ARTICLE IN PRESS

Robotics and Autonomous Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images

David Valiente *, Arturo Gil *, Lorenzo Fernández, Óscar Reinoso

Miguel Hernández University, Systems Engineering and Automation Department, 03202, Elche, Spain

HIGHLIGHTS

- Proposal to overcome the influence of the non-linear errors on traditional visual SLAM methods.
- We focus on a highly non-linear observation model: the omnidirectional.
- Comparison of traditional filters like EKF, versus SGD.
- Compact map representation, consisting of a reduced set of omnidirectional views.
- We compare accuracy, robustness against errors and speed of convergence.

ARTICLE INFO

Article history:
Received 4 July 2013
Received in revised form
6 November 2013
Accepted 22 November 2013
Available online xxxx

Keywords:
Visual SLAM
SLAM algorithm
EKF
SGD
Omnidirectional images

ABSTRACT

The problem of Simultaneous Localization and Mapping (SLAM) is essential in mobile robotics. The obtention of a feasible map of the environment poses a complex challenge, since the presence of noise arises as a major problem which may gravely affect the estimated solution. Consequently, a SLAM algorithm has to cope with this issue but also with the data association problem. The Extended Kalman Filter (EKF) is one of the most traditionally implemented algorithms in visual SLAM. It linearizes the movement and the observation model to provide an effective online estimation. This solution is highly sensitive to non-linear observation models as it is the omnidirectional visual model. The Stochastic Gradient Descent (SGD) emerges in this work as an offline alternative to minimize the non-linear effects which deteriorate and compromise the convergence of traditional estimators. This paper compares both methods applied to the same approach: a navigation robot supported by an efficient map model, established by a reduced set of omnidirectional image views. We present a series of real data experiments to assess the behavior and effectiveness of both methods in terms of accuracy, robustness against errors and speed of convergence.

1. Introduction

The solution of the SLAM problem is vital for most applications in the field of mobile robotics, for example in navigation tasks. A reliable map representation of the environment has to be built dynamically, in an incremental manner, meanwhile the mobile vehicle requires an appropriated localization inside it, which has to be calculated simultaneously. This fact poses a challenge for the SLAM techniques, since this process involves a notable complexity. The appearance of noise arises as a severe problem, which highly aggravates the achievement of a valid estimation to the problem.

Different SLAM approaches may be classified according to aspects such as the representation of the map, the solver algorithm to compute a solution and the kind of sensor which gathers information of the environment. For instance, the utilization of a laser range sensor [1] has been extensively applied to the obtention of map representations. In this area, two kinds of map representations were principally generated: 2D occupancy grid maps [2] based on raw laser, and 2D landmark-based maps [3] focused on the extraction of features, which were described thanks to laser data measurements. An interesting comparison of both representations is provided in [4].

Nowadays, the emergence of visual sensors has made the tendency to turn into the utilization of digital cameras as the main sensor to gather information. A huge number of applications benefit from the use of these sensors, whose characteristics outperform preceding sensors such as laser, in the sense of the amount of available information. In contrast to laser data sensors, vision sensors provide a wide amount of information of the scene, being as well less expensive, lighter and more efficient in terms of consumption at the price of needing a computational cost to obtain profitable information to build the map. The extraction

* Corresponding authors. Tel.: +34 96 665 9005; fax: +34 96 665 8979.
E-mail addresses: dvaliente@umh.es (D. Valiente), arturo.gil@umh.es (A. Gil), l.fernandez@umh.es (L. Fernández), o.reinoso@umh.es (Ó. Reinoso).

0921-8890/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2013.11.009

of significant feature points has been a procedure widely used in order to encode the visual information. Diverse arrangements are commonly known by their configuration in reference to the number of cameras they consist of. For instance, approaches which utilize two calibrated cameras, known as stereo-pair, in order to extract a set of 3D visual landmarks determined by a visual description [5]. Other approaches simply exploit a single camera to estimate 3D visual landmarks [6,7]. They initialize the coordinates of each 3D landmark by relying on an inverse depth parametrization, since there exists a scale uncertainty on the distance to each landmark, which cannot be directly calculated by using only a single image. Omnidirectional cameras have also been used solely [8], and even some others have arranged two omnidirectional images [9], following the line stated by stereo-based, but pursuing the major advantage associated with the wider field of view provided by omnidirectional cameras.

The estimator algorithm for a SLAM scheme has to be considered as important as the kind of sensor and the map representation. It represents the core of the system, since it is responsible for the ultimate solution. Amongst the most widely used online methods deserving to be highlighted are the EKF [10] and the Rao-Blackwellized particle filters [3,11]. Regarding the offline algorithms, one of the most effective is SGD [12].

Therefore, the correct balance in the combination of data sensors, map representation and kind of algorithm, eventually determines the effectiveness of a SLAM approach which pursues reliability and suitability for realistic applications. Great efforts have been made in this field. For example, certain approaches [10,5,13,6,14] have concentrated on the estimation of the position of a set of 3D visual landmarks in a main reference system, while dealing with the obtention of the map simultaneously. Their principle of working lays on the capability of an EKF filter to converge the estimation to an appropriate solution for the SLAM problem. In [15], an EKF algorithm also supports an approach which proposes a distinctive map representation, consisting of a reduced set of image views. These views are determined by their position and orientation in the environment. Such technique establishes an estimation of a state vector which includes the map and the current localization of the robot at each timestep.

The methods based on EKF are generally liable to become troublesome when dealing with external errors. This issue is directly deduced from the linearization of variables carried out by the EKF. In this sense, such difficulties compromise the proper convergence of the estimation. This situation normally appears in presence of gaussian noise introduced by the observation measurement, fact that usually causes injurious data association problems [16]. A visual observation model as in the case of the omnidirectional model, is susceptible to introduce non-linearities and thus it is responsible for those kind of errors. On the contrary, an offline algorithm such as SGD [17] provides more robustness to face this issue. It is worth mentioning that the vanilla SGD approach has been modified in this work to deal with omnidirectional geometry as well as with the associated observation model. Traditionally, every odometry and observation measurements are processed in an independent manner. Nevertheless, with the aim of finding a valid solution quickly, we have designed a strategy based on the simultaneous usage of a certain set of observation measurements. This proposal might seem to be likely to cause an increase of the required computational resources. However, we have concentrated on the prevention of such effect by updating several stages of the SGD's iterative optimization. According to this, some amendments have been performed so as to accomplish the avoidance of possible harmful bottleneck handicaps.

Hence, the main goal of this paper is to provide with results which help analyze the behavior of both EKF and SGD applied to a view-based SLAM approach. As it can be inferred, the solution's convergence is not trivial with EKF, neither with SGD, especially when the nature of the observation measurement is up to a scale factor. The results extracted from the experiments are intended to assess the capability of both methods to maintain a feasible estimation under different conditions. Estimation accuracy, robustness and convergence of the estimation and speed of convergence will be the most important terms to evaluate.

The structure of the paper has been divided as it follows: Section 2 introduces the most important aspects of the visual SLAM approach proposed here. The EKF principles are detailed in Section 3. Then, Section 4 concentrates on the SGD's specifications. Next, Section 5 provides a series of experiments in order to extract real data results. Finally, Section 6 pursues the analysis of the results and the discussion.

2. SLAM

The main purpose of a visual SLAM scheme is to retrieve a reliable representation of the environment explored by the robot, as well as the position of this vehicle. In this approach, the map of the environment is defined by a set of omnidirectional images acquired from different poses of the robot along the environment, denoted as views. These views do not express information about any physical landmarks as it is traditionally in the field of vision-based SLAM. By contrast, a view consists of a single omnidirectional image captured at a certain pose of the robot $x_l = (\mathrm{x}_l, \mathrm{y}_l, \theta_l)$ and a set of interest points extracted from that image. In accordance with the large field of view provided by omnidirectional images, such arrangement allows us to exploit this capability to gather a large amount of information of the scene in a single image. Thus, a highly notable reduction in terms of number of variables to estimate the solution is achieved.

The position of the mobile robot is denoted as:

$$x_v = (\mathrm{x}_v, \mathrm{y}_v, \theta_v)^T. \tag{1}$$

Each view n with $n \in [1, \ldots, N]$ is constituted by its pose:

$$x_{l_n} = (\mathrm{x}_l, \mathrm{y}_l, \theta_l)_n^T \tag{2}$$

together with its uncertainty P_{l_n} and a set of M interest points p_j, expressed in image coordinates. Each point is associated with a visual descriptor $d_j, j = 1, \ldots, M$.

Therefore, these are the variables which compose the augmented state vector:

$$\bar{x} = \begin{bmatrix} x_v & x_{l_1} & x_{l_2} & \cdots & x_{l_N} \end{bmatrix}^T. \tag{3}$$

2.1. Map building

The process of map building may be clearly understood by inspecting an example in Fig. 1. It shows the exploration procedure carried out by a robot, which starts its navigation of the environment at the origin A. At this moment, capturing an omnidirectional image I_A is required to determine the first view of the map. This view is associated with the pose x_{l_A} and it encodes the relevant information of the local area around this pose. Then, the robot moves towards the first office room. Assuming that the robot does not find any major obstruction, it will be capable of extracting correspondences between I_A and the omnidirectional image referred to the pose where it currently moves through. This procedure makes it able to localize itself. Once the robot enters in the office room, the appearance of the images vary significantly, thus, no matches are found between the current image and image I_A. In this case, the robot will initialize a new view into the map I_B at the current robot position x_{l_B}. Now, this view will facilitate the localization of the vehicle inside this office room. Finally, the

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

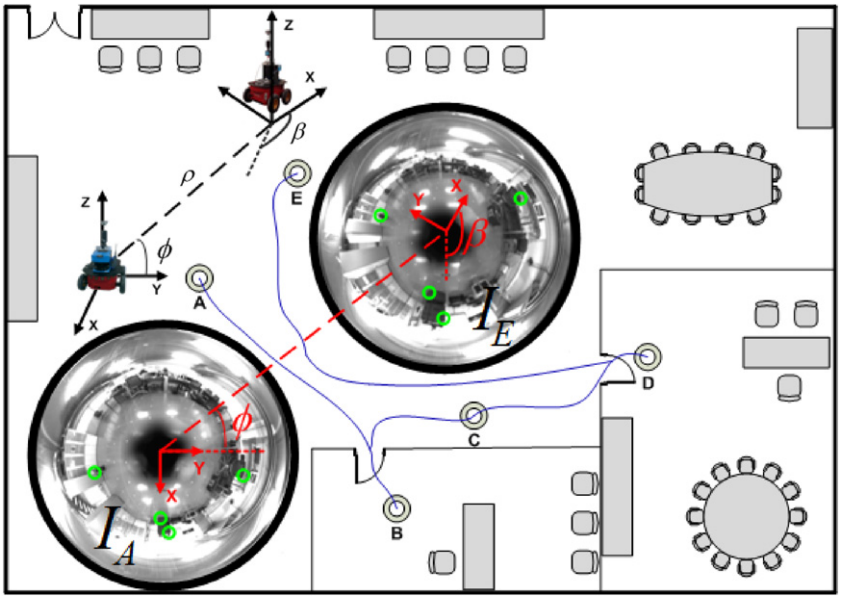

Fig. 1. Map building process. Origin is set at *A*, where a first view I_A is initiated into the map. While the robot traverses the environment, correspondences may be found between I_A and the current image captured at the current robot's pose. In case that no correspondences are found, a new view is initiated as the current image, for instance I_B at *B*. The procedure finalizes when the entire environment is represented.

robot concludes the exploration of the environment by successfully achieving a well-defined trajectory and a map representation of the different areas. As it may be seen, it has been necessary to acquire a set of views I_C, I_D, I_E to complete the final map. The size of the map in terms of the number of views initiated, directly depends on the specific appearance of the environment. Fig. 1 also depicts how the robot accomplishes the computation of its localization, by which it eventually obtains two relative angles thanks to the processing of the information provided by I_A and I_E.

The relative appearance between images is determined by a specific ratio, which it has been experimentally defined as:

$$A = k\frac{c}{p_1 + p_2} \tag{4}$$

where p_1 and p_2 are the interest points detected on each image and c are the corresponding points found between them. The value of k has also been experimentally determined according to the visual appearance of the environment. The ratio A represents a measure of similarity and it is the factor which ease the robot to decide whether to initialize a new view in the map. In particular, the robot will initialize a new view whenever the ratio A drops a certain threshold.

2.2. Data association

The data association problem is posed in the following way: given a set of observations $z_t = [z_{t_1}, \ldots, z_{t_B}]$ at each t, the views which generate each observation have to be discerned. In the approach presented here, the data association process is tackled through the computation of the appearance ratio A. First, we select a subset of candidate views from the map, based on the euclidean distance between the current pose of the robot and the position of each candidate, $D_n = \sqrt{(x_v - x_{l_n})^T (x_v - x_{l_n})}$. The maximum observation range of the robot is established as the maximum distance at which any view can be observed at each t. Then we extract corresponding points between the image acquired at the current pose of the robot and the rest of the candidate views. This allows to find the view which provides the maximum appearance ratio A, defined in (4), which will eventually be chosen as the data association. The view with maximum A reveals the highest similarity with the current image. However, if none of the candidate views provide a value for A higher than a predefined threshold, this will mean that the appearance of the current image of the robot differs substantially from the set of candidate views. Therefore it will be necessary to initialize a new view into the map at the current robot's position.

2.3. Observation model

In consequence with the view-based representation, the formulation of a new observation model is required. The intention is to retrieve a motion transformation between two poses. As observed in Fig. 1 a comparison involving two images provides a motion transformation between two poses. In fact these poses represent the positions where the robot acquired these two specific images. To that effect, only two images with a set of corresponding points between them are required to obtain the transformation. So that the observation measurement may be expressed as:

$$z_t = \begin{pmatrix} \phi \\ \beta \end{pmatrix} = \begin{pmatrix} \arctan\left(\dfrac{y_{l_n} - y_v}{x_{l_n} - x_v}\right) - \theta_v \\ \theta_{l_n} - \theta_v \end{pmatrix} \tag{5}$$

where ϕ and β are the relative angles which express the bearing and orientation at which the view i is observed. Please notice that the structure of the view i follows (2), whereas the pose of the robot is described in (1). Both measurements (ϕ, β) are shown in Fig. 1. Please note that the feature point detector chosen is SURF [18] due to its success and robustness when working with omnidirectional images [19].

3. EKF

The EKF [20] is the first algorithm which has been considered in this work to be applied to the case of visual SLAM with the intention of generating a valid estimation for the problem.

The basis of this filter lays on the estimation of the augmented state vector which is constantly updated in real time. In this framework of a view-based representation, the variables to estimate are the map itself, consisting of views and their poses, and

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

the pose of the robot inside it. Hence the state vector defined in (3) can be adapted to introduce t:

$$\bar{x}(t) = [x_v, x_{l_1}, x_{l_2}, \cdots, x_{l_N}]^T. \tag{6}$$

Once the state vector is defined, the transformation relation between $\bar{x}(t)$ and $\bar{x}(t+1)$ is:

$$\bar{x}(t+1) = F(t)\bar{x}(t) + u(t+1) + v(t+1) \tag{7}$$

where $F(t)$ contains the information pertinent to the transition between states, $u(t+1)$ is the vector related to the movement generated by the odometry of the wheels of the robot, and $v(t+1)$ represents the noise introduced in the system, which has gaussian uncorrelated nature.

Similarly, a linear relation may be defined so as to connect the observation measurement $z_i(t)$ with the current state vector:

$$z_i(t) = H_i(t)\bar{x}(t) + w_i(t) \tag{8}$$

where $H_i(t)$ encodes the relation between $\bar{x}(t)$ and $z_i(t)$. Here, $w_i(t)$ represents the random noise generated by the sensors, which is gaussian and with covariance $R(t)$.

Then, the filter's procedure has to be divided into three fundamental stages well differentiated. Firstly, a prediction of the state $\hat{x}(t)$ is carried out, and based on it, a prediction for the observation measurement $\hat{z}_i(t)$ is also proposed in the following terms:

$$\hat{x}(t+1|t) = F(t)\hat{x}(t|t) + u(t) \tag{9}$$

$$\hat{z}_i(t+1|t) = H_i(t)\hat{x}(t+1|t) \tag{10}$$

$$P(t+1|t) = F(t)P(t|t)F^T(t) + Q(t) \tag{11}$$

where $P(t|t)$ and $P(t+1|t)$ are the covariance matrices which represent the uncertainty of the estimation at instants t and $t+1$ respectively.

The second stage performs the real observation $z_i(t)$ at the current instant t, of a specific view i of the map. Now the concept of innovation has to be introduced to explain the deviation between the prior prediction $\hat{z}_i(t)$ and the current measurement $z_i(t)$:

$$v_i(t+1) = z_i(t+1) - \hat{z}_i(t+1|t) \tag{12}$$

$$S_i(t+1) = H_i(t)P(t+1|t)H_i^T(t) + R_i(t+1) \tag{13}$$

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

Finally, the third stage takes into account the refinement of the estimation obtained during the first stage, seen as an updating step. The value of the innovation is significantly relevant in the computation of the final solution provided by the filter. This solution estimation at instant $t+1$, is finally obtained as:

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + K_i(t+1)v_i(t+1) \tag{14}$$

$$P(t+1|t+1) = P(t+1|t) - K_i(t+1)S_i(t+1)K_i^T(t+1) \tag{15}$$

where in this case $K_i(t+1)$ plays a role of weighting, and corresponds to the gain of the EKF. It is calculated in the following manner:

$$K_i(t+1) = P(t+1|t)H_i^T(t)S_i^{-1}(t+1). \tag{16}$$

It is worth mentioning that the matrices referred to the noise's covariance $Q(t)$ y $R(t)$ have to be initialized. $Q(t)$ is established by means of the noise parameters which characterize the odometry of the wheels of the vehicle. On the other hand, $R(t)$ is determined by experimental accuracy thresholds associated with the visual sensor. The odometry $u(t)$ is required as an initial seed for the prediction obtention, together with the previous state, as deduced from (9). The uncertainty matrix of the map, $P(t)$, considers the noise introduced by the odometry in the form presented in (11), and the noise introduced by the visual sensor when carrying out an observation measurement, as detailed in (13) and (15).

3.1. Correspondence of interest points

With the aim of obtaining a set of feasible correspondences between two views, some restrictions have to be taken into account. Considering the use of epipolar constraints is generally agreed to delimit the search for correspondences [21]. 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.

$$p'^T E p = 0 \tag{17}$$

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

$$E = \begin{bmatrix} 0 & 0 & \sin(\phi) \\ 0 & 0 & -\cos(\phi) \\ \sin(\beta-\phi) & \cos(\beta-\phi) & 0 \end{bmatrix} \tag{18}$$

being ϕ and β the relative angles that determine a planar motion transformation between two different views, as shown in Fig. 1 and (5).

The avoidance of false correspondences has been studied extensively so as to mitigate bad effects on the final estimation for the SLAM problem. Techniques such as RANSAC and Histogram voting have been widely used, and mainly applied to visual odometry approaches [21]. Together with the epipolar constraint (17), they reveal good results in the achievement of false positive rejection. In such context of visual odometry, consecutive images are close enough to disregard high errors in the pose from where images were taken, so that the epipolar constraint is highly likely to be satisfied. Nevertheless, concentrating on the framework of our SLAM problem, the accumulative uncertainties are substantially higher, either in the pose of the robot or in the pose of the views which compose the map. This fact requires to define a reliable strategy to accomplish with a correct data association. We rely on the information provided by the predicted state vector extracted from the Kalman filter, by which we are able to obtain a predicted observation measurement $\hat{z}_t$, as stated in (5). Then it is also necessary to consider the current map uncertainties so as to deal with a realistic search for valid corresponding points between images. The map uncertainties are propagated in accordance with (17) by introducing a dynamic threshold δ. In an idealistic case, the epipolar constraint may equal a fixed threshold, implying that the epipolar curve defined between images always presents a little static deviation. On the contrary, a realistic SLAM approach, should consider that this threshold depends on the existing error on the map, which dynamically varies at each step of the SLAM algorithm. Since this error is correlated with the error on $\hat{z}_t$, we rename δ as $\delta(\hat{z}_t)$. In addition, it has to be noted that (18) is defined up to a scale factor, which is another reason to keep $\delta(\hat{z}_t)$ as a variable value. Therefore, given two corresponding points between images, they must satisfy:

$$p'^T \hat{E} p < \delta(\hat{z}_t). \tag{19}$$

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. The procedure is depicted in Fig. 2, where a detected point $P(x, y, z)$ is assumed, and it is represented in the first image reference system by a normalized vector $\vec{p_1}$ due to the unknown scale. To deal with this scale ambiguity, we suggest a point distribution to generate a set of multi-scale points $\lambda_i \vec{p_1}$, being

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

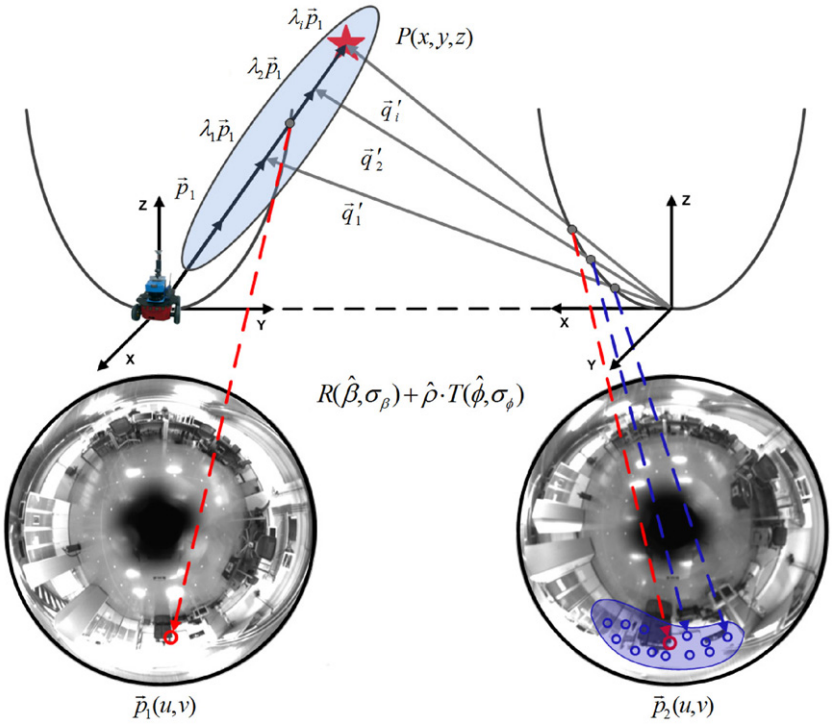

Fig. 2. Given a detected point $\vec{p}_1$ in the first image reference system, a point distribution is generated to obtain a set of multi-scale points $\lambda_i\vec{p}_1$. By using the Kalman prediction, they can be transformed into $\vec{q}_i'$ in the second image reference system by means of $R \sim \mathrm{N}(\hat{\beta}, \sigma_\beta)$, $T \sim \mathrm{N}(\hat{\phi}, \sigma_\phi)$ and $\hat{\rho}$. Finally $\vec{q}_i'$ are projected into the image plane to determine a restricted area where correspondences have to be found. Circled points represent the projection of the normal point distribution for the multi-scale points that determine this area.

representative for the lack of scale in $\vec{p}_1$. This distribution considers a valid range for λ_i according to the predicted $\hat{\rho}$. Please note that the error of the current estimation of the map has to be propagated along the procedure. To that end, we look back to the Kalman filter theory, where the innovation is defined as the difference between the predicted $\hat{z}_t$ and the real z_t observation measurement as stated in (12), and the covariance of the innovation defined in (13). So that $S_i(t+1)$ presents the following structure:

$$S_i(t+1) = \begin{bmatrix} {\sigma_\phi}^2 & \sigma_{\phi\beta} \\ \sigma_{\beta\phi} & {\sigma_\beta}^2 \end{bmatrix}. \tag{20}$$

As the predicted $\hat{E}$ can be decomposed in a rotation $\hat{R}$ and a translation $\hat{T}$, we can transform the distribution $\lambda_i\vec{p}_1$ into the second image reference system, obtaining $\vec{q}_i'$. The introduction of (20) allows to propagate the error, and thus it redefines a transformation between images through the normal distributions $R \sim \mathrm{N}(\hat{\beta}, \sigma_\beta)$ and $T \sim \mathrm{N}(\hat{\phi}, \sigma_\phi)$. Therefore $\vec{q}_i'$ is a gaussian distribution correlated with the current map uncertainty. Once obtained $\vec{q}_i'$, they are projected into the image plane of the second image, seen as circled points in Fig. 2. This projection of the normal multi-scale distribution determines the predicted area which is drawn with a continuous curve line on the omnidirectional image. This area establishes the specific image pixels where correspondences for $\vec{p}_1$ must be searched for. The shape of this area depends on the error of the prediction, which is directly correlated with the current uncertainty of the current map estimation. Dash lines represent the possible candidate points located inside the predicted area. Hence the problem of matching is simplified to the search for the correct corresponding points for $\vec{p}_i$ amongst those candidates inside a restricted area, instead of a global search along the whole image.

4. SGD

4.1. Structure

The SGD algorithm has been the second method considered in this work to be applied to the case of visual SLAM and it is responsible for generating a feasible estimation for the problem.

In this case, the problem is dealt with a graph-oriented map, which contains a set of nodes to define the poses traversed by the robot and the views initialized into the map. It is considered as a maximum-likelihood estimator, and it seeks a least squares minimization [22]. The state vector s_t encodes this representation through a set of variables which are expressed in the following manner:

$$s_t = \left[(x_0, y_0, \theta_0), (x_1, y_1, \theta_1) \cdots (x_n, y_n, \theta_n)\right] \tag{21}$$

being (x_n, y_n, θ_n) the 2D position and orientation of each node in a general reference system. Despite the fact that this kind of representation seems the most natural and intuitive, such global encoding has the main drawback of not being capable to update more than one node and its adjacents per constraint. This aspect has led to a general agreement in the use of the incremental representation:

$$s_t^{inc} = \begin{bmatrix} (x_0, y_0, \theta_0) \\ (dx_1, dy_1, d\theta_1) \\ \vdots \\ (dx_n, dy_n, d\theta_n) \end{bmatrix} \tag{22}$$

where $(dx_n, dy_n, d\theta_n)$ represents the deviation between two consecutive poses in the global reference system. According to the

formulation defined in (1) and (21), x_v and each x_{l_n} correspond with (x_0, y_0, θ_0), $(x_1, y_1, \theta_1) \cdots (x_n, y_n, \theta_n)$, and thus:

$$s_t^{inc} = \begin{bmatrix} (x_0, y_0, \theta_0) \\ (x_1 - x_0, y_1 - y_0, \theta_1 - \theta_0) \\ (x_2 - x_1, y_2 - y_1, \theta_2 - \theta_1) \\ \vdots \\ (x_n - x_{n-1}, y_n - y_{n-1}, \theta_n - \theta_{n-1}) \end{bmatrix}. \quad (23)$$

Now, the state vector is differentially encoded and each single update has influence on the whole map reestimation.

Regarding the observation measurements, a complementary subset of edges are introduced to relate nodes to each other. That is to say, they express the observation measurements between poses, either from odometry of the wheels or visual sensors. The nomenclature commonly refers to the observations as constraints, and it denotes them as δ_{ji}, where j indicates the observed node, seen from node i. The general objective stated by these kind of methods [23,12] is to minimize the error likelihood expressed as:

$$P_{ji}(s) \propto \eta \exp\left(-\frac{1}{2}(f_{ji}(s) - \delta_{ji})^T \Omega_{ji}(f_{ji}(s) - \delta_{ji})\right) \quad (24)$$

being $f_{ji}(s)$ a function dependent on the state s_t and both nodes j and i. The difference between $f_{ji}(s)$ and δ_{ji} expresses the error deviation between nodes, which in this case are views of the map and poses traversed by the robot. Such error term is weighted by the information matrix:

$$\Omega_{ji} = \Sigma_{ji}^{-1} \quad (25)$$

where Σ_{ji}^{-1} is the inverse covariance matrix responsible for the uncertainty of the observation measurements. After taking the logarithm we have:

$$F_{ji}(s) \propto (f_{ji}(s) - \delta_{ji})^T \Omega_{ji}(f_{ji}(s) - \delta_{ji}) \quad (26)$$

$$= e_{ji}(s)^T \Omega_{ji} e_{ji}(s) = r_{ji}(s)^T \Omega_{ji} r_{ji}(s) \quad (27)$$

being $e_{ji}(s)$ the error resultant from $f_{ji}(s)$–$\delta_{ji}(s)$, which is also named as $r_{ji}(s)$ to emphasize its condition of residue. Finally, the global problem seeks the minimization of the objective function which represents the accumulated error on the map:

$$F(s) = \sum_{\langle j,i \rangle \in G} F_{ji}(s) = \sum_{\langle j,i \rangle \in G} r_{ji}(s)^T \Omega_{ji} r_{ji}(s) \quad (28)$$

where $G = \{\langle j_1, i_1 \rangle, \langle j_2, i_2 \rangle \ldots\}$ defines the subset of particular constraint conforming the map, either pertaining to odometry or visual observation measurements.

4.2. Estimation

Once the formulation of the problem has been stated, the SGD algorithm develops an iterative process to reach a valid estimation for the SLAM problem. The basis of a SGD method lays on the minimization of (28) through derivative optimization techniques such as mean square estimators, so that the estimated state vector is obtained as:

$$s_{t+1} = s_t + \Delta s \quad (29)$$

where Δs updates s_t, by means of an adaptive constraint's optimization. It is worth noting that in a general case, this update is calculated independently at each step by using only a single constraint, that is to say $\Delta s = f(\delta_{ji})$. The general expression for the transition between s_t and s_{t+1} has the following form:

$$s_{t+1} = s_t + \lambda \cdot H^{-1} J_{ji}^T \Omega_{ji} r_{ji}. \quad (30)$$

- $J_{ji}(s)$ is the Jacobian of $f_{ji}(s)$ with respect to s_t. It translates the error deviation into a spacial variation.
- H is the Hessian matrix, calculated as $J^T \Omega J$, and it shapes the error function through a preconditioning matrix to scale the variations of J_{ji}:

$$H \approx \sum_{\langle i,j \rangle} J_{ji} \Omega_{ji} J_{ji}^T. \quad (31)$$

- Ω_{ji} is the information matrix associated with a constraint, and equals Σ_{ji}^{-1}.
- λ is a learning factor to re-scale the term $H^{-1} J_{ji}^T \Omega_{ji} r_{ji}$. Normally, λ follows a decreasing criteria such as $\lambda = 1/n$, where n is the iteration step. This strategy pretends to achieve a final estimation by using higher values of λ at first steps, and presuming that lower values of λ will be useful in preventing from oscillations around the final solution.

This method updates the estimation by computing the rectification introduced by each constraint at each iteration step respectively. Despite the fact that the learning factor reduces the weight by which each constraint updates the estimation, the procedure may be inefficient as it may lead to an unstable solution. Undesired oscillations may occur due to the stochastic nature of the constraints' selection. For this reason, we propose an optimization process which takes into account several constraints at the same iteration step. Such idea might cause undesired overloads of time. However, we also propose some amendments to avoid this effect, which succeed in maintaining the time requirements and even reduce them.

4.3. Adaption to omnidirectional images

Regarding the observation measurements provided by an omnidirectional camera, some assumptions have to be contemplated in the structure of the SGD algorithm.

Note that in this approach we are dealing with a visual observation given by an omnidirectional camera. This fact requires the adaption of the equations defined in the previous section, since the nature of the constraints are not only metrical like odometry's constraints. Following, we detail the terms related to the observation measurements, emphasizing on the visual observation, which has been redefined in consequence with (5):

- The first adaption was referred to $f_{ji}(s)$, differentiating between odometry and visual observation constraints:

$$f_{j,i}^{odo}(s) = \begin{pmatrix} dx_j \\ dy_j \\ d\theta_j \end{pmatrix} + \begin{pmatrix} dx_{j-1} \\ dy_{j-1} \\ d\theta_{j-1} \end{pmatrix} + \cdots + \begin{pmatrix} dx_i \\ dy_i \\ d\theta_i \end{pmatrix} \quad (32)$$

$$f_{j,i}^{visual}(s) = \begin{pmatrix} \phi \\ \beta \end{pmatrix} = \begin{bmatrix} \arctan\left(\frac{\mathrm{dy}_j - \mathrm{dy}_i}{\mathrm{dx}_j - \mathrm{dx}_i}\right) - d\theta_i \\ d\theta_j - d\theta_i \end{bmatrix} \quad (33)$$

  where ϕ and β express the relation between views and the pose codification (21), and are directly computed as defined in [15]. Visual inspection of Fig. 1 may ease to define (33).
- Then, it is necessary to recalculate $J_{ji}(s) = \frac{\partial f_{ji}(s)}{\partial s}$, accordingly with the previous reformulation. It has to be noticed the importance of considering the value of each node's index, being either $j > i$ or $j < i$, since the derivatives vary its form considerably. Furthermore, as seen above, the dimensions of $f_{ji}(s)$ are different, fact which has also to be considered in order to resize the rest of the terms involved in the SGD algorithm.

$$J_{ji}(s) = \frac{\partial f_{ji}(s)}{\partial s} = \left[\frac{\partial f_{ji}(\phi)}{\partial s}, \frac{\partial f_{ji}(\beta)}{\partial s}\right]. \quad (34)$$

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

- Lastly, we also propose that the estimation of the new state s_{t+1} reflects the usage of several constraints at the same time. We seek more relevance of constraints' weight when searching for the optimal minimum estimation. Obviously, computing more than one constraint at each step may cause a certain overload. Contrarily, in this approach we reduce the expensive estimation of H. In a general case, at every step, H is computed as many times as constraints exist in the map. In opposition with this, we only compute H once for each subset of constraints introduced simultaneously into the system at each step. Thus we dramatically reduce the number of times that H is calculated, so that we proceed in a more efficient manner which compensates possible time overloads.

5. Results

We have performed different real data experiments in an office environment. The equipment utilized in the experiments consisted of a Pioneer P3-AT indoor robot equipped with a firewire 1280 × 960 camera and a hyperbolic mirror to build the omnidirectional image. The optical axis of the camera is installed approximately perpendicular to the ground plane, as described in Fig. 3. As a result, a rotation of the robot corresponds to a rotation of the image with respect to its central point. In addition, we used a SICK LMS range finder in order to compute a ground truth by means of the method presented in [2]. The exposition of the results is structured as it follows: First in Section 5.1 we show SLAM results obtained with both methods EKF and SGD when the dimension of the map in terms of N views is variable. Then in Section 5.2, we also compare both methods by testing their accuracy and robustness on the estimation when data association errors arise. Finally, in Section 5.3 we present results with regard to the speed of convergence.

5.1. SLAM results with EKF and SGD

This experiment has been conducted in an indoor environment which corresponds to an office area of 42 × 32 m. The robot navigates this area while it acquires omnidirectional images and laser data along the trajectory. The laser data is an auxiliary reference to aid in generating a ground truth for fair comparison.

In the EKF's case, as mentioned above, the procedure of map building is accomplished in an incremental manner. Fig. 4 shows the results obtained in this experiment, where the robot starts the SLAM process by adding the first view of the map. Next, it keeps moving along the trajectory while capturing omnidirectional images. The image at the current robot pose is compared with the views stored in the map so as to extract some corresponding points that allow the robot to compute a relative measurement of its position, as explained in Section 2. The robot decides to initiate a new view whenever the relative appearance of the current image compared to the appearance of the map's views drops below a specific similarity threshold R. The ellipses indicate the uncertainty in the pose of each view and the robot. The dash-dotted line represents the solution obtained with the EKF approach, indicating with crosses the points along the trajectory where the robot decided to initiate new views in the map. The continuous line represents the ground truth whereas the odometry is drawn with dash line. The modification of R, leads to a variation of the size of the map in terms of N. As it can be observed in Fig. 4(a), a map for an environment of 42 × 32 m may be perfectly generated by a reduced set of $N = 5$ views, thus leading to a compact representation. However, the same environment may also be represented with a different number of views N as shown in Fig. 5(a). Figs. 4(b) and 5(b) compare the errors for the estimated trajectory, each one associated with the maps composed by $N = 5$ and $N = 20$ views respectively. Based on the ground truth comparison, the solution error is shown with dash-dotted line and the odometry's with dash line at every step of the trajectory. The validity of the solution is confirmed due to the accomplishment of the convergence requirements. It may be noticed that the solution error is inside the 2σ interval, drawn in continuous line, whereas the odometry error grows out of bounds. According to these results, it should be noticed that the higher values of N the lower the resultant error in the map.

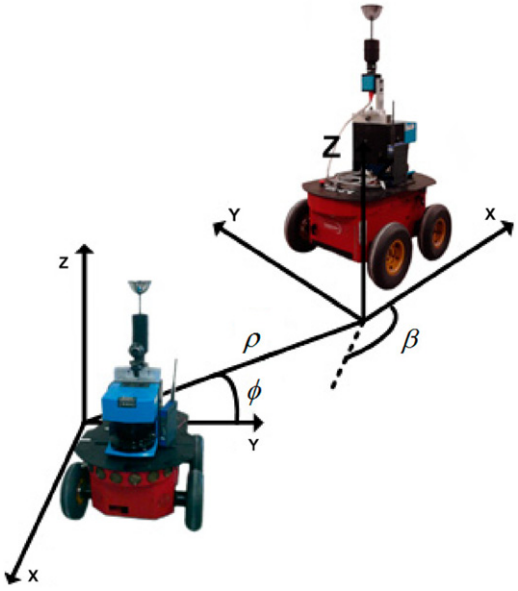

Fig. 3. Robot Pioneer P3-AT used in the experiments. Two poses are indicated with their corresponding relative angles which determine the motion transformation.

On the other hand, we run the same experiment with a SGD estimator. Fig. 6(a) and (b) represent the same two situations with $N = 5$ and $N = 20$ views previously performed. The placement of the views is exactly the same. The main difference in the manner to proceed with respect to EKF is that SGD processes the observations offline. Inspecting Figs. 4(a), 5(a), 6(a) and (b) reveals that EKF estimations are more accurate than the SGD estimations. To generalize, Fig. 7 establishes a fair comparison between both methods, where the RMS (Root Mean Square) error along the path is represented versus the number of views N. The continuous line shows the RMS error for SGD and the dash line shows the EKF's. The results of EKF outperforms in this case SGD's. However, this experiment has dealt with a desirable situation where non-linear errors, if any, were low enough so that the EKF response was able to ensure convergence. The following experiment will show the results obtained when the visual information is damaged and corrupted by significative noise errors.

5.2. Comparing accuracy

Now we intend to compare the behavior of both methods in a more realistic situation, that is to say, when they are expected to suffer non-linear errors introduced by the observation measurements and it consequently causes wrong data association errors. We have conducted the same real experiment shown above but assuming a highly relevant presence of non-gaussian errors. To that end, we have modeled a random generator scheme which introduces wrong data associations. At each estimation step, the robot computes the observation measurements for the entire set of views which is able to observe. However the robot fails to associate the observation measurement with its corresponding view at a

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

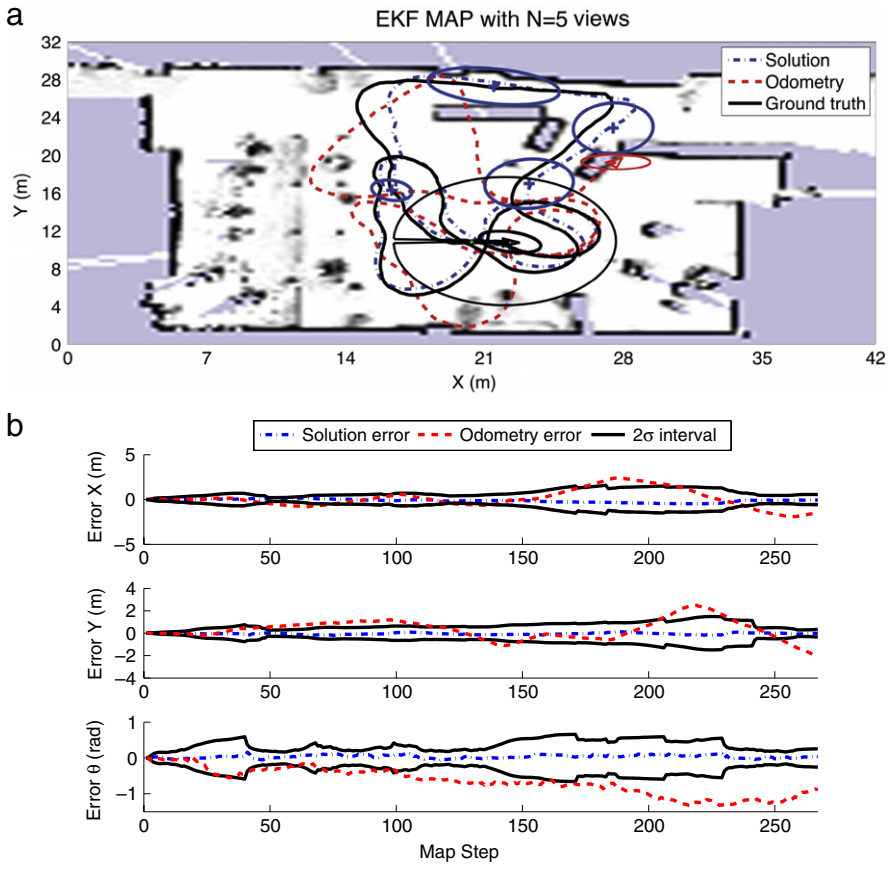

Fig. 4. (a) presents results of SLAM using an EKF algorithm with real data. The map representation of the environment is formed by $N = 5$ views. The position of the views is presented with error ellipses. (b) shows the solution and the odometry error in X, Y and θ at each time step.

certain probability, meaning that a percentage out of the total data association are wrong, and thus the observation measurement as well.

Fig. 8(a) and (b) describe the RMS error tendency of both methods, when data association fails with a given probability. The experiment has been repeated 200 times in order to retrieve consistent and coherent mean values. Again, the environment has been represented with different values of N in order to show differences. The results provided by EKF reveal that the resultant RMS error grows out of bounds when the probability of data association error is apparently low. This fact demonstrates the low reliability of the EKF when it has to deal with non-linearities and thus non-gaussian errors. Despite the fact that maps with more views provide a larger number of observation measurements to enable the rectification of the estimation, the error continuously increases. The results prove that once the solution diverges, the EKF is unable to recover it, despite the fact that N is higher. Consequently, the difficulties experienced by the EKF to keep the convergence of the estimation are evidenced.

Contrary to the EKF's results, and according to Fig. 8(b), the SGD provides a lower RMS error under the same conditions. Moreover, it ensures convergence, as the RMS's tendency only increases slightly. It is worth noting the importance of selecting a suitable value for λ, so that new updates to s_{t+1} do not lead the estimation to diverge when there is evidence of errors. In this case, the SGD proves its capability to rectify the solution even in presence of non-linearities and the consequent non-gaussian errors. Therefore, in the case of SGD, as it could be intuitively expected, the more N views in the map, the more observations gathered, and thus the better results provided.

5.3. Comparing speed of convergence

As it may be seen in the previous subsection, the SGD outperforms EKF in terms of robustness and accuracy when the system is considerably affected by non-gaussian errors. However, one should think about the speed of convergence of both methods. A compromising solution will have to be agreed so as to ensure a balance which provides robustness against the influence of noisy terms and speed of computation. With this experiment we would like to compare the speed ratios by which EKF and SGD compute a valid solution. Fig. 9 represents the time consumption to reach a valid solution versus the number of views N of the map. Since we look for a fair comparison, the y-axis, has been transformed into a normalized time variable which achieves a trustworthy comparison between both schemes. This adoption has been considered since the stochastic nature of the SGD method may lead each experiment to last a different number of iterations, and consequently a different time. Therefore the mean values for each iteration step have to be considered, so that the final estimation time can be obtained. Hence this normalization allows a fair and simpler comparison between methods.

In this sense, it may be proved that the solution provided by EKF outperforms the solution given by a basic SGD for each N-view map, since its gradient is definitely lower. However, it is also worthwhile to analyze these results together with the tendency of each corresponding RMS error. Fig. 10(a) and (b) show the normalized RMS error, versus the total time consumption to reach the final estimation. Now it can be clearly confirmed that quicker speed of convergence is assured by EKF.

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

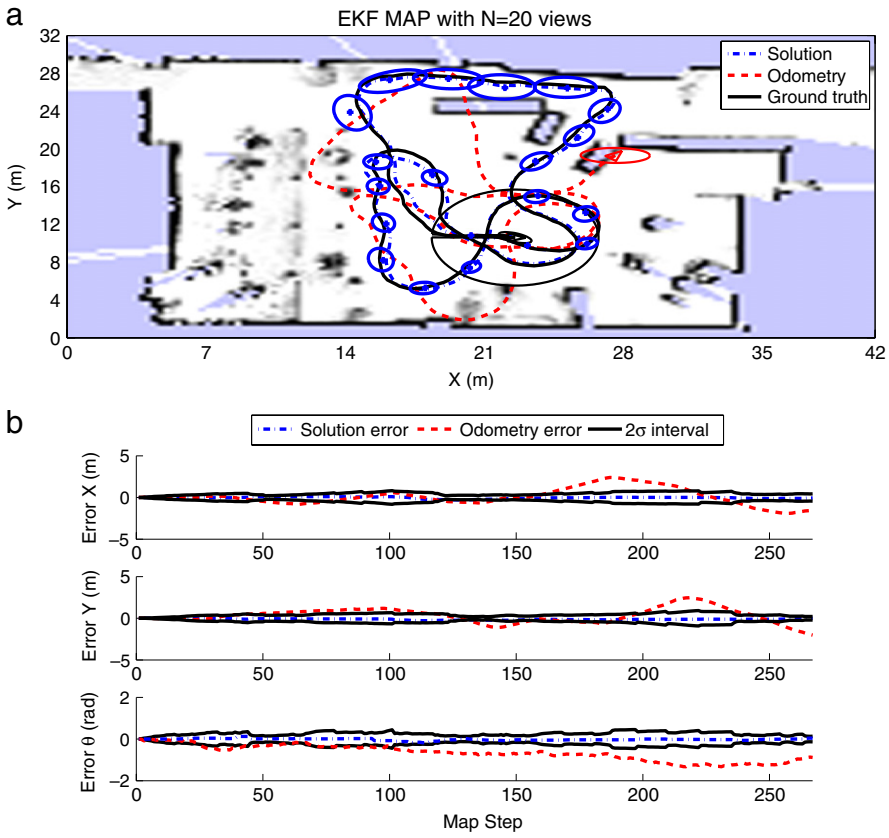

Fig. 5. (a) presents results of SLAM using an EKF algorithm with real data. The map representation of the environment is formed by $N = 20$ views. The position of the views is presented with error ellipses. (b) shows the solution and the odometry error in X, Y and θ at each time step.

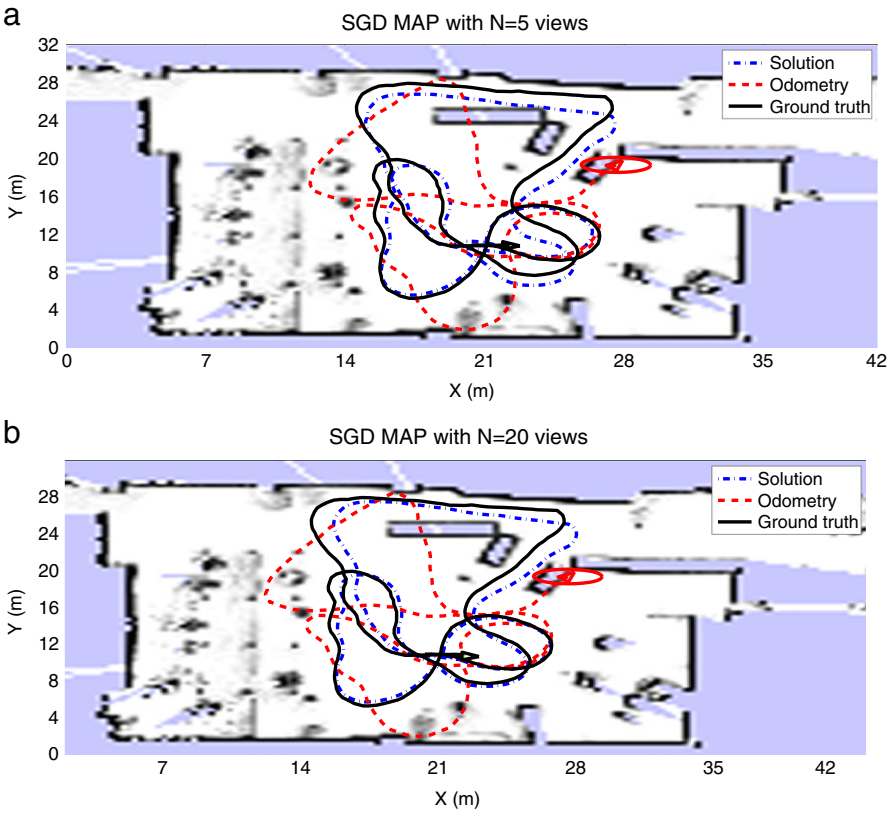

Fig. 6. (a) and (b), present results of SLAM using a SGD algorithm with real data. These map representations of the environment are formed by $N = 5$ and $N = 20$ respectively. The dash-dotted line represents the solution obtained with the SGD approach, the continuous line represents the ground truth whereas the odometry is drawn with dash line.

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

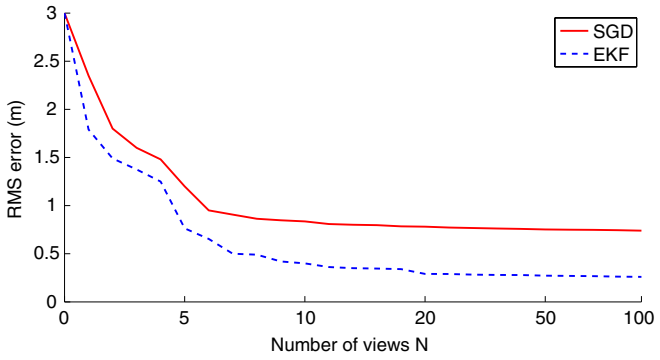

Fig. 7. RMS error (m) versus the number of views N of the map. The continuous line shows the error for the solution provided by SGD, meanwhile the dash line shows the error for the solution obtained with EKF.

6. Conclusions

We have presented a comparison between EKF and SGD algorithms, according to their provided solution to the Simultaneous Localization and Mapping (SLAM) approach. The main issue to analyze has been the influence of non-linear errors, which are a clear indicator of added noise by the visual sensor's measurements, especially associated with the omnidirectional observation model. We have presented a real data experimental set, which has considered different modifications so as to test the behavior of both methods under different conditions. The approach to the map representation relies on an efficient view-based map model, which is built by means of a reduced set of omnidirectional image views. Bearing in mind the results presented in this work, a key aspect to remark about EKF is definitely its capability to provide a suitable estimation in real time, thanks to its adequate speed of

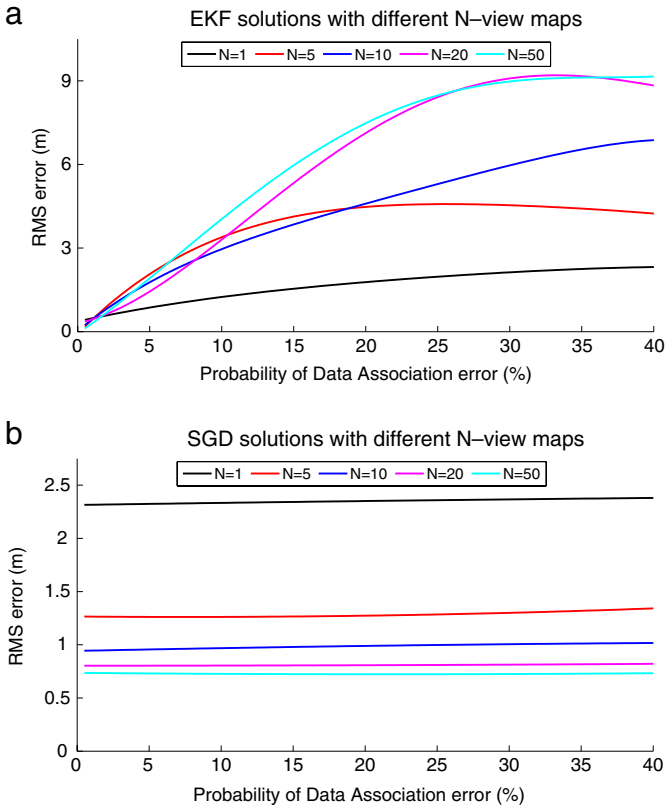

Fig. 8. (a) and (b) presents the RMS error (m) versus the probability of data association error (%) for EKF and SGD respectively. Errors for maps with different number of views N are indicated.

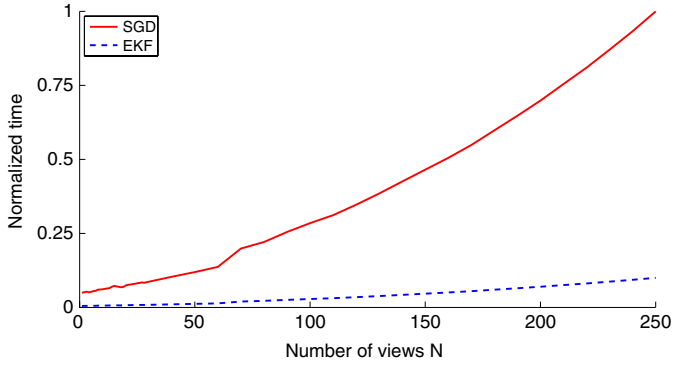

Fig. 9. Normalized time consumption versus number of views N of the map. The continuous line shows the time consumed by SGD, meanwhile the dash line shows the time consumed by EKF.

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

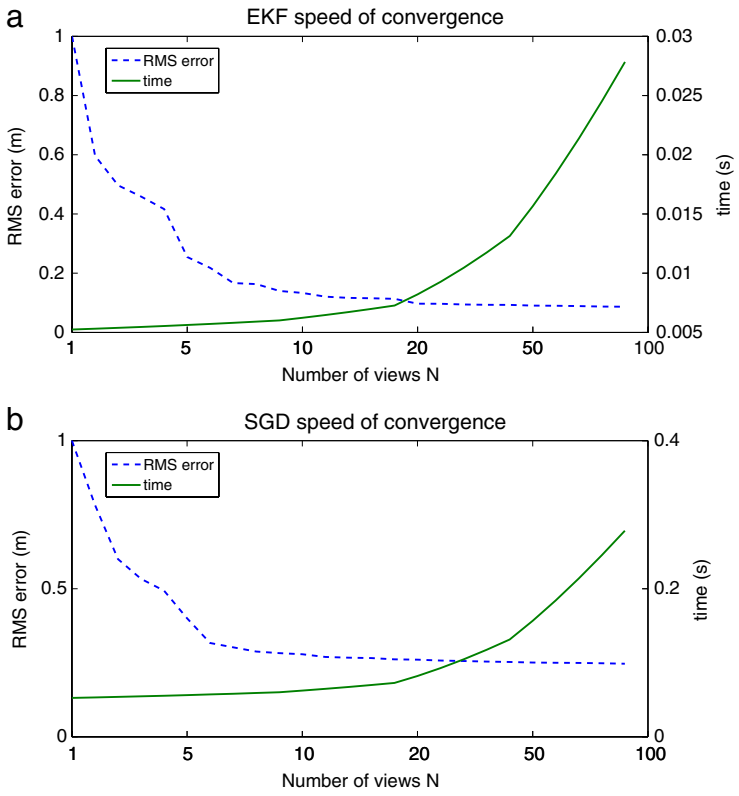

Fig. 10. (a) and (b) present the normalized RMS error (m), and time consumption (s) versus the number of views N of the map for EKF and SGD respectively. The dash lines show the RMS error, meanwhile the continuous lines show the time consumed by EKF and SGD respectively.

convergence. Moreover, other favorable aspect in case of an idealistic situation without clear evidence of non-linearities, is that EKF provides a more accurate estimation in contrast to SGD. On the other hand, contrary to EKF, the SGD has evidenced to be more reliable when a robust solution is required. Despite the fact that SGD's accuracy in an idealistic situation is lower than the EKF's, the results obtained in presence of non-linear noise effects, indicate that SGD provides a solid and stable solution which prevents the system from diverging. As it is well known, this is not accomplished by EKF, since is highly sensitive to errors due to the linearization of the variables of the filter. However, the SGD reveals a lower speed of convergence.

Therefore it has been proved that the effectiveness of each method depends on the assumed conditions. Assuring and approach to SLAM which achieves the avoidance of the effects of non-linearities and non-gaussian errors, would lead to select a SGD method. Nevertheless, in case of dealing with a more desirable situation, such as in a low-noise environment, would indicate that an EKF method would be more appropriated in order to succeed in providing a more precise solution with a higher rate of convergence.

Acknowledgments

This work has been supported by the Spanish government through the project DPI2010-15308, and the grant program FPI2011.

References

[1] Y. Chou, L. Jing-Sin, A robotic indoor 3D mapping system using a 2D laser range finder mounted on a rotating four-bar linkage of a mobile platform, Int. J. Adv. Robot. Syst. 10 (2013), http://dx.doi.org/10.5772/54655.
[2] C. Stachniss, G. Grisetti, D. Haehnel, W. Burgard, Improved Rao–Blackwellized mapping by adaptive sampling and active loop-closure, in: Proceedings of the Workshop on Self-Organization of Adaptive Behavior (SOAVE), Ilmenau, Germany, 2004, pp. 1–15.
[3] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit, FastSLAM: a factored solution to the simultaneous localization and mapping problem, in: Proceedings of the 18th National Conference on Artificial Intelligence, Edmonton, Canada, 2002, pp. 593–598.
[4] K. Wurm, C. Stachniss, G. Grisetti, Bridging the gap between feature- and grid-based SLAM, Robot. Auton. Syst. 58 (2010) 140–148.
[5] 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.
[6] J. Civera, A.J. Davison, J.M. Martínez Montiel, Inverse depth parametrization for monocular SLAM, IEEE Trans. Robot. 24 (2008) 932–945.
[7] 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, 2010, pp. 281–288.
[8] S.-E. Yu, D. Kim, Image-based homing navigation with landmark arrangement matching, Inform. Sci. 181 (2011) 3427–3442.
[9] 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, 2011, pp. 4084–4090.
[10] A.J. Davison, D.M. Murray, Simultaneous localisation and map-building using active vision, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 24 (2002) 865–880.
[11] F.A. Moreno, J.L. Blanco, J. Gonzalez, Stereo vision specific models for particle filter-based SLAM, Robot. Auton. Syst. 57 (2009) 955–970.
[12] G. Grisetti, C. Stachniss, S. Grzonka, W. Burgard, A tree parameterization for efficiently computing maximum likelihood maps using gradient descent, in: Proceedings of the Robotics: Science and Systems (RSS), Atlanta, USA, 2007, pp. 1–8.
[13] 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.
[14] S. Park, S. Kim, M. Park, S.-K. Park, Vision-based global localization for mobile robots with hybrid maps of objects and spatial layouts, Inform. Sci. 179 (2009) 4174–4198.
[15] 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, 2012, pp. 48–57.

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009

[16] J. Neira, J.D. Tardós, Data association in stochastic mapping using the joint compatibility test, IEEE Trans. Robot. Automat. 17 (2001) 890–897.
[17] C. Berger, Weak constraints network optimiser, in: Proceedings of the International Conference on Robotics and Automation (ICRA), Saint Paul, USA, 2012, pp. 1270–1277.
[18] H. Bay, T. Tuytelaars, L. Van Gool, Speeded up robust features (SURF), Comput. Vis. Image Underst. 110 (2008) 346–359.
[19] A.C. Murillo, J.J. Guerrero, C. Sagüés, SURF features for efficient robot localization with omnidirectional images, in: Proceedings of the International Conference on Robotics and Automation (ICRA), San Diego, USA, 2007, pp. 3901–3907.
[20] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, J. Basic Eng. 83 (1961) 95–107.
[21] D. Scaramuzza, Performance evaluation of 1-point RANSAC visual odometry, J. Field Robot. 28 (2011) 792–811.
[22] L. Bottou, Stochastic Learning, in: Lecture Notes in Artificial Intelligence (LNAI), vol. 3176, Springer Verlag, Berlin, 2004.
[23] D. Olson, J. Leonard, S. Teller, Fast iterative optimization of pose graphs with poor initial estimates, in: Proceedings of the International Conference on Robotics and Automation (ICRA), Orlando, USA, 2006, pp. 2262–2269.

David Valiente received the M. Eng. degree in Telecommunications Engineering from the Miguel Hernández University (UMH), Elche, Spain, in 2009, receiving also the Best Academic Student award in Telecommunications Engineering by the UMH. From 2009 to 2011, he worked as a researcher at the Communications department and at the Systems Engineering and Automation department of the UMH. Since 2011 he has a research position as scholarship holder in the area of Systems Engineering and Automation of the UMH, receiving a grant (FPI) by the Spanish Government. His research interests are focused on mobile robots, omnidirectional vision, visual feature extraction and visual SLAM.

Arturo Gil received the M. Eng. degree in Industrial Engineering from Miguel Hernández University (UMH), Elche, Spain, in 2002, receiving also the Best Student Academic award in Industrial Engineering by the UMH. He obtained the Ph.D. degree in 2008, entitled: "Cooperative construction of visual maps by means of a robot team". Since 2003, he works as a lecturer and researcher at the UMH, teaching subjects related to Control and Computer Vision. His research interests are focused on mobile robotics, visual SLAM and cooperative robotics. He is currently working on techniques to build visual maps using teams of mobile robots.

Lorenzo Fernandez received the M. Eng. degree in Telecommunications Engineering from the Miguel Hernández University (UMH), Elche, Spain, in 2008. He joined the Systems Engineering and Automation department of the UMH in 2008, where he is involved in research projects. In 2010 he obtained a Ph.D. fellowship (VALi+d) from the Valencian Regional Government to enroll in a Ph.D program at the Systems and Automation department. His research interests are mobile robots, visual appearance and visual SLAM.

Óscar Reinoso received the M. Eng. degree from the Polytechnical University of Madrid (UPM), Madrid, Spain, in 1991. Later, he obtained the Ph.D. degree in 1996. He worked at Protos Desarrollo S.A. company in the development and research of artificial vision systems from 1994 to 1997. Since 1997, he works as a professor at Miguel Hernández University (UMH), teaching subjects related to Control, Robotics and Computer Vision. His research interests are in mobile robotics, climbing robots and visual inspection systems. He is member of CEA–IFAC and IEEE.

Please cite this article in press as: D. Valiente, et al., A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images, Robotics and Autonomous Systems (2013), http://dx.doi.org/10.1016/j.robot.2013.11.009