular STH-DCSG-VAR with an additional Videre Design DCSG camera (the same camera used by the STH-DCSG-VAR). Although CMOS, these cameras feature a global shutter, which is important for shooting moving scenes or from a moving observer (both things happen in our case). Web: http://www.videredesign.com/sthdcsgvar.htm, http://www.videredesign.com/Templates/dcsg.htm
3. Each of the three cameras of the trinocular system provides a B/W monocular data stream. Color monocular vision is covered by an Unibrain Fire-i 400 camera (FireWire, color, 640x480 pixel). Web: http://www.unibrain.com/Products/VisionImg/Fire_i_400_Industrial.htm

4. Omnidirectional color vision is obtained by using a Unibrain Fire-i 400 camera fitted with a hyperbolic mirror built by Vstone. Web: http://www.vstone.co.jp/e/EVSC15MR15MR37.pdf
5. (*) 2 Hokuyo URG-04LX LRSs, mounted on the front and the back of the robot. The LRSs will be tilted down, pointing at the floor about 2 to 3 meters away from the robot (precise orientation will be chosen later). Unfortunately, the tilting base that we hoped to use under one of the Hokuyo sensors has proved to be unsuitable to RAWSEEDS' requirements (please note that the construction of that base was not part of the RAWSEEDS project). Web: http://www.hokuyo-aut.jp/products/urg/urg.htm
6. Sick LMS291 and LMS200 LRSs, mounted on the front and the back of the robot. Web: http://mysick.com/partnerPortal/eCat.aspx?c=1&go=FinderSearch&Cat=Row&At=Fa&Cult=English&Category=Produktfinder&FamilyID=267&Selections=8641,0,0,8775,0 http://mysick.com/partnerPortal/eCat.aspx?go=FinderSearch&Cat=Row&At=Fa&Cult=English&Category=Produktfinder&FamilyID=267&Selections=8644,0,0,8775,0
7. (*) Sonar belt composed of 12 to 16 Polaroid 600-series sensors (positioned all around the robot) and associated control electronics built by POLIMI. The number of Polaroid sensors actually used depends on availability at the moment of the data acquisition.
8. Xsense MTi IMU (USB, 1,7g full scale acceleration, 150deg/s full scale rate of turn). Web: http://www.xsens.com/index.php?mainmenu=products&submenu=machine_motion&subsubmenu=MTi

The mounting system for both the above suites is a rigid framework composed of aluminium profiles and joints by Item (http://www.item.info). In the case of RAWSEEDS' *sensor frame*, designed to be mounted on Robocom, the framework is already finalized and currently in use; for the additional platform Alpaca a new, larger framework is under construction. The name *sensor frame* will not be used for the latter as it will not house all the equipment associated to the sensors: part of it will in fact be located on the loading platform of the Alpaca vehicle.

3.3 Setup of the data-acquisition robots

In this document the term **data-acquisition robot** is used to define the union of a platform (Robocom or Alpaca) with its own specific sensor complement and the associated data elaboration systems and power supplies. A *data-acquisition robot* is therefore a complete system used to acquire the datasets on which the RAWSEEDS toolkit is based.
The following sections give a description of how RAWSEEDS' data-acquisition robots have been designed and assembled. As much of the information about this issue has already been given in Deliverable D1.1, only a cursory reference

to the main issues will be made.

3.3.1 Data processing and storage

Processing power is not a prominent issue for RAWSEEDS, as most of the work on sensor data is limited to acquisition and storage. However, the overall data stream from the whole sensor suite is massive, so the setup of the data processing platform is important. Further constraints were given, through the design phase, by the fact that power consumption had to be as low as possible, because the battery power supplies have limited capacity.
As explained in D1.1, we chose to use multiple low-power, small-footprint x86-compatible PCs. These machines, called *PCBricks*, are based on a VIA EN15000 motherboard. They combine very low power consumption (20W maximum) with a rich set of data interfaces, a fast hard disk and a rather low (but sufficient for RAWSEEDS) computing power.
The PCBricks use the Linux operating system; all the software used for acquisition and synchronization of sensor data, as well as for synchronization between the PCBricks and between them and the ground truth collection system (which will be described in the following sections) has been written by POLIMI.
Acquisition tests have shown that two PCBricks are sufficient for RAWSEEDS; however, we have built five PCBricks, so additional PCBricks could be added to the acquisition setup with little effort. It must be stressed that we deliberately made the choice of a modular design, and used processing facilities thoroughly independent from those governing the platforms (in the case of Robocom; Alpaca has no on-board computers). In this way the passage from one platform to the other and (possibly) on-site modification of the system during acquisition campaigns are greatly simplified.

3.3.2 Networking and synchronization

As explained in D1.1, the computers onboard each data-acquisition robot are interconnected by a 100Mbps wired LAN. Acting as a central hub for the LAN is a router/switch, mounted on the robot. The router communicates with external systems (such as the portable PCs used by the operators or, most importantly, the ground truth collection system) through a 54Mbps IEEE 802.11g wireless network.

The above network connections are mainly used for synchronization between the machines. This is in fact a very critical issue, as RAWSEEDS requires that each single data element (image frame, laser scan, odometry block of data, and so on) produced by the sensors is precisely timestamped with reference to a central clock. To reach the high quality and precision that RAWSEEDS seeks, timestamping errors must be kept very low. Precisely, the errors have to be well below the period of the data having the fastest upgrade rate: as shown in D1.1, this means that they should not exceed a few milliseconds. Moreover, this level of precision must be guaranteed over the whole time span (hours, or even days) of a complete data-gathering session.

Initial versions of the data-acquisition software were based on the Network Time Protocol (http://www.ntp.org/) synchronization system. However, preliminary tests showed that this system was difficult to integrate into our software and had a barely sufficient synchronization accuracy, so it was abandoned. We then switched the software design towards the much more efficient and precise Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, or *PTP* (http://ieee1588.nist.gov/). This protocol ensures peak time errors of less than 0.1ms (for comparison, NTP is at least 10 times less accurate in the best cases) and is now well integrated into our software framework.

3.3.3 Interconnections

Interconnection between the elements of RAWSEEDS' setup for outdoor acquisition is very similar to that used for indoor acquisition. There will be only one change, related to the fact that the Mac mini computer will not be present on the Alpaca platform. Therefore, when using that platform, the only function related to data acquisition performed by the Mac mini, i.e. acquisition of odometry data, will be assigned to one of the PCBricks. This is not a problem, as the location where the Alpaca is appropriate does not require the use of the Hokuyo and sonar sensors, thus freeing communication ports and processing power on the PCBricks.

3.3.4 Power usage and operating life

As the sensor suite chosen for outdoor operations is mostly coincident with that used for indoor activities, the considerations about power usage and operating life made in Deliverable D1.1 are still valid and will not be repeated here. It could be observed that by removing the short-range sensors from the sensor suite the overall power consumption will be lower: in practice, power usage of such sensors was low, and thus the operating time (estimated in at least two hours of continuous operation) is unchanged. This is true even if a GPS sensor will be added to the suite, as this kind of devices have very low power consumption.
Of course, what was written in D1.1 assumed the use of the Robocom platform. The Alpaca platform has a much more capable power supply than Robocom: therefore its operating life greatly exceeds that of the sensor system, and thus is not an issue.

3.3 Complete data-acquisition robot

A first data-acquisition robot has already been designed and built. It is the Robocom robot equipped with RAWSEEDS' sensor frame, as shown in Figure 1 and, from a different point of view, in Figure 21. This robot is sufficient in itself for RAWSEEDS' purposes, as it is capable of exploring all the locations chosen for the outdoor scenarios.

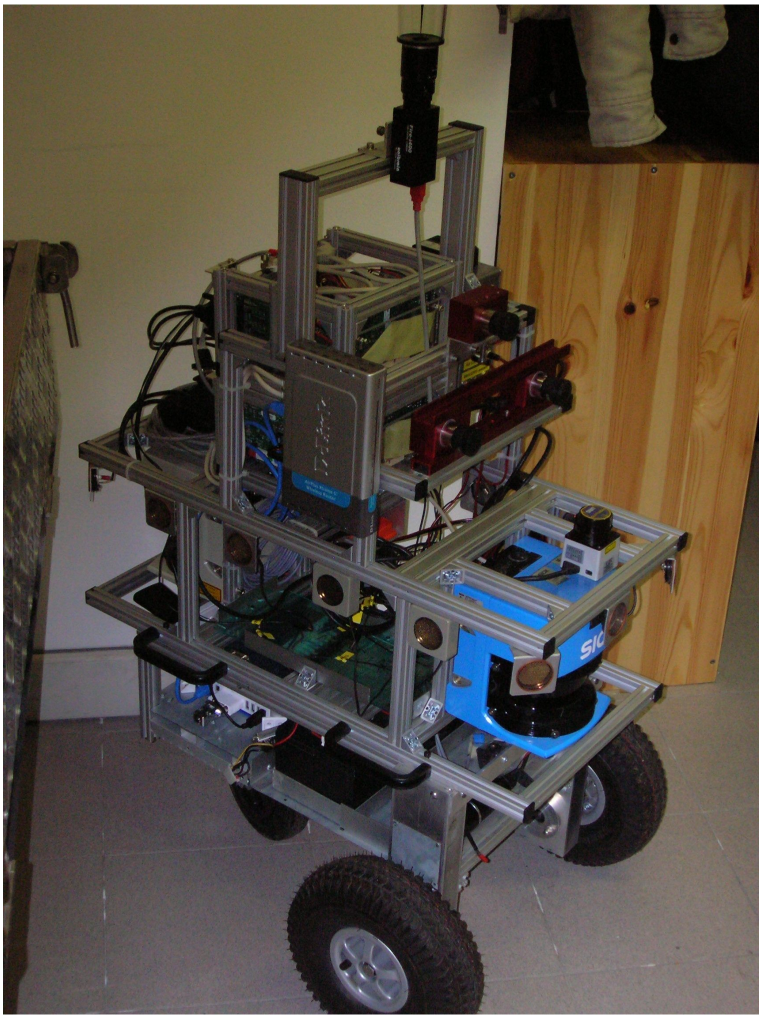

Figure 21: Robocom platform with RAWSEEDS' sensor frame.

However, as said in previous sections, we are working on the setup of a second vehicle, based on a completely different platform: the Alpaca four-wheel electric cart. We are doing that because we think that automotive applications of robotic technologies will gain a significant importance in the next years, and would like to provide RAWSEEDS' users with a dataset suitable for the test of software for automotive applications. For that reason we chose a secondary platform (the Alpaca) with the same kinematics of an automotive vehicle, and are fitting it with a suitable odometry system.

4. Outdoor scenarios

Outdoor environments are much less uniform than indoor ones: the range of terrain types, obstacles and structures, lighting conditions, etc. which can be found outdoor is extremely wide. Moreover, each of the above listed aspects of the environment can be subject to large variations even within a single location: for example, lighting or terrain changes in the woods can be enormous. Therefore, given the limited resources available, the first thing to be done in the planning of RAWSEEDS' outdoor operations has been a drastic narrowing of the range of environments that will be actually explored and used to build the RAWSEEDS benchmarking toolkit.
The first choice we made was to remain within urban areas, thus excluding much of the variation range outlined above. In this way we wanted to concentrate on the kind of environments that foreseeable robotic applications, to be developed in the short and medium term, are more likely to be targeted to. Then we extracted, from the spectrum of typical urban environments, two specific outdoor scenarios:

- a *mixed* scenario, i.e. a scenario including parts where the trajectory of the robot is surrounded by walls and/or roof and parts where the trajectory is located in the open;
- an *open-air* scenario, i.e. a scenario where the robot moves in the open (between buildings) and the obstacles are comparable with those found along urban roads.

Please remember that a scenario is defined as "*the complete set of conditions defining a single data-gathering session*".
The above outdoor scenarios have been chosen to include a comprehensive set of the kind of environments where a mobile robot designed for outdoor operation in urban (or urban-like) context will likely need to operate. Please note that the dataset associated to the *mixed* scenario includes data gathered both in indoor and outdoor environments, and has been designed to stimulate the development and test of novel software algorithms capable of satisfactory performance in both conditions, while being as well useful for indoor-only or outdoor-only algorithms.
The following sections describe the scenarios for the outdoor data-gathering activity of RAWSEEDS.

4.1 Locations

The locations chosen for the outdoor scenarios are two:

- **POLIMI-Leonardo**, i.e. the old, historic campus of the Politecnico di Milano. It is a set of low buildings, separated by small roads and interconnected by a variety of passages, from partially walled corridors to wide-open squares.
- **POLIMI-Durando**, one of the newest campuses of the Politecnico di Milano. It is a refurbished factory site, characterized by large buildings of

different shape and kind, arranged in a disomogeneous and irregular way and separated by rather wide asphalt roads.

These locations have many advantages:

- they possess all the characteristics of a wide range of urban environments, and specifically of urban roads, while not presenting the difficulties that data gathering in a real urban environment poses (e.g. the necessity of blocking the roads);
- they are part of the Politecnico di Milano (one of the Partners), and therefore allows for easy access to RAWSEEDS' personnel, easy storage of all the experimental gear over the many days that the data-gathering sessions will span on, and even convenient electric outlets to power all the equipment and/or recharge its batteries;
- they have much less limitations on the installation of structures and apparatus (such as the ground truth collection system that will be described in the following sections) than generic public places;
- we have access to detailed CAD drawings and maps;
- we can obtain all the authorizations necessary to perform the acquisition experiments directly from the University, without having to deal with external public bodies (e.g. Local Police) and so greatly speeding up the activities.

Details about the outdoor locations will be given in the following sections.

4.2 Scenarios

For indoor data-gathering, RAWSEEDS selected a single location and defined multiple scenarios, covering different conditions of lighting and dinamicity (in particular the presence or absence of people). This was made possible by the fact that the chosen environment could be put under our complete control during the course of the data acquisition sessions.
For the outdoor data-gathering sessions, a different approach has been chosen. First of all, it is difficult that very large, open, public environments such as the ones we selected can be put under the complete control of RAWSEEDS; moreover, in open spaces the lighting is largely uncontrollable and for the most part due to the Sun. For this reason we chose to explore two different locations, perform a dynamic data acquisition session for each one of them, and then try to organize and perform corresponding static sessions. As the latter require that the locations are devoid of people (e.g. during the week-end) but open to us, the actual possibility to perform such data-gathering sessions depends on specific authorizations.

It must be noted that the complexity of the data acquisition gear and setup and the dimensions of the environments to be explored will lead to data-acquisition sessions spanning a whole day or more (not considering preparation activities performed in advance). Therefore lighting conditions will certainly vary during the acquisition, not only because different environments will be visited by the robots within each location, but simply due to the change in sunlight during the day. This is not a problem, as this is one of the typical characteristics of outdoor environments and as such must be represented into the datasets, but will make the outdoor datasets more *difficult* than the indoor ones. By the way, this is one of the reasons why robotic applications are still confined, for the most part, to indoor environments. We hope that the availability of RAWSEEDS' datasets will give to the operators in the field of robotics a way to develop and test outdoor applications without having to endure the difficulty and the cost of setting up their own data-acquisition campaigns.

4.3 Data-acquisition methods

Two methods for the acquisition of sensor data are commonly used in mobile robotics: *stop-and-go acquisition* and *continuous acquisition*. In Deliverable D1.1 we discussed them and explained why continuous acquisition is best for the indoor data-acquisition activity of RAWSEEDS. As the reasons given for that are even more valid for outdoor acquisition, here we only say that the outdoor data-acquisition operations of RAWSEEDS will follow the continuous acquisition approach as well.
The choice of the frame rate and exposure time for the cameras, which is subject to the same issues and limitations outlined in D1.1, will be made after suitable acquisition tests in outdoor environments. We expect, though, to be able to adopt frame capture parameters similar to those chosen for indoor acquisition.

4.4 Data-gathering sessions

4.4.1 Location POLIMI-Leonardo

The map of the Politecnico di Milano campus set in Piazza Leonardo da Vinci (Milan), i.e. of the POLIMI-Leonardo location, is shown in Figure 22. The campus has roughly the shape of a square with 400m-long sides.
As previously observed, this location is a wide, well-integrated complex, and is the historical location of Politecnico di Milano since 1927. The buildings have been modified through the years, but mantain an overall "classical" structure, sporting columns in many areas. The campus includes many low buildings, often having inner cloisters of some kind, separated one from the other by narrow roads. These buildings are connected by a variety of different passages, including open (but roofed) promenades flanking inner cloisters, corridors (more or less enclosed by walls: from those that are completely surrounded to other that only have a roof), and a wide central square with trees and grass. Floors are partly asphalt and partly polished stone, with occasional gravel; they are for the most part level and smooth, but sometimes ramps or steps are found.
As the following figures will show, POLIMI-Leonardo sports an impressive

collection of different outdoor or mixed (i.e. only partially enclosed by walls) environments, although lacking "wide-open" spaces.
This location includes many narrow passageways: the Robocom platform is therefore the only one suitable for it.

The map also shows (in red) a possible trajectory for the robot: in practice the actual trajectory has still to be defined, and the one shown here is only intended as a guide to the passages traversable by the robot.

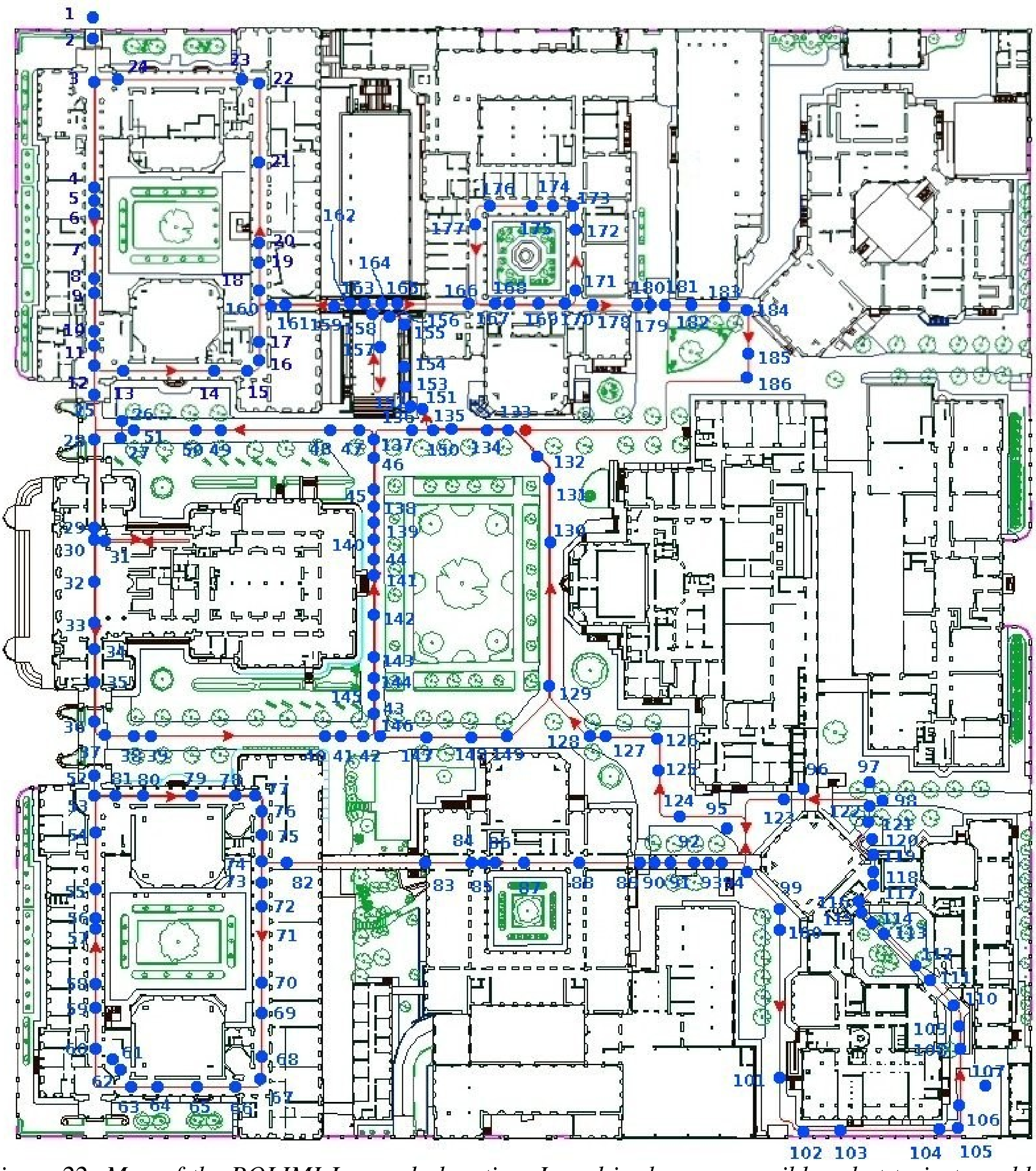

Figure 22: Map of the POLIMI-Leonardo location. In red is shown a possible robot trajectory; blue dots represent the places where pictures have been taken.

Figures 23 to 54 show some vistas of the POLIMI-Leonardo location. Please note that the numbers in the pictures' captions make reference to those in the map of Figure 22.

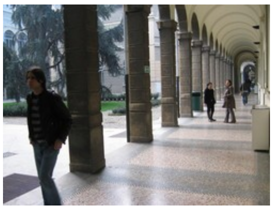
Figure 23: Picture 4.


Figure 24: Picture 6.


Figure 25: Picture 11.


Figure 26: Picture 13.

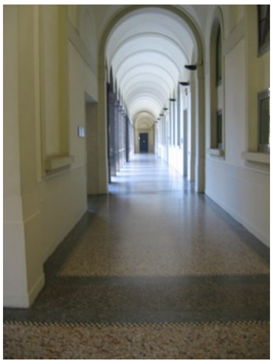
Figure 27: Picture 17.


Figure 28: Picture 20.


Figure 29: Picture 24.


Figure 30: Picture 26.


Figure 31: Picture 28.


Figure 32: Picture 29.


Figure 33: Picture 37.


Figure 34: Picture 38.


Figure 35: Picture 39.


Figure 36: Picture 44.


Figure 37: Picture 46.


Figure 38: Picture 48.


Figure 39: Picture 83.

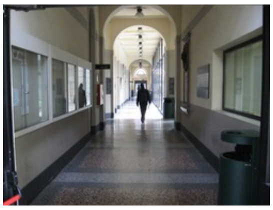
Figure 40: Picture 85.


Figure 41: Picture 87.


Figure 42: Picture 89.


Figure 43: Picture 99.


Figure 44: Picture 102.


Figure 45: Picture 105.


Figure 46: Picture 110.


Figure 47: Picture 111.


Figure 48: Picture 127.


Figure 49: Picture 129.


Figure 50: Picture 134.


Figure 51: Picture 137.


Figure 52: Picture 142.


Figure 53: Picture 154.


Figure 54: Picture 162.

4.4.2 Location POLIMI-Durando

shows the map of the POLIMI-Durando location, i.e. of the Politecnico di Milano campus set in via Durando (Milan). As previously said, it comprises a set of buildings that once housed a factory. It has been modified for its new use trying to retain as much as possible of its original character, so it has a very composite (and interesting for RAWSEEDS) nature, which closely mimics that of a typical (small) city. As the following pictures will show, buildings of different kind and style are included in the campus, as well as characteristics such as slopes, passages of various widths, external stairs, and so on. The presence of parked (and occasionally moving) cars, sidewalks, poles and of moving people (students) closely approaches that of a typical urban road. As the following figures will show, the passages in the POLIMI-Durando location are wide enough to allow easy movement to the Alpaca data-gathering robot.
The map also shows (in red) a possible trajectory for the robots: in practice the actual trajectory has still to be defined, and the one shown here is only intended as a guide to the passages traversable by the robots.
Figures 56 to 64 show some vistas of the POLIMI-Durando location. Please note that the numbers in the pictures' captions make reference to those in the map of Figure 55.

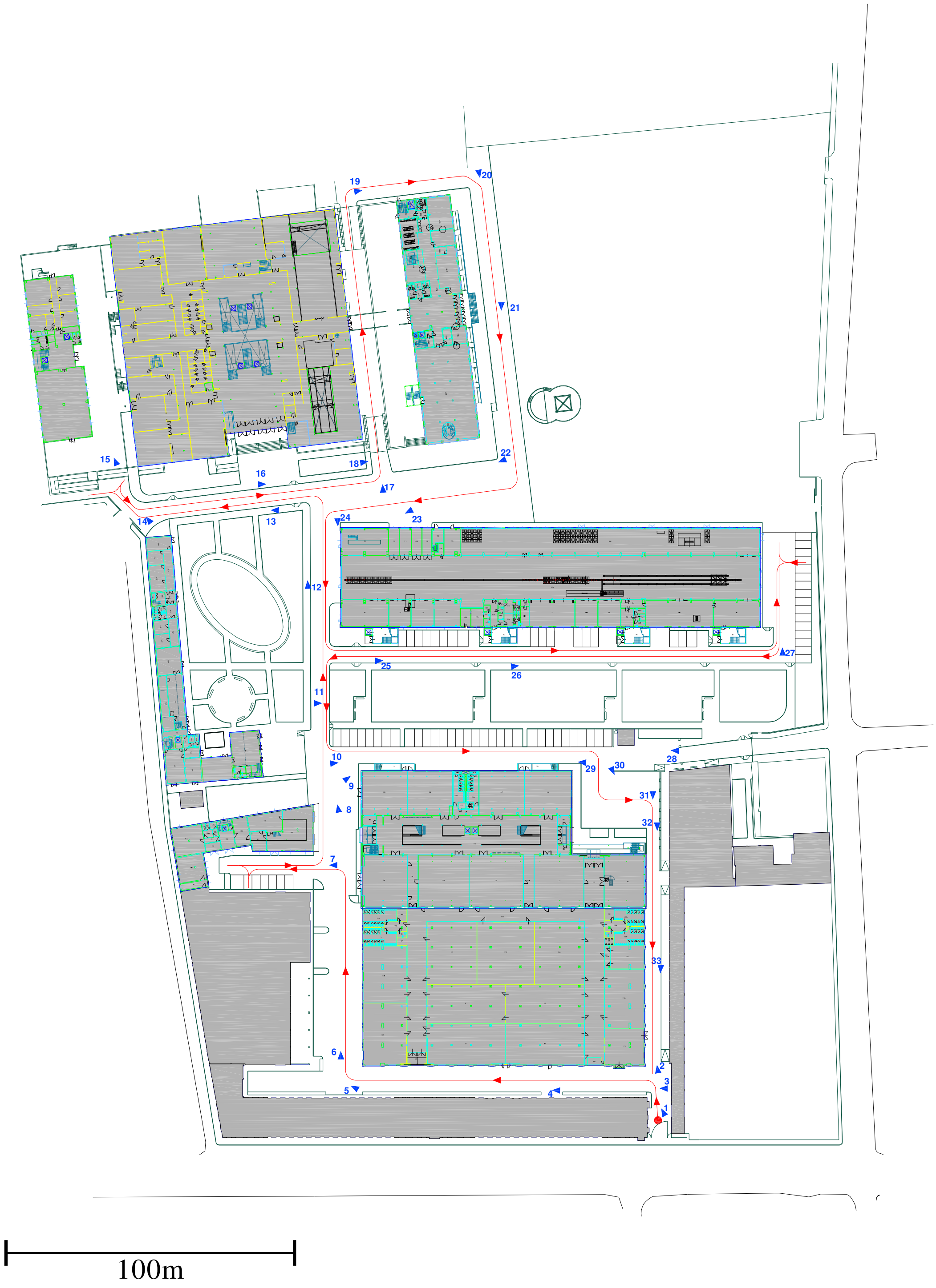

Figure 55: Map of the POLIMI-Durando location. In red is shown a possible robot trajectory; blue triangles represent the places where pictures have been taken.

Figure 56: Picture 2.

Figure 57: Picture 3.

Figure 58: Picture 5.

Figure 59: Picture 6.

Figure 60: Picture 8.

Figure 61: Picture 10.


Figure 62: Picture 12.


Figure 63: Picture 13.


Figure 64: Picture 17.


Figure 65: Picture 19.

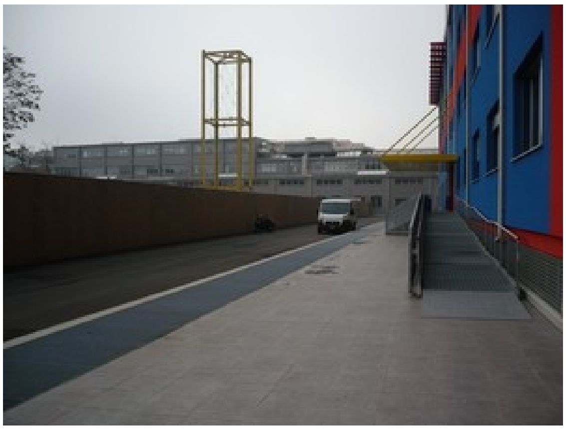
Figure 66: Picture 20.


Figure 67: Picture 21.

Figure 68: Picture 22.

Figure 69: Picture 25.

Figure 70: Picture 30.

Figure 71: Picture 32.

4.4.3 Session schedule

The RAWSEEDS outdoor data-gathering campaign with platform Robocom will be executed shortly after the end of the indoor acquisition activities. Data-gathering activities involving the Alpaca platform require the fitting of that platform with a suitable odometry system to be finished: thus it is impossible to give a schedule for that at the moment.

5. Data validation

The problems and solutions associated to the validation of the data generated by RAWSEEDS' outdoor data gathering activities are exactly the same described, with reference to indoor data gathering, in D1.1. In fact, they do not depend from the *content* of the data streams. For that reason this Chapter of Deliverable D1.2 is limited to a reference to Chapter 5 of D1.1.

6. Ground truth

The collection of a (sufficiently precise) *ground truth* associated to the datasets is very important for RAWSEEDS. The project's main goal is the definition of advanced tools for the development of solutions to three important robotics problems:

- map building from the data collected by the robot sensors (mapping problem);
- robot pose estimation given a known map and sensor readings (self-localization problem);
- robot pose estimation and map building at the same time (Simultaneous Localization And Mapping problem, or SLAM problem).

Each of these problems can be solved (with different degrees of success) by suitable algorithms. To test the performance of such algorithms (i.e., in practice, of the software implementation of them) it is necessary that the real data that they are meant to reconstruct (i.e., the ground truth) is known in the first place. This requires that such ground truth has been collected and recorded. As part of the ground truth is given by the actual trajectory of the robot during the data-acquisition operations, the collection of such part must be necessarily performed during the acquisition.
To be significant, the ground truth must be collected by means that are independent from the sensors mounted aboard the robot. To be useful, it must have a precision that is at least comparable with that of the results of the better algorithms, when applied to the datasets associated to the ground truth. These constraints lead to the fact that ground truth collection for a project such as RAWSEEDS is a very complex and difficult problem, which does not have a simple or unique solution and that has to be examined in depth before acting.

In Deliverable D1.1 we analyzed the general ground truth issue, and the problem of collecting it for RAWSEEDS' indoor data-gathering. Much of that analysis is valid for outdoor data-collection too, and thus will not be repeated here. However, the technical means suggested in D1.1 are not suitable for large areas such as the ones that will be covered by RAWSEEDS' outdoor *open-air* data-collection sessions. Fortunately alternative technologies are available, and will be described in the following sections.

6.1 Ground truth for localization

Ground truth for the robot position and orientation is necessary for the evaluation of the solutions to both the self-localization and the SLAM problems.
As we discovered during the work of WP-1, no suitable commercial solutions exist for the problem of indentifying the position of an object in indoor environments (this, indeed, led RAWSEEDS to the design and setup of a

custom system). Conversely, such problem has many commercial solutions when outdoor environments are considered: the best known of which is the use of GPS-based navigators (GPS stands for Global Positioning System). Unfortunately, the real-world precision of such devices is much lower than required by RAWSEEDS: localization is generally affected by an error of about 15 metres. This error is due to the limited theoretical accuracy of the system (a best-case error of 3m is given) and to other issues, such as the effect of Earth's athmosphere on the signals from the GPS satellites.
Fortunately, GPS-based systems with better precision do exist, although they are complex and costly. For RAWSEEDS we selected one of these, called Real Time Kinematic GPS (RTK-GPS; a good introduction to RTK systems is available at http://en.wikipedia.org/wiki/Real_Time_Kinematic). It is a system that, like GPS, measures distances using the time offset between multiple copies of the same signal; unlike GPS systems, though, RTK-GPS detects the time offset between copies of the *carrier wave* of the GPS signal, not of the payload transported by that wave. In this way precision is greatly improved, and theoretical error in *absolute* positioning falls to about 20cm.
Commercial RTK-GPS systems are composed of two devices: a *base station*, which must be positioned in a fixed location with reference to Earth's surface, and a *mobile unit*, which can be moved in the vicinity. If needed, more than one mobile unit can be present for each base station. The base station receives the GPS signal, evaluates the phase of its carrier and transmits this information to the mobile unit, thus greatly simplifying the task of the latter. A side effect of this splitting is that in a two-station RTK-GPS system the absolute position of the mobile unit is known with an error of 20cm, but the *relative position of the mobile unit with reference to the base station* has an error of about 1cm (horizontal) or 2cm (vertical).

For RAWSEEDS we will rent such a two-unit RTK-GPS system. By carefully measuring the position of the base station on the map of the explored environment, we then expect to be able to localize the mobile unit - mounted aboard the data-gathering robot - on the map (or, generally, w.r.t. the base station) with a horizontal error of 1cm. This is more than sufficient for RAWSEEDS' ground truth.
Of course, like any GPS-based localization system, RTK-GPS needs that both the base unit and the mobile unit perceive the signal coming from GPS satellites. This means that (if the base unit is correctly positioned) spaces where the satellites are not in sight of the mobile unit cannot be covered by the system. We expect this problem to occur mainly for the POLIMI-Leonardo location. Anyway, this limitation does not constitute a damage to RAWSEEDS: exactly as in the indoor data-gathering sessions only a fraction of trajectory of the robot was covered by the ground truth collection system, we accept that a similar situation can occur for outdoor locations. We will choose the course of the robot through the locations in such a way that coverage and usefulness of the ground truth collection system is maximized.

6.2 Ground truth for mapping

For the mapping problem, RAWSEEDS' stance towards the issue of ground truth is the same that we have described in Deliverable D1.1. In fact this specific ground truth collection problem is not influenced by the fact that the environment is indoor or outdoor. We then make reference to the conclusions reached in D1.1, i.e., that we will mainly use executive drawings, possibly integrated by specific measurements. The latter, by the way, could be performed by using the mobile unit of the RTK-GPS system described in previous sections.

7. Benchmark Problems

A Benchmark Problem, or BP, is defined as the union of: (i) the detailed and unambiguous description of a task; (ii) an extensive, detailed and *validated* collection of multisensorial data, gathered through experimental activity, to be used as the input for the execution of the task; (iii) a rating methodology for the evaluation of the results of the task execution. The application of the given rating methodology to the output of an algorithm or piece of software designed to solve a Benchmark Problem produces a set of scores that can be used to assess the performance of the algorithm or compare it with other algorithms.
[*From RAWSEEDS' Description Of Work (Annex I to the Contract between Partners and EU).*]

The creation of the BPs is the work of WorkPackage 4. At the present stage in the development of the RAWSEEDS project it is only possible to give a general account of the kind of problems that RAWSEEDS' Partners expect to build. As these general information has already been given in Chapter 7 of Deliverable D1.1, here we only make a reference to that document.
In addition to that, it can be said that we expect the BPs built on the outdoor datasets to be more "difficult" to solve than corresponding BPs using indoor data only, i.e. to require more sophisticated algorithms. This is due to the fact that in outdoor environments the variations in lighting, kind of surfaces, distance from robot to obstacles, kind of terrain, and so on are generally much more pronounced than in indoor ones.

8. Benchmark Solutions

A Benchmark Solution, or BS, is defined as the union of: (i) a BP; (ii) the detailed description of an algorithm for the solution of the BP (possibly including the source code of its implementation and/or executable code); (iii) the complete output of the algorithm applied to the BP; (iv) the rating of this output, calculated with the methodology specified in the BP.
[*From RAWSEEDS' Description Of Work (Annex I to the Contract between Partners and EU).*]

The choice of the specific Benchmark Solutions that RAWSEEDS will generate to solve its Benchmark Problems is a task that can be done only when the latter are known, i.e. after the work of WorkPackage 4. What can be defined at this preliminary stage, e.g. the generic kind of solutions that the Partners expect to deliver or RAWSEEDS' policy for the management of Intellectual Property Rights (please remember that the users of the RAWSEEDS website will be able to publish on it their own BSs), has been extensively described in Deliverable D1.1. We therefore make reference to Chapter 8 of D1.1 also for this document.

9. Documentation and manuals

RAWSEEDS' website (http://www.rawseeds.org) is online. It acts as the main source for information about RAWSEEDS. In particular, its public section includes:

- a presentation of the project and a description of its objectives;
- explanations of what the RAWSEEDS Benchmarking Toolkit is, of its utility for different kind of users (researchers, companies) and of the way in which users can contribute to RAWSEEDS (i.e. by submitting for publication Benchmark Solutions based on their own algorithms);
- an overview of the workplan of the project;
- a repository for all the public documents produced by the project, such as Deliverables, publications, and presentations;
- detailed information about the hardware used for data gathering, i.e. platforms and sensors;
- a description of the locations of the data gathering sessions.

The website is also used as a repository for the descriptions and manuals of all the hardware and software systems used by RAWSEEDS, both commercial and custom-built.

As soon as the parts of the Benchmarking Toolkit are ready for publication, the section of the website dedicated to their download (and to the upload of user-generated BSs) will be opened. The *Forum* for the discussions between users of the website is already active, as is the F.A.Q. (Frequently Asked Questions) section, but of course they will start to be really useful only after the Benchmarking Toolkit will be published.