Information-based view initialization in visual SLAM with a single omnidirectional camera

David Valiente[a], Maani Ghaffari Jadidi[b], Jaime Valls Miró[b], Arturo Gil[a]
and Oscar Reinoso[a]

[a]*Miguel Hernández University*
System Engineering and Automation Department, 03202, Elche, Spain
Tel:+34 96 665 9005, Fax:+34 96 665 8979
[b]*Faculty of Engineering and IT, University of Technology Sydney, NSW 2007 Australia*

Abstract

This paper presents a novel mechanism to initiate new views within the map building process for an EKF-based visual SLAM (Simultaneous Localization and Mapping) approach using omnidirectional images. In presence of non-linearities, the EKF is very likely to compromise the final estimation. Particularly, the omnidirectional observation model is induces non-linear errors, thus it becomes a potential source of uncertainty. To deal with this issue we propose a novel mechanism for view initialization which accounts for information gain and losses more efficiently. The main outcome of this contribution is the reduction of the map uncertainty and thus the higher consistency of the final estimation. Its basis relies on a Gaussian Process to infer an information distribution model from sensor data. This model represents feature points existence probabilities and their information content analysis leads to the proposed view initialization scheme. To demonstrate the suitability and effectiveness of the approach we present a series of real data experiments conducted with a robot equipped with a camera sensor and map model solely based on omnidirectional views. The results reveal a beneficial reduction on the uncertainty but also on the error in the pose and the map estimate.

Keywords: Visual SLAM; EKF; Omnidirectional images, Gaussian

Email address: dvaliente@umh.es, Maani.Ghaffarijadidi@uts.edu.au, Jaime.VallsMiro@uts.edu.au, arturo.gil@umh.es and o.reinoso@umh.es (David Valiente[a], Maani Ghaffari Jadidi[b], Jaime Valls Miró[b], Arturo Gil[a] and Oscar Reinoso[a])

1. Introduction

The problem of SLAM poses a challenge in the framework of mobile robot applications. It involves a laborious process that deals simultaneously with the mapping and robot's localization. This fact brings a challenge with regards to complexity, as the procedure is expected to work incrementally and to return a coherent representation of the environment. Besides, the existence of noise sources become accountable for undesired effects which aggravate and jeopardize the final estimation.

Lately, visual sensors have reached a great emergence as the main tool for collecting information in the field of SLAM. They represent a promising option compared to classic sensors such as laser or sonar. They allow us to take the best advantage of cameras due to their low cost, light weight and low consumption principally. Nonetheless, their major benefit turns to be their capability to collect a large amount of visual information. Such a quality is especially remarkable in the case of omnidirectional cameras, whose field of view is maximum. Many approaches have exploited this aspect of single cameras by means of visual descriptors to encode 3D visual landmarks [1, 2, 3]. Omnidirectional cameras have also been used within different contexts successfully [4, 5, 6].

Over the last years, great efforts have been made on the study and research of the EKF-based SLAM methods sustained by visual sensors [7, 1, 8, 2, 9, 10]. The main efforts have been concentrated on the position estimation of a 3D visual landmarks set in a common reference system. These approaches are liable to encounter difficulties in assuring the convergence of the solution, particularly in the presence of non-linear errors. Such errors are usually provoked by sensory input. Omnidirectional sensors are significantly susceptible to cause this issue [11], due to its highly non-linear nature. Their correspondent effects tend to affect severely the data association problem in SLAM [12]. Other offline algorithms [13, 14, 15] may be seen as an alternative technique to keep stability under non-linear circumstances for SLAM problems [16]. Within this last group, there are some other authors who take advantage of other iterative optimization techniques embedded in the core of the EKF filter [17, 18].

In this approach we rely on an improved version of the EKF which demonstrates its ability to face these common shortcomings commented above. In

particular, we employ EKF in a non-iterative way. The most relevant characteristic of this approach is the definition of the map by omnidirectional images (denoted as views), which are captured along the path of the robot and stored with their poses and visual descriptors. This idea is closely related to the concept of submap. Here, a reduced set of views constitutes a compact and simpler representation of the environment compared to traditional 3D landmark map models. The main novelty is a new mechanism for the initialization of views within the map building process, aimed at uncertainty reduction. We make use of the visual information provided by the visual sensor data in order to construct an information distribution model which accounts for information gain and losses. This task is carried out by means of a Gaussian Process (GP), which is included within the field of non-parametric Bayesian learning techniques. Application of non-parametric methods, such as GPs have recently proven great enhancements on the mapping tasks within the context of autonomous navigation. Continuous frontier maps are obtained by optimizing the process parameters, which reveal important uncertainty reduction [19, 20]. Therefore, we propose the training of a GP as a tool to establish a bounded uncertainty scheme for our approach. By adopting such a technique, we pursue a positive impact on the uncertainty, which we intend to minimize. As a result, harmful effects that are likely to appear under high uncertainty conditions, such as errors induced by non-linearities and consequently instabilities and convergence difficulties are mitigated. As a consequence, a more robust and consistent map and trajectory estimate is obtained for the visual SLAM problem.

Summarizing, the fundamental aspects and contributions of this approach may be listed as it follows:

- A new view initialization mechanism is presented for the map building process within the problem of EKF-based visual SLAM.

- This strategy accounts for information gain and losses more efficiently.

- Probabilistic representation of features and learning their correlations through Gaussian processes regression.

- Bounding the uncertainty leads to the mitigation of harmful effects induced by non-linearities in the framework of EKF-based visual SLAM.

This section has introduced the scope and it has also given a brief outline of the related work. Next, the rest of this paper has been structured

in the following manner: Section 2 briefly presents the basic theory of an EKF filter within this framework. Section 3 provides a general explanation to our EKF-based visual SLAM approach. Next, Section 4 exposes the key points of this contribution, which is supported by Gaussian Processes and Information theory. Finally, Section 5 shows the results extracted from real data experiments. They are aimed at testing the validity and reliability of this approach in terms of accuracy and robustness, but they are especially seeking the uncertainty reduction, which is obviously translated into an improvement on the solution convergence. Comparison between this proposal and a former SLAM approach has also been included to support these results. Further discussion and conclusions are addressed in Section 6.

2. EKF

The principle of the EKF [21] is based on the iterative update of an augmented state vector which represents the real time estimate to the problem. Considering our specific visual SLAM case, constituted by a view-based map, the estimate returns the pose of the views in the map and the pose of the robot. Then, the state vector can be defined as:

$$\bar{x}(t) = [x_v, x_{l_1}, x_{l_2}, \cdots, x_{l_N}]^T \tag{1}$$

where x_v represents the current pose of the robot and x_{l_N} the pose of the N-th view in the map. Two linear relations are defined by $F(t)$ and $H_i(t)$ so as to encode the dependency between $\bar{x}(t)$ and the observation measurement $z_i(t)$ respectively. In addition, it is essential to bear in mind the information provided by the odometry of the robot $u(t+1)$, the uncorrelated gaussian noise introduced into the system $v(t+1)$, and the noise generated by the sensors, $w_i(t)$, being also gaussian and with covariance $R(t)$.

Three fundamental stages are well differentiated by the EKF to operate. Firstly, a prediction for $\hat{x}(t)$ and $\hat{z}_i(t)$ is proposed. Then the second stage makes use of this prediction to determine the deviation between the prior $\hat{z}_i(t)$ with respect to the real observation $z_i(t)$. This concept is commonly known as the innovation, and its meaning is of paramount significance in the computation of the final solution provided by the filter. Finally, the third stage takes into account the second stage's output to produce the refinement of the estimation obtained during the first stage, seen as an updating step. These three stages may be described by their analytic expressions in the following terms:

- Prediction

$$\hat{x}(t+1|t) = F(t)\hat{x}(t|t) + u(t) \tag{2}$$
$$\hat{z}_i(t+1|t) = H_i(t)\hat{x}(t+1|t) \tag{3}$$
$$P(t+1|t) = F(t)P(t|t)F^T(t) + Q(t) \tag{4}$$

being $P(t|t)$ and $P(t+1|t)$ the covariance matrices which correspond to the uncertainty of the estimation at instants t and $t+1$ respectively. $Q(t)$ is constituted by the noise parameters which characterize the odometry of the wheels of the vehicle.

- Innovation
$$v_i(t+1) = z_i(t+1) - \hat{z}_i(t+1|t) \tag{5}$$
$$S_i(t+1) = H_i(t)P(t+1|t)H_i^T(t) + R_i(t+1) \tag{6}$$

where $S_i(t+1)$ represents the innovation's covariance.

- Update
$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + K_i(t+1)v_i(t+1) \tag{7}$$
$$P(t+1|t+1) = P(t+1|t) - K_i(t+1)S_i(t+1)K_i^T(t+1) \tag{8}$$

being $K_i(t+1)$ the gain matrix of the filter which plays the role of weighting. It is computed in the following manner:

$$K_i(t+1) = P(t+1|t)H_i^T(t)S_i^{-1}(t+1) \tag{9}$$

It is worth noting that $Q(t)$ y $R(t)$ have to be initialized. The noise parameters which characterize the odometry are introduced into $Q(t)$ and the experimental accuracy parameters associated with the visual sensor into $R(t)$. In addition, the odometry $u(t)$ is required as an initial seed for the prediction obtention defined by Eq. (2).

3. Visual SLAM

The main goal of a visual SLAM technique is to retrieve a reliable estimate of the position of the robot inside a certain explored environment, which has to be estimated simultaneously. In this approach, the map is constituted by a set of omnidirectional images acquired at different poses in the environment,

denoted as views. Please note that these views do not correspond with information from any specific physical landmark as it is traditionally understood in vision-based SLAM. They are constituted by an omnidirectional image captured at the pose $x_l = (\mathrm{x}_l, \mathrm{y}_l, \theta_l)$ and a set of points of interest extracted from that image. Such arrangement allows us to take the best advantage of the wider field of view provided by omnidirectional cameras. A large amount of information may be encoded by a simple image, fact that leads to a significant reduction of the required computational resources. In consequence, the compactness of the map is enhanced thanks to the less number of variables needed for estimating the solution.

As briefly mentioned in Section 2, the pose of the mobile robot at time t is expressed as $x_v = (\mathrm{x}_v, \mathrm{y}_v, \theta_v)^T$, and each view n with $n \in [1, \ldots, N]$ is constituted by its pose $x_{l_n} = (\mathrm{x}_l, \mathrm{y}_l, \theta_l)_n^T$, its uncertainty and a set of feature points expressed in image coordinates along with their visual descriptors.

Finally, as shown in Eq. (1), the augmented state vector for our EKF view-based SLAM is:

$$x\bar{(t)} = \begin{bmatrix} x_v & x_{l_1} & x_{l_2} & \cdots & x_{l_N} \end{bmatrix}^T \tag{10}$$

3.1. Map Building

The map building task is depicted in Figure 1. The robot starts exploring the environment at the origin point A, where it captures an omnidirectional image I_A, stored in the map as a view with pose x_{l_A} and being representative of the relevant visual information around this position. Now I_A is assumed to be the first part of the map, then the robot moves towards the first office room. As long as there is not any major obstruction, the robot extracts corresponding points between I_A and the omnidirectional image at its current pose. This fact makes it able to localize itself. However, once the robot enters the office room, the appearance of the images varies significantly, so no matches are found on I_A. At this point, the robot initializes a new view named I_B at the current robot's position x_{l_B}. This view aids the robot in localizing itself inside the office room. Finally, the robot concludes the exploration of the environment with an accumulated map defined by views I_C, I_D, I_E. The number of views initiated in the map directly depends on the sort of environment and its visual appearance. Figure 1 also provides a synthesis of the localization procedure. A comparison between I_A and I_E is presented, where corresponding points and the motion transformation given by the relative angles between the pose of the images are indicated.

It is important that the initialization of a new view in the map is clearly explained. Our former approach [22] relied on a relative measurement between images so as to define an initialization ratio, which was experimentally defined as:

$$A = k \frac{c}{p_1 + p_2} \tag{11}$$

being p_1 and p_2 the feature points detected on each image and c the correspondences between them. The value of k was aimed at weighting the measurement according to the visual appearance of each particular scenario.

Although this strategy assesses the similarity of the environment, it is empirically suited for scenarios where the matching process is feasible and few outliers are introduced. This fact leads us to seek a more reliable and general mechanism. According to the objectives stated in this work, in order to achieve the uncertainty reduction, new contributions and added value have to be implemented at this stage. We suggest an initialization ratio which adapts to information gain and losses. Thus it is based on the current uncertainty of the map. To that end, we propose the use of a regression technique represented by a GP in order to infer a data information distribution which aids in the definition of this ratio. In particular, the robot will initialize a new view whenever there is a high change in the inferred information distribution from the sensor data, in other words, when there exists relevant changes in the environment's visual appearance. As a result, the final estimation will benefit from this idea since any new view will be initiated at an optimum pose in the sense of uncertainty. The arrangement of new views will now assure that the uncertainty of the estimation will be bounded. This proposal reinforces the value that comes along with our view-based approach: major information changes on the environment are encoded by new views in the map. Further explanation and details are given in Section 4

3.2. Observation Model

The observation model has be to designed in accordance to the view-based map model presented above. The goal is to retrieve a motion transformation between images which aids in the robot localization. The versatility of omnidirectional images enables to apply epipolar constraints [23] to extract the observation measurement between two images as shown in Figure 1. The position of these images corresponds to the pose where the robot captured them. To that effect, only two images with a reduced set of corresponding
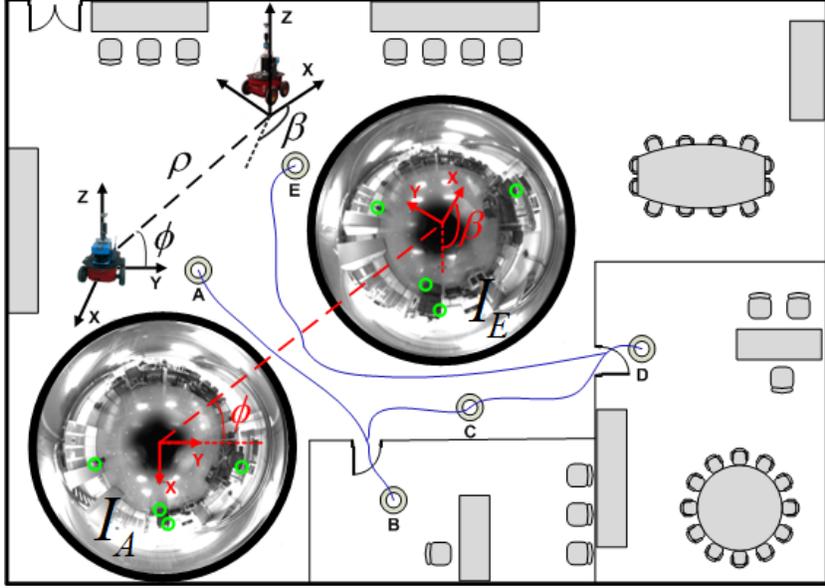
Figure 1: Map building process. First view in the map, I_A, is initiated at the origin A. While the robot traverses the environment, correspondences may be found between I_A and the current image captured at the current robot's pose. In case that no correspondences are found, a new view is initiated as the current image, for instance I_B at B. The procedure finalizes when the entire environment is represented.

points between them are required to obtain the motion transformation. The observation measurement may be expressed as:

$$z_t = \begin{pmatrix} \phi \\ \beta \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{y_{l_n} - y_v}{x_{l_n} - x_v}\right) - \theta_v \\ \theta_{l_n} - \theta_v \end{pmatrix} \tag{12}$$

where ϕ and β are the relative angles which express the bearing and orientation at which the view n is observed from the current robot's pose. Please note that the structure of the view n was presented as $x_{l_n} = (x_l, y_l, \theta_l)_n^T$, whereas the pose of the robot is given as $x_v = (x_v, y_v, \theta_v)^T$. Figure 1 graphically exposes the meaning of these measurements (ϕ, β).

3.3. False correspondences avoidance

Matching correct feature points between images it is crucial for retrieving a reliable and consistent observation model. In this sense, we set some restrictions which prevent from false correspondences appearance. We make use of

epipolar constraints to delimit the search for correspondences [24]. The same point detected in a first camera reference system, denoted as $p = [x, y, z]^T$, may be expressed as $p' = [x', y', z']^T$ in the second camera reference system. Then, the epipolar condition is used to state the relationship between both 3D points p and p' seen from different views.

$$p'^T E p = 0 \tag{13}$$

where the matrix E is the essential matrix and it can be computed from a set of corresponding points in two images.

$$E = \begin{bmatrix} 0 & 0 & \sin(\phi) \\ 0 & 0 & -\cos(\phi) \\ \sin(\beta - \phi) & \cos(\beta - \phi) & 0 \end{bmatrix} \tag{14}$$

being ϕ and β the relative angles that determine a planar motion transformation between two different views, as shown in Figure 1 and Eq. (12). We rely on the information provided by the EKF at its prediction stage to devise a realistic search for valid corresponding points between images. In an idealistic case, the epipolar constraint defined by Eq. (13) should equal a fixed threshold, implying that the epipolar curve defined between images always presents a little static deviation. However, we consider the propagation of uncertainties in the map into Eq. (13) by introducing a dynamic threshold $\delta(\hat{z}_t)$. This implies a more realistic SLAM approach, since this threshold depends on the existing error on the map, which dynamically varies at each step of the SLAM algorithm. This error is correlated with the error on $\hat{z}_t$. Therefore, given two corresponding points between images, they must satisfy:

$$p'^T \hat{E} p < \delta(\hat{z}_t) \tag{15}$$

Finally, as it may be observed in Figure 2, the information provided by a predicted state is used to generate a projection of a normal multi-scale distribution on the sensor frame, where corresponding points are expected to be detected. This projection determines a predicted area in where to search for. The shape of this area depends on the error of the prediction, which is directly correlated with the current uncertainty of the map. This approach not only mitigates the undesired harmful effects associated with false positives, but also simplifies the search for corresponding points between images as it restricts the area where correspondences are expected, instead of a global search along the whole image.
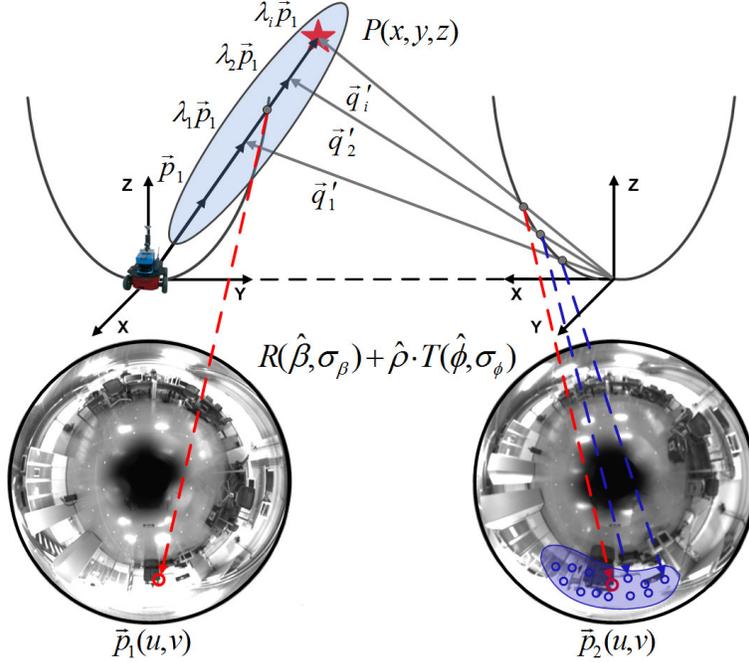
Figure 2: Given a detected point $\vec{p_1}$ in the first image reference system, a point distribution is generated to obtain a set of multi-scale points $\lambda_i \vec{p_1}$. By using the EKF prediction, they can be transformed into $\vec{q_i'}$ on the second image reference system by means of a rotation $R \sim N(\hat{\beta}, \sigma_\beta)$, translation $T \sim N(\hat{\phi}, \sigma_\phi)$ and scale factor $\hat{\rho}$. Finally, $\vec{q_i'}$ are projected into the image plane to determine a restricted area where correspondences have to be found. The circled points represent the projection of the normal point distribution for the multi-scale points that determine this area.

4. Gaussian Processes

A GP has been introduced in this work in order to establish a sensor data distribution, which can be mapped into a global reference system. GPs entail a non-parametric Bayesian regression method, which statistically infer the dependencies between points in a data set [25], in contrast to conventional functions which analytically relate inputs and outputs. A GP, denoted as $f(x)$, is constituted by its mean, $m(x)$, covariance $k(x, x')$, and the training
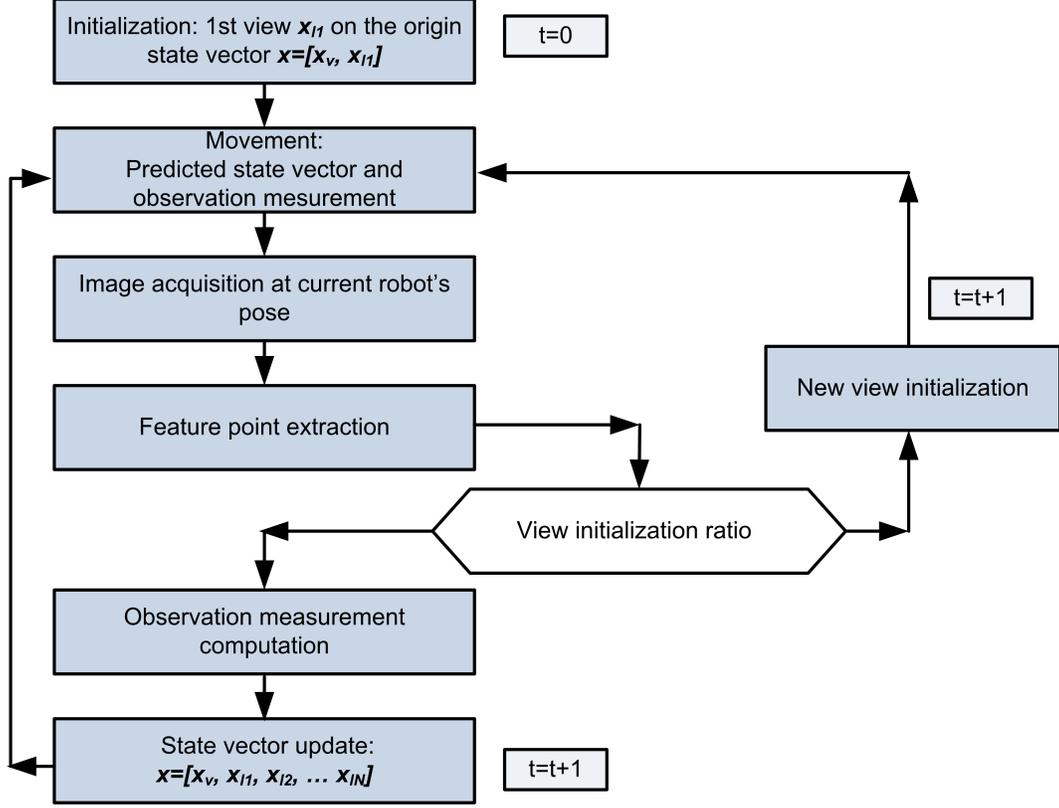
Figure 3: Block diagram summary for the EKF-based visual SLAM approach.

and test input vectors, x and x' respectively.

$$f(x) \sim \mathcal{GP}[m(x), k(x, x')] \tag{16}$$

$$f(x') \sim \mathcal{N}(\mu, \sigma^2) \tag{17}$$

$$\mu = E(f' \mid x, y, x') = k(x', x)[k(x, x) + \sigma_n^2 I]^{-1} y \tag{18}$$

$$\sigma = k(x', x') - k(x', x)[k(x, x) + \sigma_n^2 I]^{-1} k(x, x') \tag{19}$$

being σ_n^2 the variance of the Gaussian observation noise and f' the output values at the test points.

4.1. Sensor data distribution based on GP

Once the GPs basis and formulation have been presented, we are able to devise a model to represent our sensor data information distribution. The

inference procedure through a GP takes the visual information gathered from the environment in the form of feature points detected on the image frame, as it was explained in Section 3.

While the robot navigates, a certain observation measurement is performed at time t. Then, the feature points detected on the image acquired at the current robot's pose are considered as our training data set x_i for the GP. The test points x_i' are determined by sampling uniformly the space defined in a global reference system. Finally, the GP returns the mean values μ_i and variances σ_i^2 inferred for these test points. The most straightforward outcome of $GP's$ output is the probability of existence of a feature point at the locations specified by the test points.

There are several steps involved in the construction of the sensor data information distribution: Firstly, the feature points are locally processed on the camera reference system. Secondly they are back-projected into a global reference system by means of the calibration parameters of the sensor [26]. Next, they become the input to the GP, which returns the probability distribution. Ultimately, when new points are extracted from images acquired at new poses, the distribution is fused into the general information reference system. We can assume that significant variations on the scenario will lead to the detection of new feature points which will cause substantial changes on the information distribution representation, what in the end it is the key point to analyze.

Figure 4 shows an example of the sensor data distribution generated by the GP in terms of probability of existence associated to a bunch of feature points.

Figure 5 presents an illustrative example where the robot explores the environment, meanwhile the visual information varies along the path and so does the feature points and the output returned by the GP. Poses A and B are relatively close so that the scene captured by the robot at these points is quite similar, and thus many feature points are coincident between images. By contrast, when the robot approaches the second room, the appearance of the environment changes substantially. Therefore, at pose C, new feature points are detected with respect to images at poses A and B. Figure 6 provides an extended explanation to this example. Figure 6(a) represents the motion transformation between poses A, B and C. Then Figure 6(b) shows the images acquired at A, B and C respectively. The feature points are projected on the image plane and indicated with crosses. Green crosses mark corresponding feature points between images and blue crosses mark new
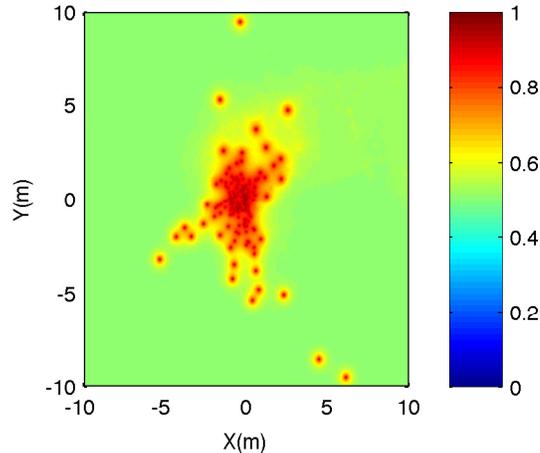
Figure 4: Sensor data information distribution: probability of existence of feature points on the 2D reference system.

feature points detected. The change on the appearance of the environment at C is notable, as it is confirmed by the detection of unseen feature points shown in blue. This fact involves a variation on the information distribution, in Figure 7. Here, the current probability of existence of feature points at poses A, B and C is represented. This figure depicts the evolution of the sensor data information distribution along these poses. Please note that a relevant variation appears between poses B and C. Overlapped areas of high probability denote locations where feature points have been repeatedly observed several times from different poses. This region encapsulates the visual information that has remained steady, being more significant between poses A and B. By contrast, at pose C the information distribution confirms that the weight of this overlapped area decreases versus the areas which represent the probability of existence of new feature points.

As seen in the example, at every new pose the robot traverses, a set of new observation measurements are obtained between its current pose's image and the views stored in the map. The visual information provided from the camera is more likely to overlap at the nearby poses, thus the comparison of images results in a higher number of corresponding points. By contrast, the information varies considerably when the robot discovers unknown areas, and then the corresponding points decrease dramatically.

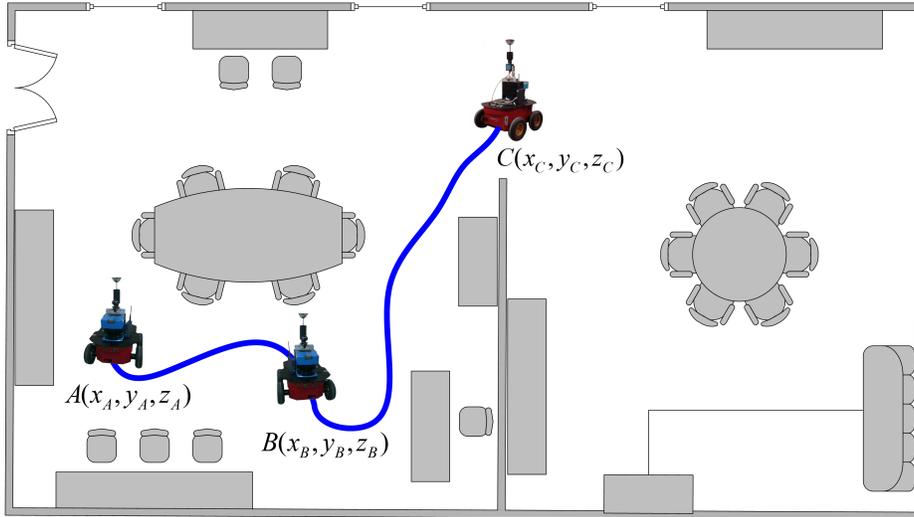A profitable use can be made of this variation of visual information. We

13

Figure 5: Map building process using an omnidirectional sensor. The robot explores the environment while simultaneously initiates image views in the map at poses A, B and C. These images aids in its localization.

propose an efficient map building process in terms of uncertainty. The purpose is to study these variations of visual information in order to decide the initialization of new views in the map. With this assumption, every new view will encode the most relevant changes in the environment, according to their visual informative characteristics. The main expected result is the reduction of the total uncertainty of the estimated map. The placement of the new views will be efficiently selected according to a bounded uncertainty procedure which will be defined by an updated version of the former initialization ratio defined by Eq. (11).

To that end, we adopt the tool given by Kullback-Leibler divergence [27], which is commonly known as Information Gain in probability theory. The aim is to evaluate the fluctuation expected in the entropy when a new sample set is introduced to a certain distribution. In this context, we use the entropy to measure the uncertainty associated to the feature points given by our GP in terms of probability of existence. The Kullback-Leibler divergence (KL) represents the difference of entropy between the information distribution of current feature points, F_1, and the new inferred feature points, F_2, from new images. In other words, the higher value of KL divergence means the newly introduced feature points are less similar. Thus higher is the amount of new
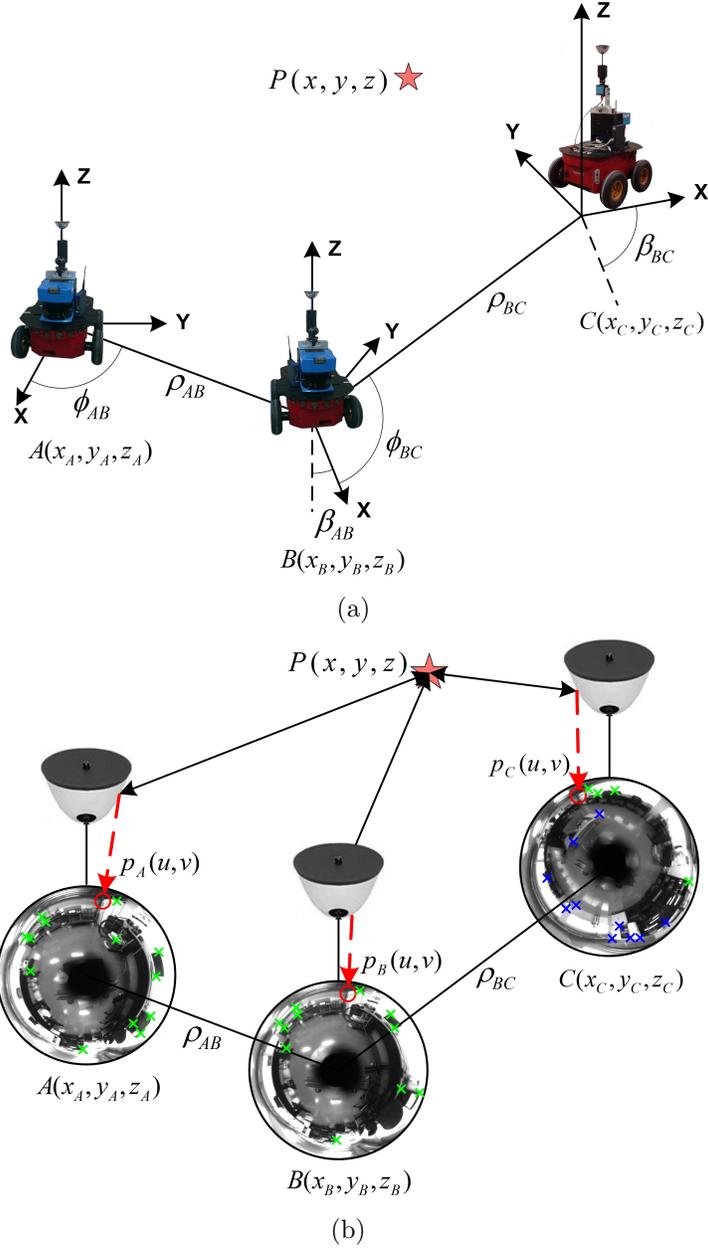
(a)

(b)

Figure 6: Detailed description to example presented in Figure 5: 6(a) represents the motion transformation between poses A, B and C and a 3D point, $P(x, y, z)$; 6(b) shows the images acquired at poses A, B and C, where the projection of $P(x, y, z)$ on every image is indicated by $p_A(u, v)$, $p_B(u, v)$ and $p_C(u, v)$ respectively. Corresponding points between images are shown by green crosses whereas the new feature points are shown by blue crosses.
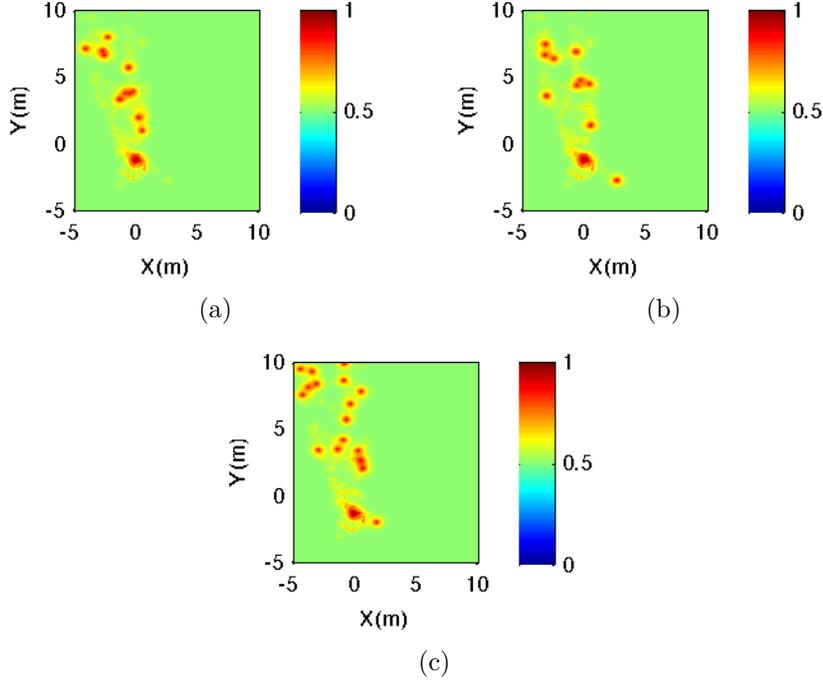
Figure 7: Evolution of the sensor data information distribution along poses shown in the example presented in Figure 5: 7(a) A; 7(b) B; 7(c) C. Variation on the probability of existence of feature points on the 2D reference system.

visual information discovered by the robot. Likewise, the uncertainty on the estimated map will increase as areas of the environment with new visual appearance are being explored by the robot. The following equations define their structure:

$$H(F_1) = -\sum_i F_1(i) \log F_1(i) \qquad (20)$$

$$KL(F_1 \parallel F_2) = H(F_1, F_2) - H(F_1) = \sum_{i=1}^{k} F_1(i) \log \frac{F_1(i)}{F_2(i)} \qquad (21)$$

where $H(F_1)$ is the entropy of the information distribution of the current feature point set at time t. $F_1(i)$ represents the probability of existence of the current i-feature points at time t and so does $F_2(i)$ with new points added at time $t + 1$ from a new image. The relevance of their information contribution is directly proportional to $1/\sigma_i^2$.

16

The strategy to initialize a new view seeks an upper bound for the uncertainty so as to get an efficient map in this sense. To that aim, as we keep the information distribution of the points referred to a global system, we consider the KL value in its accumulative format. Then, we measure the accumulated increases in the uncertainty, given by the addition of new visual information. The previous initialization ratio shown by Eq. (11) has been substituted by the following one:

$$\gamma = \sum_t KL(P_t \parallel P_t + 1) \tag{22}$$

where P_t refers to the data information distribution obtained up to time t and $P_t + 1$ refers to the new data information fused into the global reference system at time $t+1$. Establishing different thresholds for γ leads us to obtain different view initializations and thus different map versions. Obviously, the associated uncertainty also fluctuates differently depending on the placement of the views. A more detailed explanation with real results is presented in the next section.

5. Results

We have performed two different sets of real data experiments in an office-like environment. The equipment consists of a Pioneer P3-AT indoor robot with a 1280×960 firewire camera and a hyperbolic mirror. The optical axis of the camera is installed approximately perpendicular to the ground plane, as it can be seen in Figure 6. We used a SICK LMS range finder in order to compute a ground truth representation. Firstly, in Section 5.1 we present results to examine the behavior of this proposal in terms of its associated uncertainty. Then, in Section 5.2 we show different map solutions obtained with this SLAM approach. All the set of results presented here are also compared with a former SLAM version [22] which does not use GP nor data information distribution in order to initialize views.

5.1. Initialization ratio and sampling resolution

Here we evaluate the mechanism for view initialization under different conditions. The first parameter to consider is the threshold value for the new initialization ratio γ, defined in Eq. (4.1). Intuitively, the less views are initialized in the map when the higher is this ratio. This means that larger changes on the visual environment are encoded by less views, without
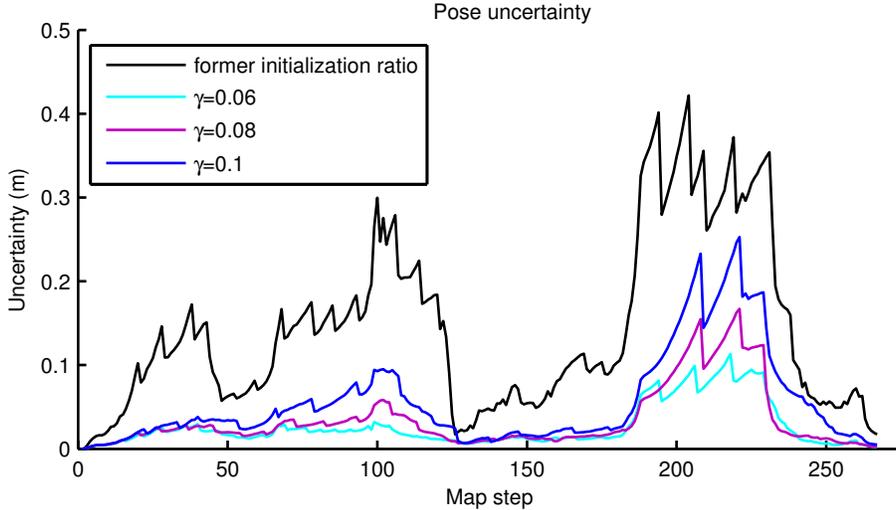
17

Figure 8: Evolution of the uncertainty along the robot's path. Different values for γ are shown and compared to the uncertainty obtained with the initialization ratio (Eq. 11) employed in the former SLAM approach.

implying any new initialization. A real experiment has been conducted in a scenario with dimension $25m \times 25m$. Figure 8 presents the current uncertainty along the path followed by the robot at each time step. It is now confirmed that lower values for the initialization threshold help in the reduction of uncertainty. It is worth noting that the results provided by this proposal outperform the uncertainty associated with the former SLAM approach at every case. The main reason for what this is a more feasible mechanism is the consideration of information gains and losses rather than the amount of feature points matched. Figure 9 shows the mean uncertainty accumulated on the total map at each time step. Obviously, the shape and the evolution is quite similar. Again, the results obtained with this approach confirm a better performance with regards to the uncertainty. This means that the view initialization strategy accomplishes with the proposed scheme for obtaining a bounded uncertainty. Nonetheless, it is worth mentioning that low uncertainty values imply larger number of views in the map, and obviously at a higher computational cost. Hence a trade-off solution is needed, which usually depends on the specific application.

Secondly, it is necessary to state the same analysis but now looking at the estimation error. To that aim, we extract values for the RMS error
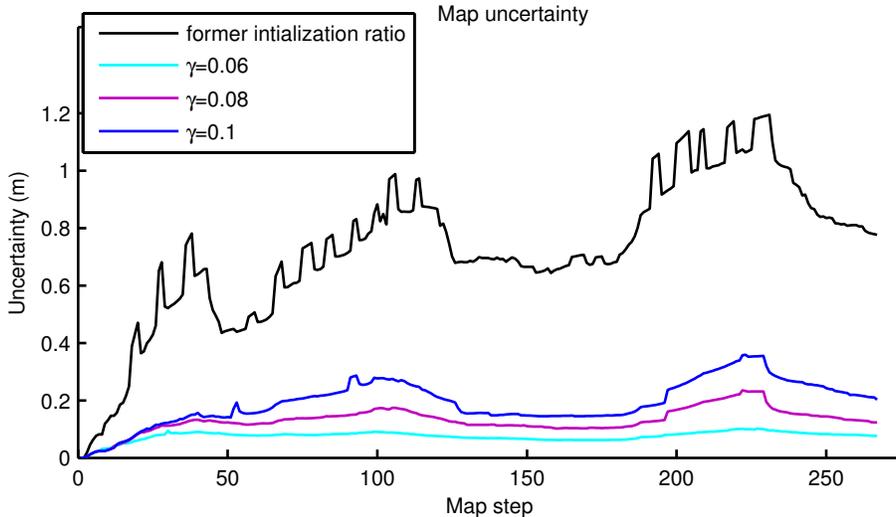
18

Figure 9: Evolution of the mean uncertainty accumulated on the total map. Different values for γ are shown and compared to the uncertainty obtained with the initialization ratio (Eq. 11) employed in the former SLAM approach.

(Root Mean Square). Figure 10 plots RMS values associated to the different initialization ratios γ. Once again, comparing with the former approach, the obtained error with the new approach is lower at any case.

Finally, another parameter which has a considerable importance on the efficiency of this approach is the sampling size for the test points selection. As it was mentioned in Section 4, the global reference system is sampled uniformly by means of these test points x'. Then, the data information distribution inferred by the GP will have a specific resolution which is directly linked with this sampling size determined by x'. Now, Figure 11 represents the RMS (Root Mean Square) error when the sampling size is varied. It can be observed that higher resolutions ensure better results as the probability areas are more accurately determined. However, a high resolution inference from the GP become very expensive in terms of computation. It is worth noting that the dimension of the grid size is up to scale, according to the scale factor of the map.

5.2. Map building based on data information distribution

At this point, a trade-off parameter set can be extracted from the study of the results presented above. Hence, we carry out a complete SLAM exper-
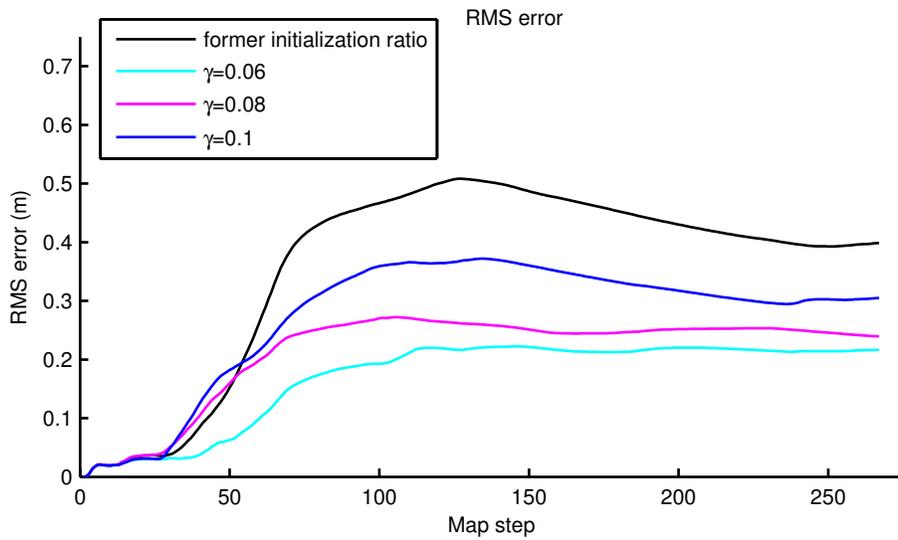
Figure 10: RMS error for different initialization ratios γ. The RMS value obtained with the former SLAM approach has been also plotted for comparison.
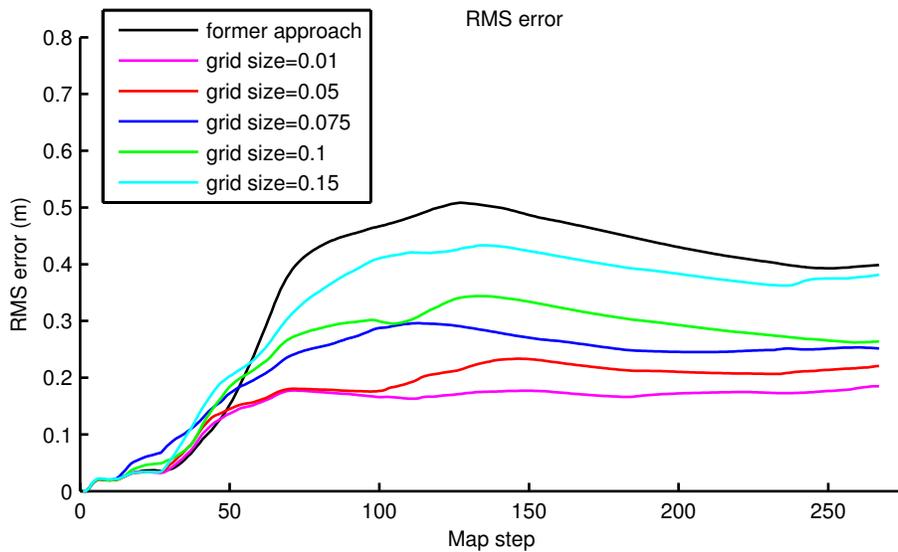

Figure 11: RMS error for different grid size resolutions. The grid size resolutions are expressed up to the scale factor of the current map. The RMS value obtained with the former SLAM approach has been also plotted for comparison.

iment. Figure 12 shows the final map and path estimation for an office-like environment. Figure 13 presents a different solution for the same scenario where a different initialization ratio has been considered. In order to compare and prove the benefits of this proposal, Figure 14 presents results obtained with the former EKF-based SLAM approach [22]. By inspecting Figure 12(a) and Figure 13(a), it can be confirmed that lower initialization ratios ensure a more robust solution with a larger number of views, but obviously at a higher computational cost. Figure 12(b) and Figure 13(b) show the behavior of the error along the path. Both estimations confirm their improvements in comparison with the former approach as seen in Figure 14(a) and Figure 14(b). An important reduction in terms of uncertainty is achieved.

Finally, the method has been used in a larger scenario with the aim of testing its robustness and feasibility in large environments. Figure 17 provides general details of this scenario, which corresponds to an indoor trajectory of $180m$. The areas which the robot goes through consist of office-like rooms, laboratories, corridors and open spaces. The main challenge is to deal with the big changes on the visual appearance between rooms, but with the lighting changes on the images. Some omnidirectional images are also presented, as well as the real path followed by the robot. Figure 16 provides results for this $50m \times 50m$ scenario when using the proposed approach. Again, the estimated path and map reveal their accuracy and similarity to the real path, but also its reduced uncertainty.
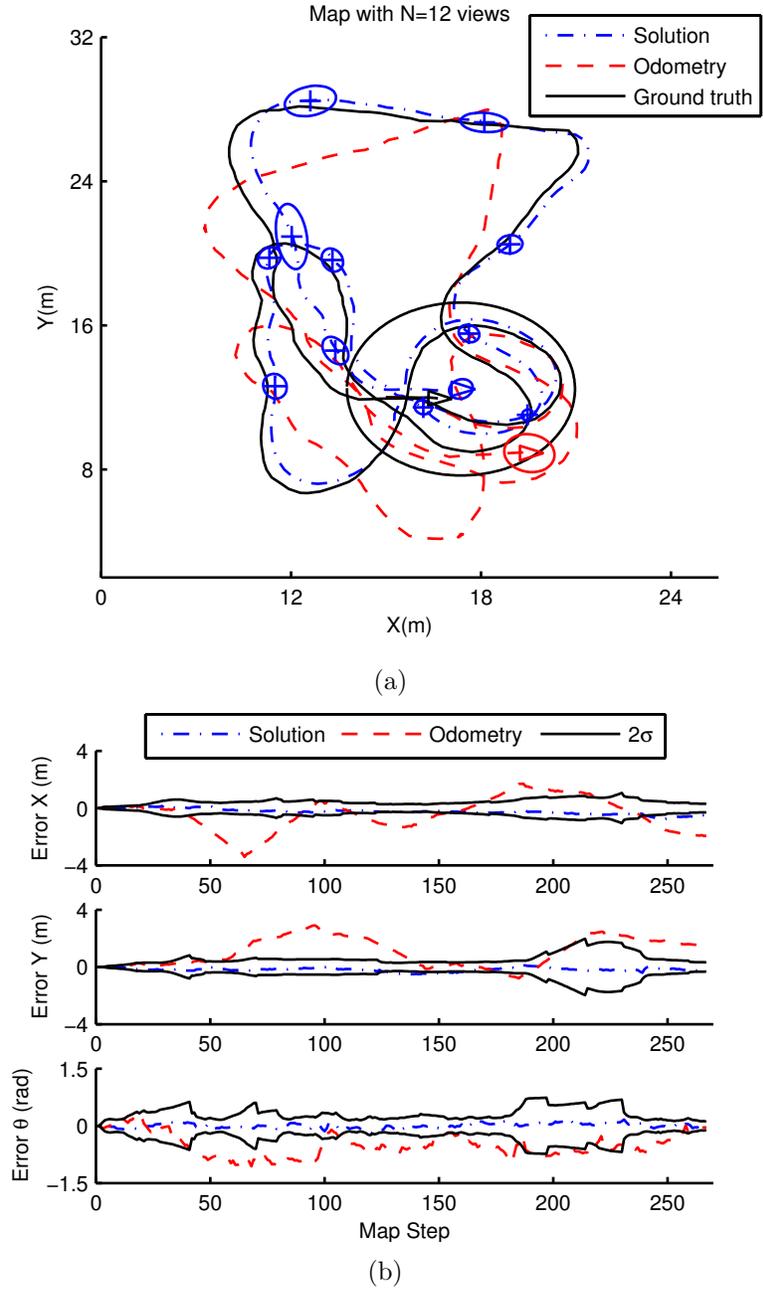
(a)

(b)

Figure 12: 12(a) presents real data results obtained with the presented EKF-based SLAM approach. The map representation of the environment is formed by N=12 views. The position of the views is presented with error ellipses. 12(b) shows the solution and the odometry error in X, Y and θ at each time step.
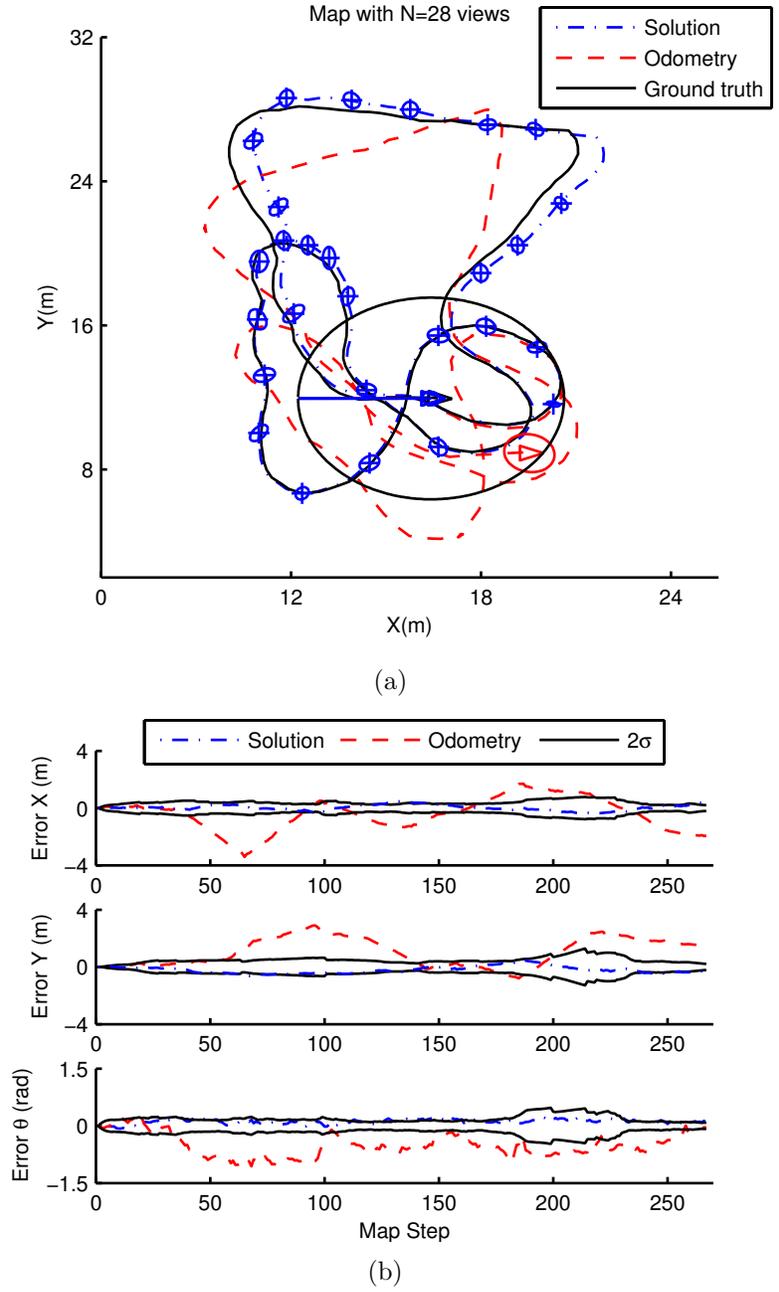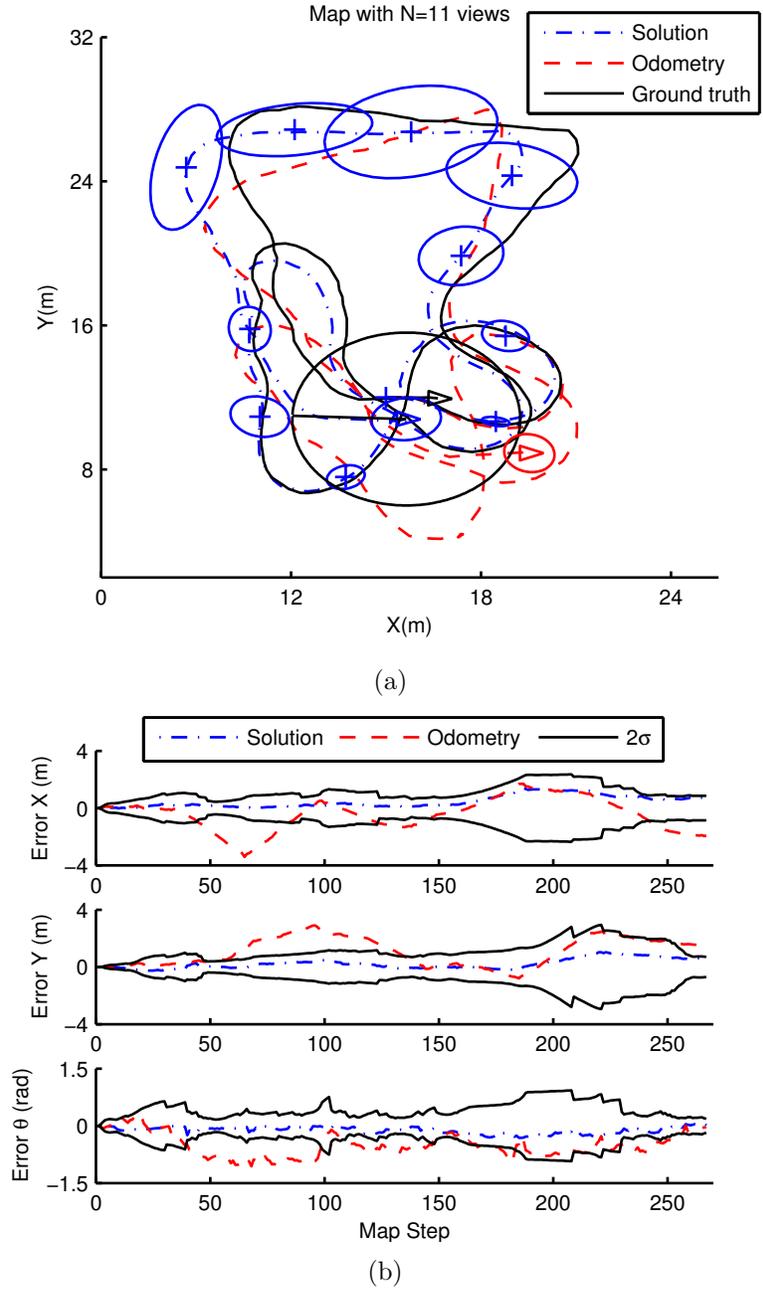
(a)

(b)

Figure 13: 13(a) presents real data results obtained with the presented EKF-based SLAM approach. The map representation of the environment is formed by N=28 views. The position of the views is presented with error ellipses. 13(b) shows the solution and the odometry error in X, Y and θ at each time step.

(a)

(b)

Figure 14: 14(a) presents real data results obtained with the former SLAM approach. The map representation of the environment is formed by N=11 views. The position of the views is presented with error ellipses. 14(b) shows the solution and the odometry error in X, Y and θ at each time step.
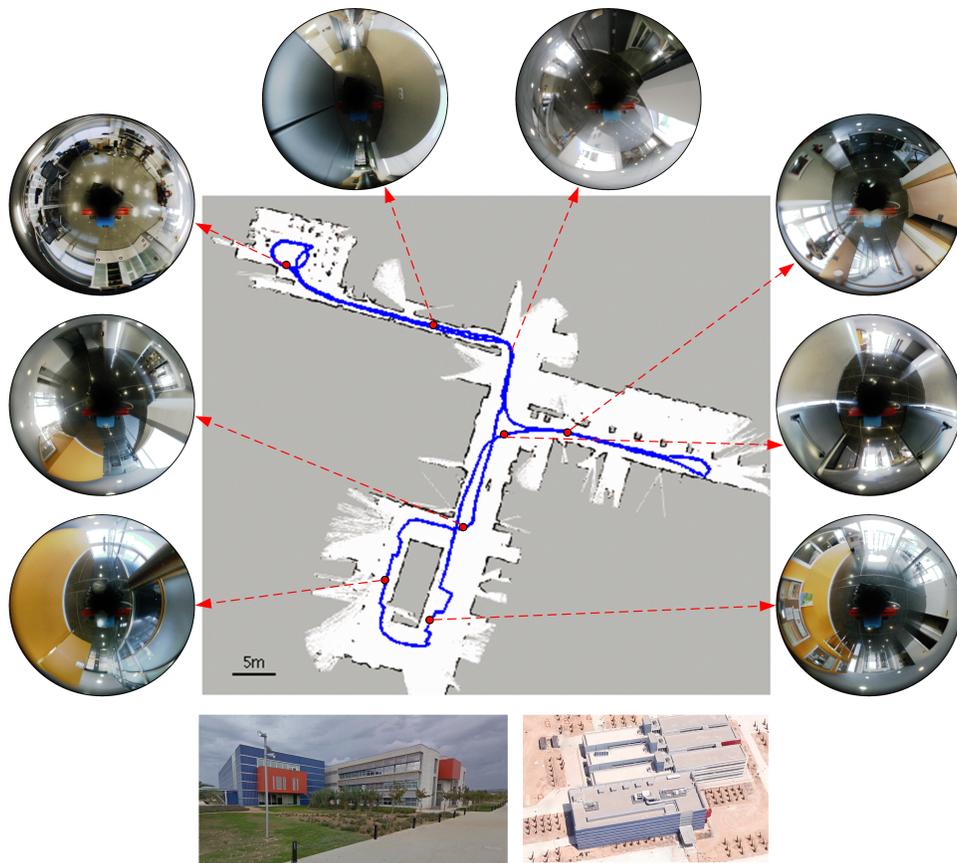
Figure 15: Main details of the large scenario where the last dataset was acquired. The layout of the building, real path followed by the robot and some omnidirectional views of different areas are indicated.

6. Conclusions

This work has presented a novel mechanism for the view initialization within the map building process applied to an EKF-based visual SLAM approach supported by a omnidirectional sensor. This contribution seeks to provide a more feasible mechanism which accounts for information gain and losses so that the harmful effects suffered by visual SLAM approaches are mitigated. Particularly, non-linearities and undesired effects induced in the observation and movement jeopardize the convergence of the estimation provided by traditional EKF-based SLAM approaches. In consequence, it is crucial to keep uncertainty bounded to deal with this issue. To that aim, we
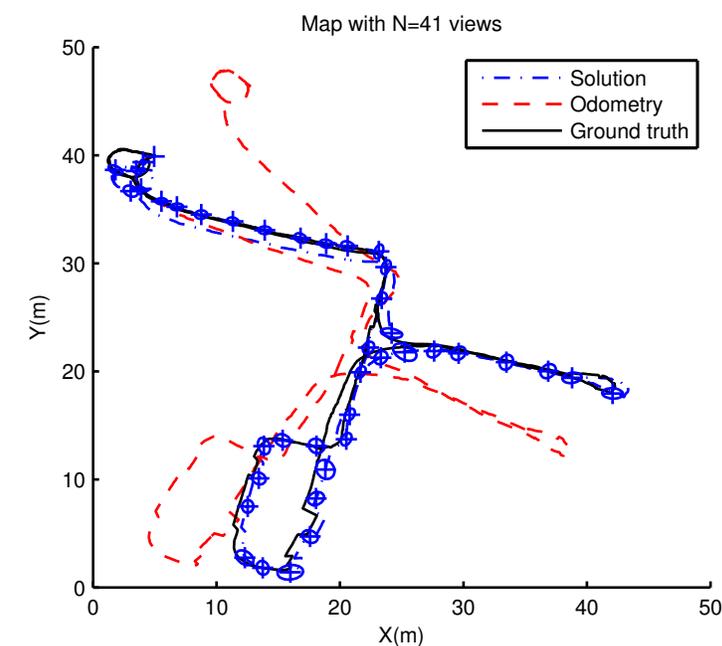
(a)
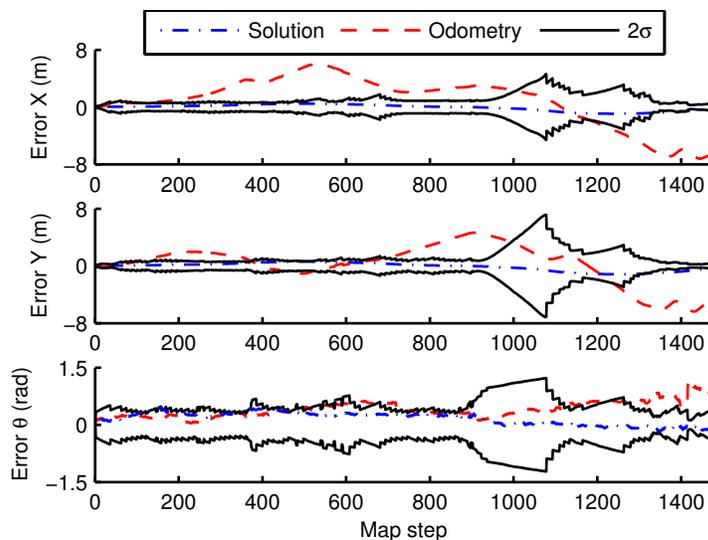
(b)

Figure 16: Real data results obtained with the presented EKF-based SLAM approach in a large scenario. The map representation of the environment is formed by N=41 views. The position of the views is presented with error ellipses. 16(b) shows the solution and the odometry error in X, Y and θ at each time step.
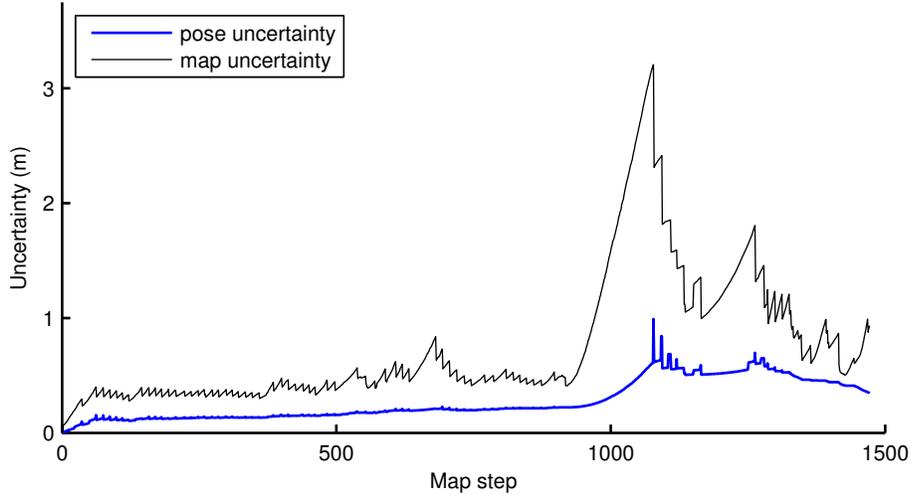
26

Figure 17: Evolution of the uncertainty for the large scenario presented above.

have focused on encoding information gain and losses to define the proposed mechanism to improve the view initialization stage in our EKF-based visual SLAM approach. This new mechanism is achieved by means of the construction of a data information distribution inferred with a Gaussian Process. This distribution represents a probability model for the existence of feature points which is exploited from an informative point of view. Thus an Information Gain method is finally introduced to come up with the intended initialization process, which confirms its capability to bound the uncertainty and to efficiently initiate new views in the map. The results presented have proven the validity of this proposal and the expected benefits regarding with the uncertainty reduction, which is translated into a more robust and consistent map and trajectory estimation. Likewise, the results demonstrate the effectiveness of this approach to set limits to the estimation error as well. In order to reinforce the value of these results and the contributions made on this research, we have also compared them with the results obtained by a former EKF-based SLAM approach which uses a more empirical initialization mechanism.

7. Acknowledgements

References

[1] A. Gil, O. Reinoso, M. Ballesta, M. Juliá, L. Payá, Estimation of visual maps with a robot network equipped with vision sensors, Sensors 10 (2010) 5209–5232.

[2] J. Civera, A. J. Davison, J. M. Martínez Montiel, Inverse depth parametrization for monocular SLAM, IEEE Transactions on Robotics 24 (2008) 932–945.

[3] C. Joly, P. Rives, Bearing-only SAM using a minimal inverse depth parametrization, in: Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO), Vol. 2, Funchal, Madeira, Portugal, 15-18 June 2010, pp. 281–288.

[4] S.-E. Yu, D. Kim, Image-based homing navigation with landmark arrangement matching, Information Sciences 181 (2011) 3427–3442.

[5] Y. Rasmussen, Y. Lu, M. Kocamaz, Integrating stereo structure for omnidirectional trail following, in: Proceedings of the International Conference on Intelligent Robots and Systems (IROS), San Francisco, USA, 25-30 September 2011, pp. 4084–4090.

[6] L. Paya, F. Amoros, L. Fernandez, O. Reinoso, Performance of global-appearance descriptors in map building and localization using omnidirectional vision, Sensors 14 (2014) 3033–3064.

[7] A. J. Davison, D. M. Murray, Simultaneous localisation and map-building using active vision, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24 (2002) 865–880.

[8] A. J. Davison, Y. Gonzalez Cid, N. Kita, Real-time 3D SLAM with wide-angle vision, in: Proceedings of the 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, Lisbon, Portugal, 2004, pp. 117–124.

[9] S. Park, S. Kim, M. Park, S.-K. Park, Vision-based global localization for mobile robots with hybrid maps of objects and spatial layouts, Information Sciences 179 (2009) 4174–4198.

[10] E. Guerra, R. Munguia, A. Grau, Monocular SLAM for autonomous robots with enhanced features initialization, Sensors 14 (2014) 6317–6337.

[11] D. Valiente, A. Gil, L. Fernández, O. Reinoso, View-based maps using omnidirectional images, in: Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO), Vol. 2, Rome, Italy, 28-31 July 2012, pp. 48–57.

[12] J. Neira, J. D. Tardós, Data association in stochastic mapping using the joint compatibility test, IEEE Transactions on Robotics and Automation 17 (2001) 890–897.

[13] R. Yang, H. Gao, P. Shi, Delay-dependent robust H_∞ control for uncertain stochastic time-delay systems, International Journal of Robust and Nonlinear Control 20 (2010) 1852–1865.

[14] L. Wu, D. W. C. Ho, Fuzzy filter design for Itô stochastic systems with application to sensor fault detection, IEEE Transactions on Fuzzy Systems 17 (2009) 233–242.

[15] Y.-T. Sun, C.-H. Wang, C.-C. Chang, Switching t-s fuzzy model-based guaranteed cost control for two-wheeled mobile robots, International Journal of Innovative Computing, Information and Control 8 (2012) 3015–3028.

[16] D. Valiente, A. Gil, L. Fernandez, O. Reinoso, A modified stochastic gradient descent algorithm for view-based slam using omnidirectional images, Information Sciences 279 (2014) 326–337.

[17] R. Kümmerle, B. Steder, C. Dornhege, A. Kleiner, G. Grisetti, W. Burgard, Large scale graph-based slam using aerial images as prior information, Autonomous Robots 30 (2011) 25–39.

[18] E. Einhorna, C. Schrötera, H. Grossa, Attention-driven monocular scene reconstruction for obstacle detection, robot navigation and map building, Robotics and Autonomous Systems 59 (2011) 296309.

[19] M. Ghaffari Jadidi, J. Valls Miro, R. Valencia, J. Andrade-Cetto, Exploration on continuous gaussian process frontier maps, in: Proceedings of the International Conference on Robotics and Automation (ICRA), Hong Kong, China, May 31 2014-June 7 2014, pp. 6077 – 6082.

[20] M. Ghaffari Jadidi, J. Valls Miro, R. Valencia, J. Andrade-Cetto, G. Dissanayake, Exploration in information distribution maps, in: RSS'13 Workshop on Robotic Exploration, Monitoring, and Information Content, Berlin, Germany, 24-28 June 2013, pp. 1–8.

[21] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, Journal of Basic Engineering 83 (1961) 95–107.

[22] D. Valiente, A. Gil, L. Fernandez, O. Reinoso, A comparison of EKF and SGD applied to a view-based slam approach with omnidirectional images, Robotics and Autonomous Systems 62 (2014) 108–119.

[23] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.

[24] D. Scaramuzza, Performance evaluation of 1-point RANSAC visual odometry, Journal of Field Robotics 28 (2011) 792–811.

[25] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning series, Massachusetts Institute of Technology, 2006.

[26] D. Scaramuzza, A. Martinelli, R. Siegwart, A toolbox for easily calibrating omnidirectional cameras, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), China, 9-15 October 2006, pp. 5695 – 5701.

[27] S. Kulback, R. A. Leiber, On information and sufficiency, Annals of Mathematical Statistics 22 (1951) 79–86.