Martin, F., Carballeira, J., Moreno, L., Garrido, S. & Gonzalez, P. (2017). Using the Jensen-Shannon, Density Power, and Itakura-Saito Divergences to Implement an Evolutionary-Based Global Localization Filter for Mobile Robots. *IEEE Access, 5*, 13922–13940.

DOI: 10.1109/access.2017.2724199

Using the Jensen-Shannon, Density Power, and Itakura-Saito divergences to implement an evolutionary-based global localization filter for mobile robots

Fernando Martín, Juan Carballeira, Luis Moreno, *Member, IEEE,* Santiago Garrido, and Pavel González

***Abstract*—One of the most demanding skills for a mobile robot is to be intelligent enough to know its own location. The global localization problem consists of obtaining the robot's pose (position and orientation) in a known map if the initial location is unknown. This task is addressed applying evolutionary computation concepts (Differential Evolution). In the current approach, the distances obtained from the laser sensors are combined with the predicted scan (in the known map) from possible locations to implement a cost function that is optimized by an evolutionary filter. The laser beams (sensor information) are modeled using a combination of probability distributions to implement a non-symmetric fitness function. The main contribution of this work is to apply the probabilistic approach to design three different cost functions based on known divergences (Jensen-Shannon, Itakura-Saito, and Density Power). The three metrics have been tested in different experiments and the localization module performance is exceptional in regions with occlusions caused by different obstacles. This fact validates that the non-symmetric probabilistic approach is a suitable technique to be applied to multiple metrics.**

***Index Terms*—Jensen-Shannon Divergence, Itakura-Saito Divergence, Density Power Divergence, Differential Evolution, Global Localization, Mobile robots**

I. Introduction

THE Global Localization (GL) problem is one of the most demanding tasks that an autonomous robot needs to accomplish in order to improve its capabilities. It is considered that the environment is known (a map is provided) and the objective is to develop an algorithm to obtain the robot's pose (position and orientation) if the initial location is unknown. A common approach is to distinguish between two classes of GL systems according to the information source. The positioning systems, where the most typical device is the Global Positioning System (GPS), use signals emitted from external sources to localize the robot. This strategy is not applicable in environments or regions where external signals are not available. The self-positioning systems rely on the robot sensor system. For example, an autonomous robot can estimate its location using laser scans provided by laser range finders. Information about the map is needed and proprioceptive data (odometry from wheel encoders) is very useful. This approach is more adequate for indoor environments where it is difficult to receive external signals. In this paper, the GL module is included in the second option. The sensor data provided by a laser range finder is the only information that is required to estimate the robot's pose in an indoor map.

Our recent work has been focused on the application of evolutionary optimization techniques to implement GL modules. The Differential Evolution (DE) method [1] has been chosen to be the basis of solutions for two-dimensional (2D) [2] and three-dimensional (3D) [3] environments. DE is a population-based procedure where each component of the set represents a candidate (pose) that is weighted by a fitness value. The population set evolves in an iterative way according to the evolutionary principles to optimize the cost value. If the fitness function is designed to represent the difference between a real observation vector obtained from the true location and an estimated one measured in the known map, the population will converge to the best solution (where the observation vector and the estimated readings are more similar), making it possible to apply this type of strategy to the aforementioned task. The sensor system is used to measure the observation vectors (laser scans).

A key factor of the optimizer is the cost function. In the basic versions of our GL module, the mean squared error was selected to compute the cost value of each candidate. The Euclidean distances between laser observations and estimates with the same bearing were calculated to compute the cost value. In [4], a completely different approach based on the Kullback-Leibler (KL) divergence [5] was applied to implement the fitness function. The same metric is applied in [6] to develop a similar algorithm where the engine is a mixture between the Markov chain Monte Carlo (MC) method and the DE technique.

According to the state of the art, the most common strategy when developing a cost function for this type of filters is to choose a symmetric metric. The KL divergence, which is non-symmetric, is included in a wider class of metrics known as Csiszár-Morimoto (CM) $f-$divergences [7]. This dissimilarity measure is also classified as a Bregman divergence [8] because it is non-symmetric and the triangle inequality is not satisfied. This metric is used to measure differences between probability distributions. One of the main advantages of using an asymmetric metric is that it allows to penalize some situations depending on the sensor information. The cost function with an asymmetric divergence can be designed to improve the

F. Martín, J. Carballeira, L. Moreno, S. Garrido, P. González were with the Robotics Lab, Carlos III University, Madrid, Spain, e-mail (corresponding author): fmmonar@ing.uc3m.es.

performance in some unexpected situations. This feature has been exploited to increase the robustness to occlusions.

A requirement to measure the KL divergence is that probability distributions have to be defined for the data provided by sensors. The laser measurements are not only defined by distances and probability distributions represent the whole beam. Different types of distributions with different weights are combined to model different events. These profiles were selected in [4] with the objective of obtaining a GL algorithm more robust to occlusions. Cases with possible occlusions are favored by the weights of the fitness function. In this way, the probability profiles and the KL-based cost function make it possible to cause a great improvement in the method performance when the robot is in an environment with significant occlusions [4], [6]. The asymmetry of the metric is exploited to increase the robustness to occlusions.

After multiple experiments, it was realized that the improvement was not limited to the KL divergence and the key factor was the probability profiles approach. Therefore, we found it necessary to validate the probabilistic technique by developing different cost functions. In recent years, multiple papers have been published about applications of dissimilarity measures in many research fields [9], [10], [11]. In addition, different works have been conducted to establish links and correspondences between metrics. An exhaustive study is given in [12]. The probability distributions that were successfully implemented can be applied to other metrics. In this paper, the same profiles have been adopted to develop other cost functions based on different metrics. Three different divergences have been chosen to implement cost functions: Jensen-Shannon (JS) [13], Density Power (DP) [14], and Itakura-Saito (IS) [15]. The objective is to test the performance of the GL module with the new cost functions, checking if the capabilities are held and examining advantages and disadvantages.

The analysis of the experiments carried out in this work lets us conclude that all metrics have the same advantages that were reported in our previous publications. In particular, the great performance in regions with significant occlusions originated by uniform noise or unmodeled obstacles is maintained for all divergences. Therefore, it is validated that the non-symmetric probabilistic approach is a suitable technique to be applied to multiple metrics. The type of divergence is important because it permits to introduce probability distributions to compare sensor measurements, but the crucial stage when developing a GL method robust to occlusions is the definition of the probability profiles.

This paper is structured as follows. Section II contains the state of the art. In Section III, the main ideas of the evolutionary GL filter are detailed. After that, the probability profiles approach and the divergences with the cost functions are explained in Sections IV and V, respectively. The tests that have been carried out are analyzed in Section VI and the most significant results are pointed out in Section VII.

II. State of the art

There are multiple classes of GL techniques that can be categorized depending on many aspects. A common approach makes a distinction between Bayesian-based, optimization-based, and hybrid filters.

The Bayesian-based methods are probabilistic techniques where the localization problem is solved in two steps. The Bayesian inference rule is used for that purpose. First, sensor information (motion and environment perception) is read and used to build the *a posteriori* density function. After that, the robot's pose is computed according to this function. Different criteria can be applied to calculate the pose. Accurate models for the density functions are needed in order to represent probabilities in the whole map. The probability distribution converges to the solution when the problem is successfully solved. Many research groups have presented interesting approaches that rely on the aforementioned ideas. For example, it is possible to cite the grid-based probabilistic filters [16], [17] or the MC localization strategies [18], [19], [20]. Different authors have proposed modifications to reduce the population requirements. Biswas *et al.* [21] propose to apply the gradient of the sensor measurements model to decrease the population size that is required to solve the problem. Their method is called Corrective Gradient Refinement. Other interesting ideas can be found in the works of Grisetti *et al.* [22] and Zhang *et al.* [23]. Lenser and Veloso [24] have developed a variation of MC called Sensor Resetting Localization. Their objectives are to reduce the population requirements and the impact of unmodeled movements and systematic errors. Their sensor is a Charge-Coupled Device (CCD) camera. Ito *et al.* [25] have proposed a modified version of MC where Wifi and RGB-D information are applied to localize the robot in indoor environments. Wang *et al.* [26] have proposed a particle filter for mobile robots situated in high-occluded and dynamic environments.

The optimization filters rely on a fitness function that is utilized to compute a cost value that shows how good the solution is. The sensor information has to be updated when the robot is moving and the algorithm returns the best estimate after convergence or a maximum number of iterations. The candidate with the best fitness value is usually chosen as the solution of the GL problem. Two different ideas result in two different strategies. The well-known Kalman filters derive the cost function to compute the solution. They present good computational performance, but they cannot deal with multi-hypotheses distributions. Therefore, they have often been applied to tracking problems (re-localization), where only a single hypothesis is needed when computing the robot's pose. The second idea is to make a stochastic search over the space of possible solutions. The evolutionary algorithms, such as DE or Particle Swarm Optimization (PSO) filters, rely on this idea. Both methods and MC are compared in [27]. Lisowski [28] has implemented an hybrid version that mixes DE and MC. A different evolutionary technique called Harmony Search algorithm [29] is the basis of the GL filter designed by Mirkhania *et al.* [30]. As cited before, the DE method was applied in our previous work [2], [3]. Ronghua *et al.* [31] have proposed a mixture between MC and a genetic algorithm optimizer. Chien *et al.* [32] have applied PSO to implement a modified version of MC that avoids premature convergence.

Algorithm 1 DE-based GL filter

1: **function** $DE\_GL(real\_dist, pop, known\_map)$
2: **for** $i = 1 : N_P$ **do**
3: $estim\_dist(i) \leftarrow dist\_est(pop\_i, known\_map)$
4: $cost(i) \leftarrow DIV\_fitness(estim\_dist(i), real\_dist)$
5: **end for**
6: **while** (CONVERGENCE CONDITIONS) **do**
7: **for** $i = 1 : N_P$ **do**
8: $MUTATION$
9: $CROSSOVER$
10: $SELECTION$ with $THRESHOLDING$
11: $estim\_dist(i) \leftarrow dist\_est(pop\_i, known\_map)$
12: $cost(i) \leftarrow DIV\_fitness(estim\_dist(i), real\_dist)$
13: **end for**
14: $DISCARDING$
15: $[solution, error] \leftarrow min(cost)$
16: **end while**
17: **end function** ▷ return solution, error, and population

There are different techniques called hybrid methods or multi-hypotheses Kalman filters [33], [34] where the possible solution is formed by a set of probability distributions, but the problem resolution is not only based on Bayesian rules but on geometric constraints and decision trees. In [35], the solution is estimated using laser sensors. They join Gaussian distributions to model the sensor information. A combination between particle and Kalman filters is presented by Jochmann *et al.* [36].

The most common cost function of the GL filters is based on the quadratic error between estimated measurements and observations provided by the sensor system. Features of the map, such as walls, columns or obstacles, are used in the cost function to distinguish between possible locations. Different ideas have been proposed in our recent work. The absolute error was chosen in [37] and the KL divergence was applied in [4], [6]. Other authors have presented different techniques. The Hausdorff distance, which can be utilized to measure differences between set of points, is minimized in [38]. The entropy of future belief distributions is the variable that is considered by Fox *et al.* [39] to minimize the expected uncertainty of possible locations. The Mahalanobis distance is adopted by the feature-based strategy implemented by Arras *et al.* [40].

III. DE-BASED GL METHOD DESCRIPTION

The GL strategy is based on evolutionary optimization techniques. The population set contains candidates to be the correct location of the robot. These candidates are weighted by a cost value that compares sensor information from the true pose to sensor estimates from the candidate using the known map. The engine of the filter is the DE method that was first developed by Storn and Price [1]. DE is an evolutionary filter that has been applied to solve optimization problems in multiple fields. The main algorithm is detailed in this section. Since it has also been explained in our previous work, a brief description is given here. The reader can consult [2], [3] for a more exhaustive explanation. The method in pseudocode is presented in Algorithm 1.

The objective is to estimate three coordinates $(\hat{x}, \hat{y}, \hat{\theta})$ that correspond to the robot's pose in a known map. There are N_P candidates (*pop*) that will evolve and converge to the best solution according to the available information. The required size can be generated by an initialization method basing on the information contained in the laser scan [41]. The first population is generated randomly to cover the whole map. The laser readings from the robot's true pose, the known map, and the DE configuration parameters are also needed.

At the beginning, the cost value of each candidate is computed (lines $2-5$ of Algorithm 1). Sensor information from the true location ($real\_dist$) and from the candidate ($estim\_dist(i)$) have to be compared in order to do that. The laser sensor has been modeled by an observation vector formed by 61 readings separated $3°$. The sensor information is modeled using probability profiles following the same concepts applied in [4] to generate a GL filter where the main advantage is the excellent performance in environments with unmodeled obstacles that generate occlusions. The JS, DP, and IS divergences are applied to implement three different cost functions ($DIV\_fitness$) that will be described in Section V.

The GL algorithm is executed in an iterative way until convergence is reached (line 6). In each iteration and for the whole population set (lines $7-13$), the evolutionary filter creates the candidates to replace the current population members. Two evolutionary operators are applied to generate the new candidates. First, the mutation combines members of the population to form mutated vectors. After that, the diversity is increased by the crossover stage. New candidates are created by mixing (combining parameters) current population members and mutated ones.

The mutated vector for pop_i^k, a candidate at iteration k, is computed in line 8:

$$mut_i^k = pop_a^k + F(pop_b^k - pop_c^k), \tag{1}$$

where pop_a^k, pop_b^k, and pop_c^k are three random population members and F controls the differential variation.

The next evolutionary operation is the crossover. $cro_i^k = (cro_{i,1}^k, cro_{i,2}^k, \ldots, cro_{i,D}^k)^T$ is the name chosen for this vector:

$$cro_{i,j}^k = \begin{cases} mut_{i,j}^k & \text{if } p_{i,j}^k < \delta \\ pop_{i,j}^k & \text{otherwise} \end{cases}, \tag{2}$$

where $p_{i,j}^k$ is generated randomly in the interval $[0, 1]$ for each parameter of the population member. The crossover control variable is the crossover probability (δ), which is in the interval $[0, 1]$.

The number of chromosomes (D) represents the dimensions of the search space, which is three in this case. More information about types of evolutionary mechanisms can be found in [42]. The population set evolves according to a fitness function that is optimized. After convergence, the candidate with the best cost value will be the solution of the problem.

The selection step consists in comparing the new candidate to the current one to choose the population member for the next generation. The element with the best cost value is selected for the population of the next iteration or generation. A thresholding mechanism is added to the selection process. This mechanism introduces a limit to choose new solutions. When the fitness values are compared, the new candidate

is only accepted if the difference between costs is larger than a pre-specified threshold. The objective is to avoid the optimization caused by the sensor noise. This mechanism has been applied to multiple problems [43].

The thresholding mechanism can cause a degradation of the convergence properties because less solutions are accepted in each iteration. In order to increase the speed, the discarding step is implemented. At the end of each iteration, the worst population members (according to their costs) are substituted by solutions which are close to the best candidates. The percentage of candidates to be discarded has to be fixed.

When the algorithm converges, the best population member is chosen to be the estimate of the robot's pose in the known map.

The method cannot be limited to a static robot. The localization task does not end when the GL problem has been solved. After that, motion and sensor information can be integrated for multiple motion-perception cycles following the same ideas proposed in [2]. The best solution is saved and the population set is displaced according to the motion model. Since the robot is correctly localized, the population size could be drastically reduced without a decrease in the performance, which is an advantage of this technique.

IV. Probability profiles for fitness functions

Different divergences are proposed to implement the fitness function of the GL filter. All of them rely on the assumption that the sensor measurements are modeled as independent probability distributions for each laser beam. In this section, the sensor measurements model that makes it possible to improve the performance in environments with occlusions is detailed.

It has been assumed that the map is represented by an occupancy grid map, which is one of the most typical approaches in GL. In this type of map, the 2D environment is discretized using cells with a fixed size. A value between zero and one is computed for each cell to represent the probability of being occupied. If each cell of the map is denoted by m_{ij}, the full map can be defined by the following set:

$$m = \{m_{ij} : \quad 1 \leq i \leq n, \quad 1 \leq j \leq o\}, \tag{3}$$

where i and j are positive integers that take the map limits into account (n and o). The occupation probability of each cell is $p(m_{ij})$.

This notation lets us define the map estimation problem as the computation of the individual probabilities of each cell. Since the map and its probabilities are known, this representation will be very useful.

The information about the environment is given by the sensor measurements. In this work, the sensor is a laser scanner that measures the cells crossed by a laser beam.

The evolutionary-based GL method relies on a cost function that computes the difference between the observation vector (information provided by the laser scan) and the simulated vector from a possible solution (readings from the estimate, according to the known map). From a statistical point of view, it is possible to use different metrics to measure this difference.

The true location at time t is represented by $x_t = (x, y, \theta)^T$ and the observation vector is $z_t = (z_{t,1}, z_{t,2}, ..., z_{t,N_s})$. The estimated vector from a population member $\hat{x}_t$ is $\hat{z}_t$ (according to the map m).

A. The perception model

The laser scan contains N_s single measurements. In beam-based models, each component of the laser vector ($z_{t,i}$) can be modeled by probability distributions that consider different events. Different authors have proposed a Bayesian model to approximate the behavior of the range sensor beam model for dynamic environments [44]. These probabilistic models provide us $p(z_t|x_t, m)$ for a laser range finder, which is the probability of obtaining the laser scan z_t for a location x_t in the map m.

There are different research groups studying how to model the sensors measurements. The simplest approaches consider discrete grid maps [45], [46] or continuous metric maps [47], [48]. Moravec [46] has "proposed non-Gaussian densities over a discrete grid of possible distances measured by a sonar". Fox *et al.* [45] have suggested to use only the distance to the closest obstacle along the direction of the sensor. In this model, the measurement is originated by a mixture of two physical causes: the laser beam hits an object modeled in the map or with an object not modeled in the map. Thrun *et al.* [47] added two more causes to this model: people and the maximum range of the sensor.

In this work, the same ideas presented in [4] are followed. The probability density function for a single value originated by a laser beam is

$$\begin{aligned} p(z_{t,k}|x_t, m) \quad = \quad & k_h p_{hit}(z_{t,k}|x_t, m) + \\ & k_o p_{occl}(z_{t,k}|x_t, m) + \\ & k_u p_{unkn}(z_{t,k}|x_t, m), \end{aligned} \tag{4}$$

where each probability represents the next events:

- p_{hit}: Modeled obstacles. The laser measurement is modeled by a Gaussian distribution that includes the measurement noise. If the true distance is denoted by $z^*_{t,k}$, this distribution can be expressed as

$$p_{hit}(z_{t,k}|x_t, m) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2_{hit}}} e^{-\frac{(z_{t,k}-z^*_{t,k})^2}{2\sigma^2_{hit}}}, \tag{5}$$

  where σ^2_{hit} is the noise variance.
- p_{occ}: Unmodeled obstacles (possible occlusions). The unmodeled obstacles are modeled by a uniform distribution which varies between zero and one. This represents the unexpected objects that can be found in the environment and are not included in the map. These objects originate measurements shorter than the expected one according to the map. A typical example of that situation occurs when there are persons that are detected by the laser readings.
- p_{unkn}: Unknown places. The unknown places probability is also a uniform distribution (between zero and one) from the measurement up to the sensor range.

k_h, k_o, and k_u are constant numbers used to give different weights to these probabilities. These constant numbers are

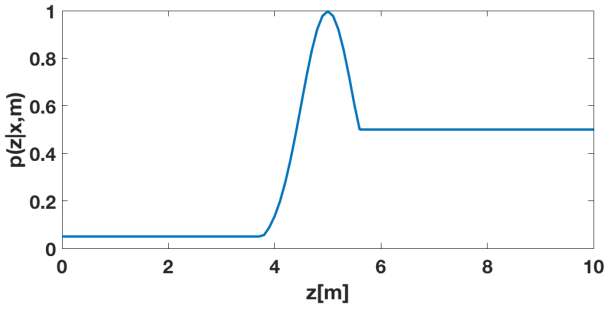

Fig. 1. Probability distribution for a single laser beam [4].

defined in the intervals in which they have influence on the density function.

Equation 4 can be understood by observing Figure 1. The laser measurement is 5 m, and p_{hit} is a Gaussian distribution with mean equal to 5 m and standard deviation equal to 0.5 m. The unmodeled obstacles are possible occlusions, and they are modeled as a uniform distribution with a low value ($k_o = 0.05$) when the measurement is shorter than the laser reading. There is no information about places with distances larger than five, thus the probability value chosen for these places is $k_u = 0.5$. The probability is defined up to the sensor range, which is 10 m in this case.

The cost function compares the real measurement with the estimated one. The probability distribution for the estimated measurement is

$$\begin{aligned} p(\hat{z}_{t,k}|\hat{x}_t, m) \quad = \quad & \hat{k}_h p_{hit}(\hat{z}_{t,k}|\hat{x}_t, m) + \\ & \hat{k}_o p_{occl}(\hat{z}_{t,k}|\hat{x}_t, m) + \\ & \hat{k}_u p_{unkn}(\hat{z}_{t,k}|\hat{x}_t, m), \end{aligned} \tag{6}$$

where equivalent variables have been chosen to represent the estimate.

The probability densities defined in Equations 4 and 6 are used to calculate the fitness value in the GL method. The cost function takes into account different situations in order to deal with possible occlusions. These situations are explained in the next section.

B. Robustness to occlusions - coefficients selection

An occlusion occurs if an unmodeled obstacle originates an error when comparing the robot's true location and the optimum solution in the map. Different cases are defined in order to deal with possible occlusions. When the real measurement is shorter than estimated reading, it can correspond to an occlusion that has to be considered by the cost function. It is not possible to have measurements from the optimum solution that are shorter than the true measurements if the robot is correctly localized. In this case, the fitness value has to present a higher value to discard this location. The penalization of its value is done choosing adequate weights for each situation.

The divergences will compute the distance between two probability distributions (P_S and $P_{\hat{S}}$, real observation and estimated one) for each laser measurement (k). The probability distributions are defined following the concepts that were detailed in Section IV-A. Different values will be given to the components of these distributions in order to deal with possible occlusions. The different options are explained in Table I, where the possible values for the variables of Equations 4 and 6 are represented.

TABLE I
COEFFICIENTS OF EQUATIONS 4 AND 6 ACCORDING TO THE DIFFERENCE BETWEEN THE REAL READING ($z_{t,k}$) AND THE ESTIMATE ($\hat{z}_{t,k}$).

	$z_{t,k} << \hat{z}_{t,k}$	$z_{t,k} \leq \hat{z}_{t,k}$	$z_{t,k} > \hat{z}_{t,k}$	$z_{t,k} >> \hat{z}_{t,k}$
k_o	0.1	0.1	0.1	0.95
$\hat{k}_o$	0.05	0.05	0.05	0.05
k_h	0.9	0.9	0.9	0.95
$\hat{k}_h$	0.95	0.95	0.95	0.05
k_u	0.15	0.5	0.9	0.95
$\hat{k}_u$	0.5	0.5	0.5	0.05

An occlusion can exist when the real measurement is much lower than the estimated reading. In this case, the cost function is not penalized and the probability profiles are similar to Figure 1. These values are represented in the left column of the table. Observe that $k_u = 0.15$. This low value does not penalize the cost function, and the robustness to occlusions is increased.

When the real measurement is similar but shorter to the estimated one ($z_{t,k} \leq \hat{z}_{t,k}$), it can be due to an occlusion, but also the difference can be caused by the sensor noise. In this situation, the fitness function is penalized by $k_u = 0.5$. Notice that this factor has no influence when $z_{t,k} = \hat{z}_{t,k}$.

It is not possible to have $z_{t,k} >> \hat{z}_{t,k}$ if the candidate solution is the true location, which corresponds to the right column of the table. This is probably a wrong measurement, thus the fitness function is strongly penalized. k_o, k_h, and k_u are fixed to a high value (0.95), and $\hat{k}_h$, $\hat{k}_o$, and $\hat{k}_u$ are equal to a lower one (0.05). Therefore, the cost value presents a high value and this estimate will be discarded.

When the estimated value is slightly higher than the real distance ($z_{t,k} > \hat{z}_{t,k}$), the estimated pose is not the correct one but it can be close to the optimum value. The penalization factor is only given to the unknown cells ($k_u = 0.9$).

It has to be said that the previous values have been empirically fixed to deal with occlusions. Different constants will result in different behaviors. The algorithm is very sensitive to these values because the probability profiles and, therefore, the cost values, will be completely different depending on the weights given to the probability densities. The good performance is shown in the experimental results section.

V. DIVERGENCE-BASED FITNESS FUNCTIONS

Although there are multiple divergences that could be applied to implement the cost function of the GL strategy, three metrics have been chosen in this work. Each one will be explained in this section. The KL divergence was adopted in our previous work [4], [6]. The JS divergence can be viewed as a symmetric version of the KL divergence. The DP divergence can be equivalent to a quadratic error version when using probability distributions. The IS distance is similar to the KL

divergence when its formula is analyzed.These divergences belong to different families of metrics with interesting properties and connections between them.

Different works have been published to establish formal relations between metrics in information theory [12], [49], [50]. Among them, Cichocki and Amari [12] have studied wide families of divergences named Alpha, Beta, and Gamma. The dissimilarity measures applied in this paper can be included in the first two categories. The researchers use power functions to generalize the KL divergence and to obtain different classes of divergences. They report that the power functions allow to increase the robustness with respect to outliers and, therefore, the performance is better or more flexible (an example can be found in [51]). Using this approach, it is possible to define three families of divergences (Alpha, Beta, and Gamma) that can be viewed as generalizations of the KL divergence. All classes are linked and it is possible to do transformations between them [9].

These families are derived from the well-known CM $f-$divergence and the Bregman divergence. The CM $f-$divergences are obtained using the following equation:

$$d_{CM}(P||Q) = \sum_i q(i) f(\frac{p(i)}{q(i)}), \tag{7}$$

where f is the generator function and p and q are the densities of two probability distributions P and Q.

The Bregman divergences are given by

$$d_{BR}(P||Q) = \sum_i [\Phi(p(i)) - \Phi(q(i)) - \frac{\delta\Phi}{\delta q(i)}(p(i) - q(i))], \tag{8}$$

where Φ is the generator function.

The Alpha divergence [52] is a special case of CM $f-$divergence defined by the following formula:

$$d_A^{(\alpha)}(P||Q) = \frac{\sum_i [p^\alpha(i) q^{1-\alpha}(i) - \alpha p(i) + (\alpha - 1) q(i)]}{\alpha(\alpha - 1)}. \tag{9}$$

This metric depends on the variable parameter α. For example, when $\alpha \to 1$, the generalized KL divergence is obtained ($\sum_i [p(i) \ln(p(i)/q(i)) + p(i) - q(i)]$).

From the CM divergences, it is possible to establish some basic properties for the Alpha divergences: non-negativity ($d_A^{(\alpha)} \geq 0$ and $d_A^{(\alpha)} = 0$ if and only if $P = Q$), convexity with respect to both P and Q, continuity (continuous function of real variable α), duality ($d_A^{(\alpha)} = d_A^{(1-\alpha)}$), etc. The reader can consult [12] to find more properties and a more detailed explanation.

The Beta divergence [14], [53] is obtained from the Bregman divergence:

$$d_B^{(\beta)}(P||Q) = \frac{\sum_i [p^\beta(i) + (\beta - 1) q^\beta(i) - \beta p(i) q^{\beta-1}]}{\beta(\beta - 1)}. \tag{10}$$

This divergence is dependent on β. The connection between the Bregman and the Beta divergences is strong, and the properties of the Bregman divergence are also valid for the Beta divergence: non-negativity ($d_B^{(\beta)} \geq 0$ and $d_A^{(\beta)} = 0$ if and only if $P = Q$), convexity with respect to P, linearity (a positive linear combination of Bregman divergences is also a Bregman divergence), invariance under affine transforms, etc. The Beta divergence has a single global minimum equal to zero for $P = Q$, and increases with the absolute value of the difference between p and q.

Since the KL divergence is utilized multiple times in this paper, a reminder of this metric is given in the next section. After that, the cost functions are detailed and connected to the families of divergences. Finally, an illustrative example for a single laser beam is shown.

A. Kullback-Leibler

The KL divergence [5] can be defined as "a non-symmetric measure of the difference between two probability distributions P and Q". The KL divergence in discrete spaces is defined as

$$d_{KL}(P||Q) = \sum_i p(i) \ln \frac{p(i)}{q(i)}. \tag{11}$$

The KL divergence exists for probability distributions in which the sum of densities is equal to one ($\sum_i p(i) = \sum_i q(i) = 1$). Only cases where both densities present positive values ($q(i) > 0$ and $p(i) > 0$) are computed by the formula. The quantity $0 \ln 0$ is assumed to be zero. The KL divergence from P to Q is not equal to the KL divergence from Q to P.

If $S(x, z) = \{m_{ij}^{x,z}\}$ is the area (in cells) that is crossed by an observation z when the robot is located at x, the KL divergence for a given orientation (k) can be expressed by

$$d_{KL}^k(P_{S(x_t,z_{t,k})}||P_{\hat{S}(\hat{x}_t,\hat{z}_{t,k})}) = \\ = \sum_{i,j \in S_T} p_{S(x_t,z_{t,k})}(m_{ij}) \ln \frac{p_{S(x_t,z_{t,k})}(m_{ij})}{p_{\hat{S}(\hat{x}_t,\hat{z}_{t,k})}(m_{ij})}, \tag{12}$$

where S_T is the maximum between $S(x_t, z_{t,k})$ and $\hat{S}(\hat{x}_t, \hat{z}_{t,k})$. The number of cells crossed by the true observation $S(x_t, z_{t,k})$ and the area covered by the estimate $\hat{S}(\hat{x}_t, \hat{z}_{t,k})$ are different. It has been considered that the limit of these areas is the first unknown cell that is reached. It has been assumed that each cell is independent from the others and it is possible to apply the additive property of the KL divergence to independent random events.

In order to simplify this formula, the following expression will be considered from now on:

$$d_{KL}^k(P_{S_k}||P_{\hat{S}_k}) = \sum_{i,j \in S_T} p_{S_k}(m_{ij}) \ln \frac{p_{S_k}(m_{ij})}{p_{\hat{S}_k}(m_{ij})}. \tag{13}$$

This formula compares laser readings for a given orientation. The KL divergence for laser scans formed by N_s observations can be expressed as

$$d_{KL}(P_S||P_{\hat{S}}) = \sum_{k=1}^{N_s} d_{KL}^k(P_{S_k}||P_{\hat{S}_k}). \tag{14}$$

The KL divergence is greater or equal to zero, zero being the value when the real laser scan is exactly equal to the estimated reading in the known map (non-negativity property of the CM f-divergences). In this adaptation of the KL metric, the probability densities are not normalized (Section IV) and

the absolute value is used if negative values are obtained for d_{KL}. The same concept is also applied to the other metrics. A correction factor is introduced to distinguish between places where the same cost value is obtained but the number of occlusions is different:

$$KLD = d_{KL}(P_S||P_{\hat{S}})e^{\frac{N_{occ}}{N_s}}, \quad (15)$$

where N_{occ} represents the number of occlusions ($z_{t,k} << \hat{z}_{t,k}$). This correction factor has been chosen empirically. KLD represents the cost value that was used in [4] to solve the GL problem.

A typical problem of the KL divergence has to be considered. When $q(i)$ is very low for a particular i, the specific term $p(i)/q(i)$ can dominate the result. This issue has not been found in the experiments that have been performed in this paper. Observing Figure 1 and Table I, $P_{\hat{S}}$ is in the interval $[0.05, 0.95]$. In this way, the difference between larger and lower probabilities is not large enough to cause the cited problem in the current implementation.

B. Jensen-Shannon

The JS divergence [13] is derived from the KL divergence. Given the KL divergence between two probability distributions (Equation 14), the JS divergence is

$$d_{JS}(P||Q) = \frac{1}{2}[d_{KL}(P||M) + d_{KL}(Q||M)], \quad (16)$$

where $M = (P + Q)/2$.

This metric can be classified as a symmetrized Alpha divergence. In general, there are two ways to symmetrize divergences. The first option is to use Equation 16. The JS divergence is a special case of Alpha divergence that is obtained when Equation 16 is applied to symmetrize the Alpha divergence (using d_A instead of d_{KL}) and the limit for $\alpha \to 0$ is computed.

The Jeffreys divergence [54] is obtained with the second symmetrization option:

$$d_{JF}(P||Q) = \frac{1}{2}[d_A(P||Q) + d_A(Q||P)]. \quad (17)$$

In this case, the limit for $\alpha \to 1$ has to be computed for Equation 17. Due to its relation with the JS divergence, some results using the Jeffreys divergence are included in the experiments.

The next formula is applied to compute the JS divergence for the whole scan:

$$d_{JS}(P_{S_k}||P_{\hat{S}_k}) = \frac{1}{2}[d_{KL}(P_{S_k}||M) + d_{KL}(P_{\hat{S}_k}||M)], \quad (18)$$

where $M = (P_{S_k} + P_{\hat{S}_k})/2$.

This divergence is also called information radius (IRad) [55] or total divergence to the average [56]. It has important differences with respect to the KL divergence. For example, a finite value is always obtained and it is a symmetric metric. The name that is usually given to the square root of the JS divergence is the JS distance [57]. This dissimilarity measure is a symmetrized and smoothed variant of the KL divergence. It is bounded between zero and two.

This metric is based on Jensen's inequality [58] and the Shannon entropy. Jensen's inequality "relates the value of a convex function of an integral to the integral of the convex function". The application of its related theorems to information theory are used to define both the KL and the JS divergences. The Shannon entropy, named after the American mathematician Claude Shannon, is a well-known measurement used in information theory that is defined as $H = -\sum_i p(i) \log_b p(i)$.

The JS divergence has been applied to multiple fields, such as computer science, machine learning, medicine, history, etc. Some examples are given in papers about quantum information theory [59], machine learning [60], and medicine [61].

Finally, the cost value when the correction factor that considers occlusions is introduced is computed by

$$JSD = d_{JS}(P_S||P_{\hat{S}})e^{\frac{N_{occ}}{N_s}}. \quad (19)$$

C. Density Power

The DP divergence was developed by Basu *et al.* [14]. Following the same ideas explained in the previous sections to develop the cost function of the GL filter, the next formula is applied to compute the DP divergence for a given orientation:

$$d^k_{DP}(P_{S_k}||P_{\hat{S}_k}) = \sum_{i,j \in S_T} [p^{1+\rho}_{S_k}(m_{ij}) + \frac{1}{\rho} p^{1+\rho}_{\hat{S}_k}(m_{ij}) - \left(1 + \frac{1}{\rho}\right) p_{\hat{S}_k}(m_{ij}) p^{\rho}_{S_k}(m_{ij})]. \quad (20)$$

This dissimilarity measure is a version of the Beta divergence. It has the same properties: single global minimum for $P_{S_k} = P_{\hat{S}_k}$, non-negativity, and increment dependent on the absolute value of the difference between densities. The authors in [12] believe that the most important motivation to study this metric from a practical point of view is to increase the robustness of the learning algorithms with respect to outliers.

It can be appreciated that this metric is dependent on a variable factor ρ. It has been reported in [14] that when $\rho = 1$ the DP divergence is equivalent to the Euclidean L2-norm:

$$d^k_{DP}(P_{S_k}||P_{\hat{S}_k}) = \sum_{i,j \in S_T} \left[p_{S_k}(m_{ij}) - p_{\hat{S}_k}(m_{ij})\right]^2. \quad (21)$$

The indefinition that appears when $\rho = 0$ is solved in [14] using the L'Hôpital's rule. The generalized KL divergence is obtained in this case.

Due to the versatility of the current method, this factor could be set to a fixed value or introduced as an additional chromosome (parameter in the population set) to be optimized. Both options have been tested. However, in the current version of the method, ρ will tend to one if Equation 20 is optimized. According to this formula, lower values are obtained for the cost function when ρ is one. For that reason, ρ is fixed to one in the experiments carried out in this paper. A different approach for the optimization problem is needed in order to exploit ρ as an additional optimization factor.

When the whole scan is considered, the next formula is computed:

$$d_{DP}(P_S||P_{\hat{S}}) = \sum_{k=1}^{N_s} d^k_{DPD}(P_{S_k}||P_{\hat{S}_k}). \quad (22)$$

Finally, the cost value when the correction factor that considers occlusions is introduced is computed by

$$DPD = d_{DPD}(P_S||P_{\hat{S}})e^{\frac{N_{occ}}{N_s}}. \quad (23)$$

It has to be remarked that one advantage of this formula is that, since ρ is fixed to one, the cost value is equivalent to a quadratic error version but using the probability profiles approach.

The DP divergence has been recently applied to point set registration in computer vision [62], multivariate analysis [63], or active learning [64].

D. Itakura-Saito

The IS divergence was proposed in the sixties by Itakura and Saito [15]. It is applied to the current problem following the same ideas described for the other divergences. For a given orientation of the laser scan, it is expressed as

$$d^k_{IS}(P_{S_k}||P_{\hat{S}_k}) = \sum_{i,j \in S_T} \left[\frac{p_{S_k}(m_{ij})}{p_{\hat{S}_k}(m_{ij})} - \ln \frac{p_{S_k}(m_{ij})}{p_{\hat{S}_k}(m_{ij})} - 1 \right]. \quad (24)$$

The IS divergence is classified as a Beta divergence. It is directly derived from Equation 10 when $\beta = 0$. Therefore, the properties of the Beta divergences are inherited.

It can be observed that the Beta divergences connect the IS divergence ($\beta = 0$), the generalized KL divergence ($\beta = 1$), and the Euclidean L2-norm ($\beta = 2$). The same can be concluded using different values of ρ because Equation 20 relies on a different version of the Beta divergence.

In [12], the authors explain that the choice of β is related to the statistical distribution of the data sets. For example, the optimal choice for the normal distribution is $\beta = 2$ and for the gamma distribution is $\beta = 0$. Although the problem studied here cannot be strictly linked to a known probability distribution, this analysis suggests that a further study about the statistical properties of the problem would help to choose the most suitable metric.

When the whole scan is considered, the divergence is equal to

$$d_{IS}(P_S||P_{\hat{S}}) = \sum_{k=1}^{N_s} d^k_{IS}(P_{S_k}||P_{\hat{S}_k}). \quad (25)$$

Finally, the cost value when the correction factor that considers occlusions is introduced is computed by

$$ISD = d_{IS}(P_S||P_{\hat{S}})e^{\frac{N_{occ}}{N_s}}. \quad (26)$$

This metric has been applied to multiple fields, such as matrix factorization in music analysis [65], machine learning [66], or malware detection [67].

E. Illustrative example

An example where the divergences are computed for a single laser orientation can be observed in Figure 2. The first measurement ($k = 1$) of the laser scan from the real pose and the estimate are represented in a grid map. The first step is to model the probability distributions of the laser beams. Different options have been defined. The values that have to be considered to compute the four divergences are shown in Table II. This case is only an introductory example to explain the concepts behind the last sections.

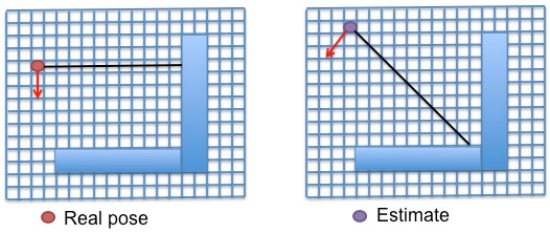

Fig. 2. First measurement obtained by the laser scanner. True location (left) and estimate (right). Laser beam in black [4].

Each cell of the laser beam is numbered in increasing order, starting from the first cell crossed by the ray. Different values are given to the free space (0.05), the obstacles (0.95), and the unknown space (0.5). The discrete probability distributions from the real location (p_1) and the estimated solution (p_2) are composed of 13 single values because the longest range is 13 cells. The probabilities are not normalized.

After adding the terms of the fourth row, the KL divergence is equal to 0.3475. The average sum of the fifth and sixth rows is the JS divergence, which is 0.3897 in this case. The DP divergence is equal to 1.2150. The IS divergence is 3.6578. It can be observed that the parameter that is used to measure the difference between probability distributions is completely different depending on the metric.

VI. Experimental Results

Different trials have been proposed to test different characteristics in simulated and real indoor maps that represent laboratories, corridors, and offices. The laser scan is formed by 61 readings separated 3° (180 field of view). The objective is to study the method performance for the cost functions that have been implemented in the current paper. In addition, it has to be checked if the same outstanding properties shown by the KL-based approach in environments with occlusions [4] are maintained when the new divergences are applied, fact that would validate the probability profiles used to model the information provided by the sensing system.

A. Accuracy

The first experiment examines the localization filter when the robot is situated at different locations of the simulated map displayed in Figure 3. The results are registered in Tables III and IV. The same population size is applied in all cases from a single location. The population requirements are only shown in the last table for simplicity. Each cell is a square of 12.1 cm side and the total map has $960 \times 300 = 288{,}000$ cells (4,200 m^2).

The robot's position and orientation are defined by a vector of three coordinates (x, y, θ). The position is defined by the first two parameters and the orientation is given by the third coordinate (zero being pointing right, horizontal direction, increasing clockwise). A Gaussian distribution over the laser measurement is used to model the sensor noise (σ). The

TABLE II
PROBABILITY DISTRIBUTIONS FOR THE EXAMPLE SHOWN IN FIGURE 2. VALUES: FREE SPACE $= 0.05$, OBSTACLE $= 0.95$, UNKNOWN SPACE $= 0.5$. $p_1(cell) = p_{S_k}(m_{ij})$, $p_2(cell) = p_{\hat{S}_k}(m_{ij})$. $p_M(cell) = [p_1(cell) + p_2(cell)]/2$. LEFT COLUMN: DIVERGENCE FOR WHICH THE PARAMETERS OF EACH ROW ARE NEEDED.

	cell	1	2	...	9	10	11	12	13
	$p_1(cell)$	0.05	0.05	...	0.05	0.05	0.05	0.95	0.5
	$p_2(cell)$	0.05	0.05	...	0.05	0.95	0.5	0.5	0.5
JS	$p_M(cell)$	0.05	0.05	...	0.05	0.5	0.275	0.725	0.5
KL	$p_1(cell) \ln \frac{p_1(cell)}{p_2(cell)}$	0	0	...	0	-0.1472	-0.1151	0.6098	0
JS	$p_1(cell) \ln \frac{p_1(cell)}{p_M(cell)}$	0	0	...	0	-0.1151	-0.0852	0.2568	0
JS	$p_2(cell) \ln \frac{p_2(cell)}{p_M(cell)}$	0	0	...	0	0.6098	0.2989	-0.1858	0
DP	$[(p_1(cell) - p_2(cell)]^2$	0	0	...	0	0.8100	0.2025	0.2025	0
IS	$\frac{p_1(cell)}{p_2(cell)} - \ln \frac{p_1(cell)}{p_2(cell)} - 1$	0	0	...	0	1.9971	1.4026	0.2581	0

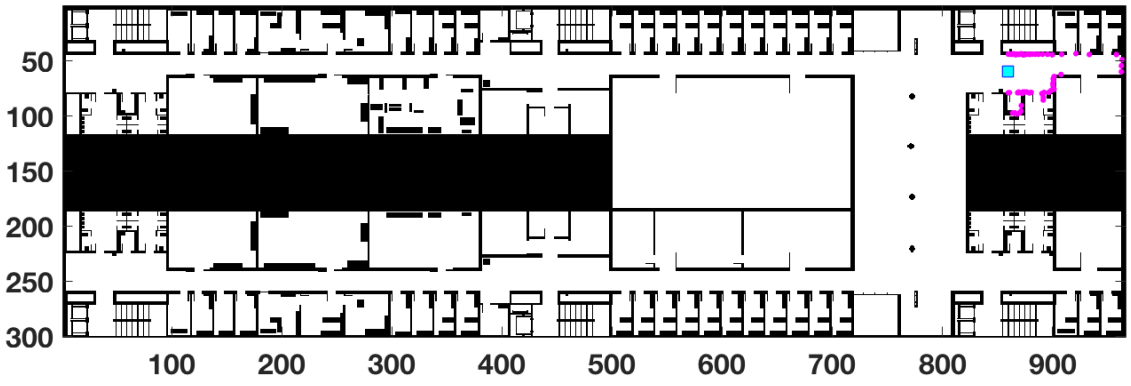

Fig. 3. GL in a simulated map. All units in cells. Robot's location in blue square.

standard deviation of this distribution is 1%. The difference between the robot's true position and the estimate is the position error. The orientation error measures differences in angles. A threshold has been fixed to decide if the robot is correctly localized.

The percentage of success or success rate is a variable that has been defined in these experiments to represent the number of algorithm executions in which the error is lower than a threshold divided by the total number of trials. This threshold is set to 50 cm in this paper. Failures are not included when computing errors. Other standard techniques such as the performance metrics proposed by Olson and Delen [68] are not applied because they are more suitable for classification models.

Observing Table III, the range of the error is [1.15, 19.60] cm. Similar results are presented in [4], [37]. The localization error is worse in larger offices because the laser error depends on the distance measured. There are not significant differences between metrics. The best accuracy is obtained with the KL divergence in most cases, but the difference with respect to the other metrics is not relevant. The orientation error (not included in tables for simplicity) is in the interval [0.00, 1.37] degrees, which is also a good result. The Jeffreys divergence has also been tested for comparison. Similar errors are obtained with this metric: [1.21, 18.90] cm and [0.06, 0.78] degrees.

The last variable that is analyzed is the success rate (Table IV). It has to be remarked that, in this experiment, motion is not considered. The GL algorithm is run for a single laser scan. Motion and scans from other locations would be needed for places with very similar sensor information, which could happen in indoor maps with symmetric corridors or similar rooms. Analyzing the results, the percentage of success is maximum in almost all cases for all metrics. However, this parameter could be worse. For example, motion will be needed in order to correctly localize the robot if it is situated in $(120, 25, 0)$ or in one of the small offices in the lower part of the map.

Similar trials have been carried out in a real map (Figure 4), which is a learned map of the Intel Research Lab provided by Dieter Fox[1]. The cell size is 5 cm and the area is 29×29 m^2. It represents a floor plan with a main hallway and different offices crowded with furniture.

The results are presented in Tables V and VI. The position error is in the interval [0.15, 13.50] cm. This error is slightly lower when compared to simulated conditions because the resolution of the real map is better. There are not significant differences between metrics. The orientation error is [0.01, 0.61] degrees and the percentages of success are optimal except in (80,50,90) (location with perceptual ambiguities).

[1]http://cres.usc.edu/radishrepository/view-one.php?name=intel_lab

TABLE III
POSITION ERROR WHEN THE ROBOT IS IN THE SIMULATED MAP OF FIGURE 3. ERRORS IN MEAN ± STANDARD DEVIATION (CM).

Robot's pose	KL	DP	IS	JS
(750,50,90)	1.69 ± 1.30	1.74 ± 0.30	1.99 ± 1.40	2.46 ± 1.40
(600,130,180)	8.29 ± 0.92	11.52 ± 5.10	12.56 ± 1.03	10.66 ± 5.40
(860,60,0)	1.36 ± 0.60	2.19 ± 0.28	1.38 ± 0.13	1.61 ± 0.16
(568,84,3)	2.87 ± 1.20	5.69 ± 0.51	3.51 ± 0.43	4.76 ± 0.54
(805,245,0)	1.15 ± 0.70	2.36 ± 1.70	1.63 ± 1.20	3.11 ± 1.60
(230,30,0)	6.24 ± 0.23	6.11 ± 3.50	6.10 ± 1.30	7.08 ± 1.50
(190,33,0)	5.86 ± 1.80	9.40 ± 3.40	9.40 ± 2.40	9.08 ± 3.40
(625,90,10)	12.64 ± 2.29	11.93 ± 3.02	10.78 ± 4.83	13.80 ± 2.95
(120,25,0)	11.82 ± 11.47	13.2 ± 11.94	17.25 ± 6.71	19.60 ± 10.68
(682,153,3)	11.30 ± 31.74	10.15 ± 3.34	12.29 ± 2.11	11.04 ± 1.67

TABLE IV
SUCCESS RATE (IN %) WHEN THE ROBOT IS IN THE SIMULATED MAP OF FIGURE 3.

Robot's pose	N_P	KL	DP	IS	JS
(750,50,90)	250	95	90	90	90
(600,130,80)	350	95	100	95	100
(860,60,0)	250	100	100	100	100
(568,84,3)	550	100	100	100	100
(805,245,0)	250	100	100	100	100
(230,30,0)	300	95	100	100	90
(190,33,0)	400	90	100	100	100
(625,90,10)	250	100	100	100	100
(120,25,0)	350	58	70	52	44
(682,153,3)	350	100	100	100	100

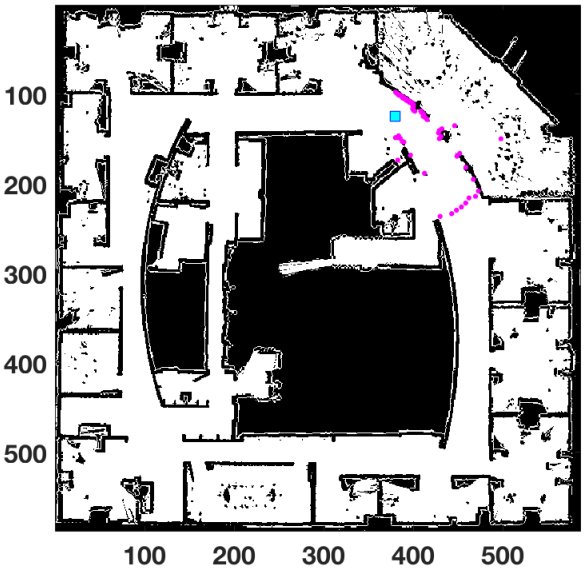

Fig. 4. GL in a real map. All units in cells. Robot's location in blue square.

Other research groups have proposed different localization methods that can be compared to our technique. The position error published in [38] by the research group of Donoso is [8, 15] cm. The average translation error of the technique proposed by Se *et al.* [69] is 7 cm. Their average angular error is one degree. Similar errors have been found when investigating the results of other research groups in the same type of environments. Since the other algorithms rely on different concepts and assumptions, it is necessary to study the configuration parameters and the experiment conditions of the other techniques to present a more exhaustive comparison.

The influence of the sensor noise is analyzed in Figure 5. The simulated map of Figure 6 is used for this experiment. The localization error and the percentage of success are represented against the sensor noise (σ is variable). Acceptable results are obtained with all metrics even with a high noise. Regarding the error, the KL-based cost function presents the best values and its performance is closer to a linear behavior. The worst error for higher noises is obtained with the IS divergence. The percentage of success is the maximum one (100%) up to 8% of sensor noise, which is above the error of the commercial sensors. With a 10% of sensor noise, the KL divergence presents the best error but the worst percentage of success. It has to be noticed that wrong cases (when the filter does not return the correct location) are not considered when computing errors.

B. Occlusions: uniform noise

The uniform noise has been chosen to model occlusions that can be caused by persons or dynamic obstacles that are not included in the known map. The same procedure was performed in previous publications to compare two different fitness functions (L1-norm and L2-norm) [37] and to show the good properties of the KL divergence [4]. In the current paper, the behavior when using different metrics in the cost function has been tested.

The objective is to study the algorithm performance when the true location is a distinguishable place of the map shown in Figure 6. The measurement provided by the laser sensor could be the real one or a random number. A uniform distribution has been defined to model the random measurements that represent the contamination of the sensor information. The minimum value is the 25% of the sensor distance and the maximum value is the 75% of the laser measurement. In this way, the contaminated laser reading can be expressed by

$$z_{k,c} = (1-\gamma)\mathcal{N}(z_k, \sigma) + \gamma\, \mathcal{U}(0.25z_k, 0.75z_k), \qquad (27)$$

where the laser measurement ($z_{k,c}$) can be given by the real sensor distance ($\mathcal{N}(z_k, \sigma)$) or the uniform distribution ($\mathcal{U}(0.25z_k, 0.75z_k)$) depending on the contamination level (γ). The mean of the normal distribution is the laser distance (z_k) and the sensor noise is represented by the standard deviation

TABLE V
POSITION ERROR WHEN THE ROBOT IS IN THE REAL MAP OF FIGURE 4. ERRORS IN MEAN ± STANDARD DEVIATION (CM).

Robot's pose	KL	DP	IS	JS
(380,125,0)	0.15 ± 0.19	0.68 ± 0.61	0.73 ± 0.58	1.04 ± 0.61
(320,120,10)	1.02 ± 1.46	0.83 ± 1.60	1.11 ± 1.51	0.93 ± 1.01
(440,258,21)	1.96 ± 4.12	2.05 ± 1.09	1.37 ± 2.02	2.45 ± 0.69
(75,520,3)	0.53 ± 0.35	2.01 ± 0.35	0.85 ± 1.04	0.72 ± 0.65
(80,50,90)	7.97 ± 7.36	13.38 ± 9.73	13.50 ± 9.15	12.10 ± 12.01
(122,120,15)	11.93 ± 9.77	7.22 ± 6.10	9.39 ± 7.67	10.20 ± 6.91
(180,60,0)	0.23 ± 0.15	0.52 ± 0.24	0.25 ± 0.17	0.30 ± 0.22

TABLE VI
SUCCESS RATE (IN %) WHEN THE ROBOT IS IN THE REAL MAP OF FIGURE 4.

Robot's pose	N_P	KL	DP	IS	JS
(380,125,0)	300	100	100	100	100
(320,120,10)	300	95	100	100	100
(440,258,21)	300	100	95	100	100
(75,520,3)	300	100	100	100	100
(80,50,90)	300	74	84	62	74
(122,120,15)	300	98	100	100	100
(180,60,0)	300	100	100	100	100

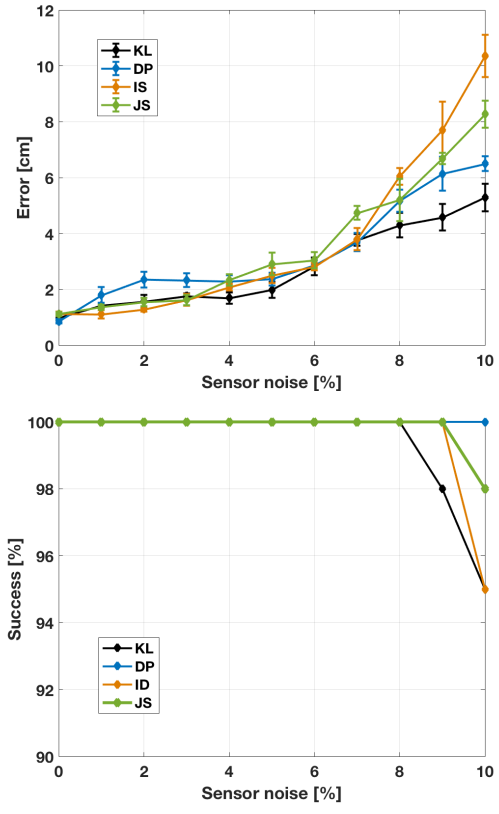

Fig. 5. Sensor noise influence. Top: position error vs. sensor noise. Standard deviation in error bars. Bottom: success rate vs. sensor noise. Robot's location: (60,50,0) in map of Figure 6.

($\sigma = 1\%$ in this experiment). The range of the uniform distribution is $[0.25z_k, 0.75z_k]$. γ is a random value that can be zero or one. The percentage of contaminated measurements is used to generate this value. The notation used in [37] has been followed. This is a particular case of occlusions because the contaminated measurements cannot be recognized in the map. The real pose is $(60, 50, 0)$. An example with a 50% of contamination is shown in Figure 7. It can be appreciated that the uniform noise causes significant changes in the laser scan.

The study of the influence of the contamination level is presented in Figure 8. The horizontal axis corresponds to the percentage of contaminated measurements (contamination level). The vertical axis is used for the position error (top). The percentage of success is in the right part.

The results are outstanding in all cases if the localization error is considered. Therefore, the probability profiles that are defined to model the different metrics cause a great improvement in the method behavior when uniform noise is included. The errors are lower than 5 cm even with a 50% of contamination. This means that the 50% of the laser readings are wrong measurements originated by the uniform noise. As reported in [4], the same error is obtained with 5% of contamination when the quadratic error cost function is used. Besides, when using the quadratic error, it is not possible to obtain the solution when the contaminated noise is greater or equal to 28%. For a higher noise (80%), the lowest error is obtained when the IS divergence is applied (13 cm) and the worst performance is presented by the DP divergence (43 cm) when the four metrics are compared. When compared to the Jeffreys divergence, this metric shows an error equal to 8.82 cm and 100% success with a 50% of contamination.

In [4], the evolution of the success rate with respect to the uniform noise was excellent when compared to traditional quadratic errors. The success rate started to decrease with a 50% of contamination. It was believed that this was the best behavior of the algorithm. However, the population size was not optimal. The population size has been increased to 250 in this experiment, and the percentage is 100% with a 68% of contamination in all cases. Therefore, the method success is even better than the previously reported one. The best results are obtained by the IS divergence and the worst option is the JS-based fitness function.

Wang *et al.* [26] have published a particle-based localization filter for high-occluded and dynamic environments. They have tested their technique in a 0.1 m resolution map. Their lower errors are 5 cm and 1.1 degrees with 2 moving people and 10 cm and 1.5 degrees with 8 moving people. The authors have defined a variable called occlusion ratio, which is similar to the percentage of uniform noise presented here. Although they do not present exact values depending on the occlusion ratio, they obtain an error of several meters when the occlusion ratio is approximately 30%. Their experiments are focused on the

Fig. 6. GL in a medium-size simulated map. All units in cells. Robot's location in blue square.

Fig. 7. Laser scan with 50% of contamination. All units in cells. Robot's location in blue square.

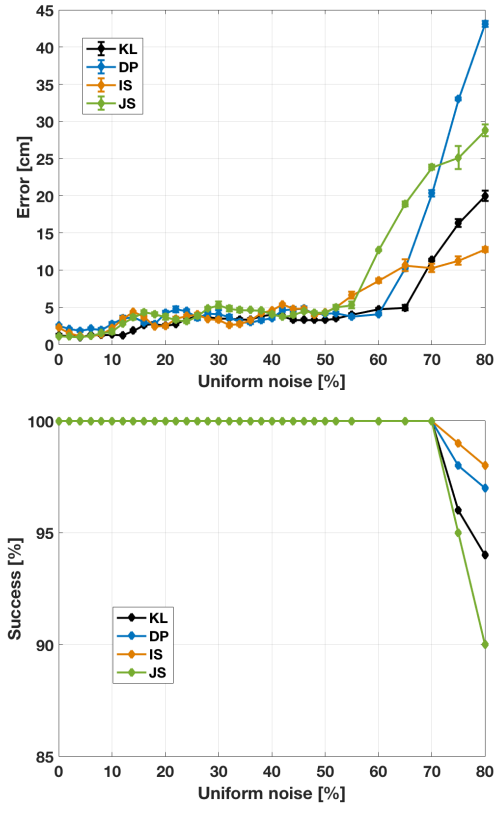

Fig. 8. Top: GL error vs. uniform noise in a simulated map (percentage of contaminated measurements). Standard deviation in error bars. Bottom: success vs. uniform noise.

tracking problem in a real scenario.

According to the results presented in this section, it can be concluded that the probability profiles that are defined to deal with occlusions produce a great improvement of the method capabilities in this type of situation.

C. Occlusions: unmodeled obstacles

In this section, the GL filter is examined when obstacles which are not modeled in the available floor plan are added. Two different experiments have been implemented when the robot is in the map of Figure 6. First, a huge obstacle is situated in front of the robot and the percentage of success is computed when the robot is approaching the obstacle. It can be appreciated that the scan is occluded and the situation is critical. Figure 9 registers the results. The scan corresponds to the location where the distance between the robot and the obstacle is 57 cells (6.9 m). The occlusion caused by the unmodeled obstacle makes the localization process harder.

The minimum distance with optimum results when using the quadratic fitness function was 87 cells (10.5 m) [4]. This distance is now in the interval $[31, 37]$ cells depending on the divergence. Since all metrics result in a similar performance, the probability profile is also excellent for this type of situation. Although no significant differences are found between metrics, the best results are obtained with the KL divergence.

Tsou and Wu [70] have published a localization method where feature matching is used to deal with dynamic obstacles. They have performed a similar experiment where a dynamic obstacle is added to the scanning area. Two different distances to the obstacle are tested: 434 cm with 2% dynamic information and 62 cm with 22% dynamic information. The average errors are 18.6 cm for the first case (2%) and 24.8 cm for the second one (22%). The success rate is approximately equal to 90% in both cases. Their algorithm is tested in a squared map (18.5 m side) where possible locations are generated using different resolutions. The results with the highest resolution, which is 49 cm, are reported here.

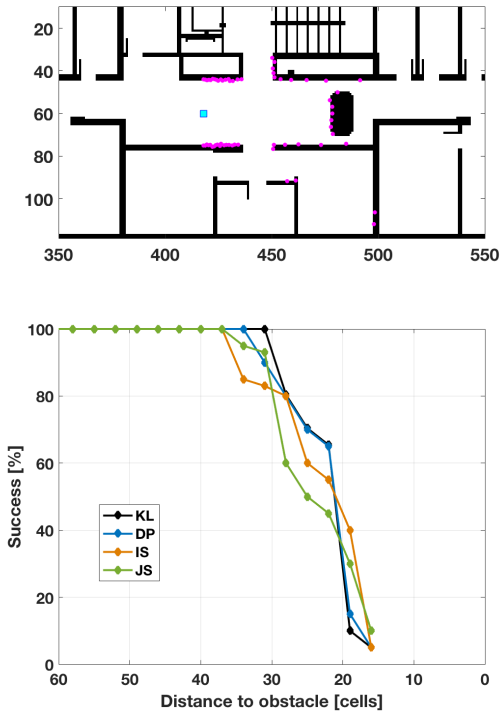

Fig. 9. Occlusions originated by a big obstacle. Top: occlusion caused by obstacle in front of the robot. Units in cells. Bottom: percentage of success vs. distance to obstacle.

In the next experiment, the algorithm is tested when different small objects are added when the robot is situated in a distinguishable place. These obstacles increase the number of occluded measurements. The results are displayed in Figure 10 and Table VII.

The true pose with the real measurements (including obstacles) is in the top row of Figure 10. The estimate in the known map (without obstacles) is in the bottom row. The number of occlusions is equal to 5, 17, and 45 cells in the left, middle, and right columns of the figure, respectively (more cases are included in the table). This number is equal to the number of measurements originated by the unmodeled obstacles. The GL errors and the percentages of success are shown in the table. The limit to consider successful cases has been increased to 1 m in this experiment.

When 37 laser readings out of 61 are wrong measurements caused by occlusions, the correct location is obtained in all cases and the worst error is 7.55 cm. The error that was reported with the quadratic cost function in the same circumstances (with occlusions) was in the interval $[10.57, 11.76]$ cm [4]. The errors of the current metrics are significantly lower in all cases. When the number of occluded measurements was greater or equal to 17, the quadratic-based method failed to reach the solution and the success rate was 0%. After this analysis, it is clear that the probability profiles technique is the optimum option in situations with unmodeled obstacles.

Comparing the results when using different metrics, the error without obstacles is similar to the error with unmodeled obstacles up to 37 occluded measurements. The accuracy is

TABLE VII
GL ERROR WITH DIFFERENT UNMODELED OBSTACLES. TRUE LOCATION: (55, 55, 0). NUMBER OF OCCLUSIONS IN THE LEFT COLUMN. TYPE OF DIVERGENCE IN THE UPPER LEFT CORNER. ERRORS IN MEAN ± STANDARD DEVIATION. SENSOR NOISE: 1%.

KL	Position error (cm)	Orientation error (deg)	Success(%)
0	3.50 ± 0.16	0.04 ± 0.04	100
5	4.43 ± 0.15	0.05 ± 0.05	100
8	0.69 ± 0.04	0.02 ± 0.04	100
17	2.40 ± 0.30	0.15 ± 0.17	100
37	7.55 ± 2.92	0.42 ± 0.18	100
43	11.74 ± 3.38	0.57 ± 0.34	90
45	50.87 ± 1.22	0.54 ± 0.69	18
DP	**Position error (cm)**	**Orientation error (deg)**	**Success(%)**
0	4.32 ± 0.26	0.05 ± 0.06	100
5	2.61 ± 0.16	0.09 ± 0.09	100
8	2.25 ± 0.22	0.06 ± 0.07	100
17	6.33 ± 0.36	0.28 ± 0.17	100
37	3.68 ± 2.10	0.33 ± 0.10	100
43	36.94 ± 13.01	2.08 ± 1.85	26
45	60.75 ± 13.11	1.32 ± 1.07	22
IS	**Position error (cm)**	**Orientation error (deg)**	**Success(%)**
0	4.35 ± 0.18	0.05 ± 0.06	100
5	2.09 ± 0.21	0.03 ± 0.04	100
8	1.90 ± 0.17	0.04 ± 0.03	100
17	2.08 ± 0.12	0.03 ± 0.04	100
37	4.13 ± 4.13	0.35 ± 0.30	100
43	13.49 ± 2.40	0.36 ± 0.27	92
45	13.54 ± 2.81	0.83 ± 0.57	18
JS	**Position error (cm)**	**Orientation error (deg)**	**Success(%)**
0	4.19 ± 0.24	0.03 ± 0.03	100
5	3.55 ± 0.23	0.04 ± 0.04	100
8	1.89 ± 0.17	0.03 ± 0.04	100
17	2.66 ± 0.17	0.03 ± 0.08	100
37	2.45 ± 1.45	0.21 ± 0.11	100
43	12.41 ± 3.26	0.54 ± 0.61	92
45	12.22 ± 3.23	0.70 ± 0.69	22

maintained when there are occlusions. Although the error with the DP divergence is slightly higher when there are 17 occluded readings, it still presents an acceptable value. No significant differences are observed between metrics. For higher levels (43 and 45 measurements), the errors are worse and the success rate decreases. The IS and the JS divergences show better errors ($[12.22, 13.54]$ cm and $[0.36, 0.83]$ degrees). If the success rate is examined, the worst performance is obtained with the DP divergence. The other three metrics present similar values.

D. Population requirements and computational cost

The study of the population requirements of the GL method is relevant because this parameter has an important influence on the computational cost. The population size that is required has been examined for different locations in the simulated map (Figure 3). The experiment consists of measuring the minimum set that is needed to obtain the maximum percentage of success. Only cases with 100% of success are shown. The aim is to check if any divergence needs lower particles than the other ones. The position errors and the required number of particles are displayed in Table VIII. The orientation error is omitted for simplicity.

Although some differences can be appreciated, there is no evidence to choose one metric among the others. In addition,

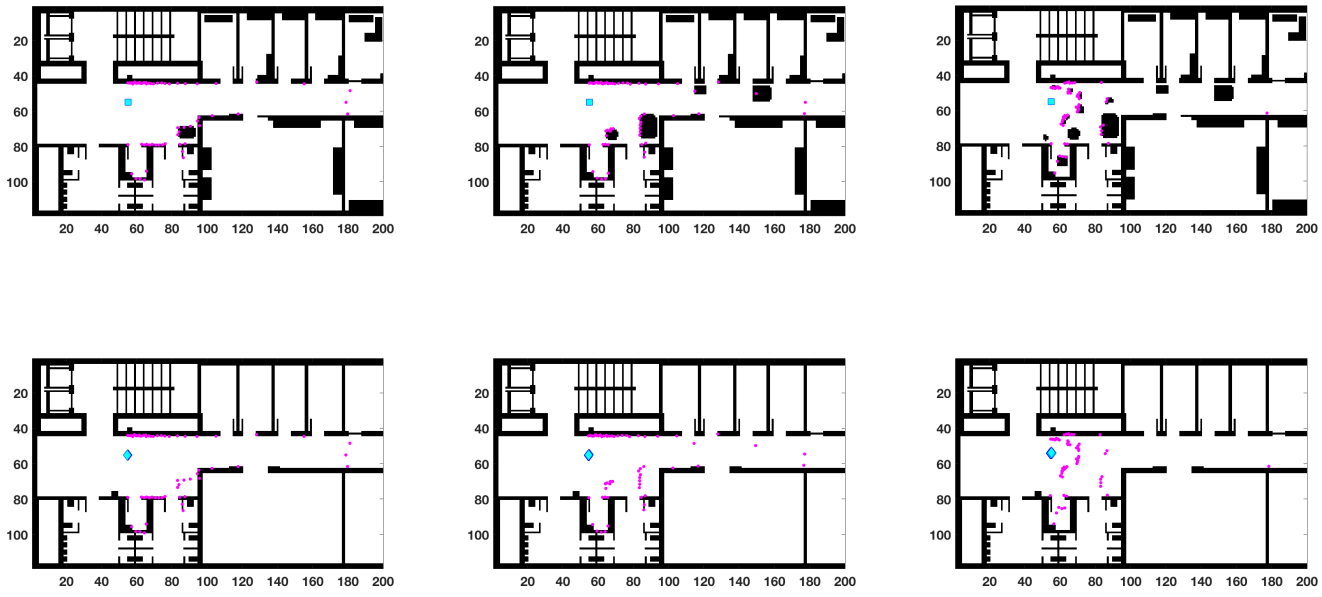

Fig. 10. Occlusions originated by small obstacles. Different obstacles in each column. All units in cells. Top: real laser scan from the true pose. Robot's true pose in blue square. Bottom: estimated pose in the known map. Estimate marked with a blue diamond.

TABLE VIII
OPTIMUM POPULATION SIZE (TOP) AND POSITION ERROR (BOTTOM) WHEN THE ROBOT IS IN THE SIMULATED MAP OF FIGURE 3. ERRORS IN MEAN ± STANDARD DEVIATION (CM).

Robot's pose	KL	DP	IS	JS
(50,60,0)	50	35	55	40
(790,40,90)	200	205	195	201
(700,80,180)	140	165	145	170

Robot's pose	KL	DP	IS	JS
(50,60,0)	4.64 ± 0.84	4.48 ± 1.08	4.84 ± 0.62	4.62 ± 0.88
(790,40,90)	4.63 ± 0.85	3.45 ± 1.05	4.30 ± 1.75	3.42 ± 2.07
(700,80,180)	8.66 ± 10.82	9.65 ± 4.96	12.23 ± 2.60	10.50 ± 5.61

this parameter is highly dependent on different variables such as the location (larger rooms need less particles because it is easier to reach the minimum) or the perceptual ambiguities (symmetries). Therefore, more experiments are needed to present a more exhaustive study of the population requirements, which is a challenging work to be accomplished in the future. This section is intended to be an introduction to the study of this parameter.

The time complexity of the filter is $O(DE\_GL) = N_{iter} \times N_P \times N_s$, where N_{iter} is the number of iterations. This parameter depends on the number of iterations, the population size, and the number of measurements of the laser scan. The computational complexity is dependent on the number of iterations when the algorithm is executed for a fixed number of particles and a laser scan with a constant resolution ($N_P \times N_s$ is constant). The time complexity of other evolutionary-based methods is similar. The MC technique has to be viewed from a different perspective because it integrates the motion information to compute the robot's pose. Many motion-perception cycles are needed to converge. Anyway, the time complexity of the MC-based algorithms is highly dependent on the number of particles. The whole search space (in GL without motion) must be covered to assure convergence to the true solution. Therefore, the population size that is required is larger. Several comparisons between DE and variants of MC are presented in [71], [28]. In both papers, the MC-based method needs more particles to succeed.

An experiment has been carried out to study the parameters that influence the computational cost. The objective is to compare the times per iteration depending on the population size for different locations and maps. The number of iterations that are needed to converge is also presented. Table IX illustrates the results. The success rate is 100% in all cases. The algorithm is implemented in MATLAB in a computer with a 2.7 GHz Intel Core i5 processor.

Different conclusions can be drawn from the table. No significant differences are found when comparing metrics. The cost highly depends on the number of particles. For 50 particles, the time per iteration is in the interval [26, 27] ms. For 100 candidates, the time is [47, 66] ms. For 200 members, the cost is [88, 112] ms.

The iterations to converge depend on the map. In general, larger maps require more iterations. The method needs [362.4, 405.5] iterations for the map of Figure 3 (288,000 cells). When the robot is in the map of Figure 6 (120,000 cells), it needs [188.9, 216.3] iterations. In Figure 4 (360,000 cells), [350.1, 438.3] iterations are required. All metrics show similar numbers. However, a more exhaustive study is necessary to compare the convergence properties.

TABLE IX
TIMES PER ITERATION FOR DIFFERENT LOCATIONS AND MAPS. RESULTS IN MEAN ± STANDARD DEVIATION.

Map/Pose	N_P		KL	DP	IS	JS
Figure 3	200	time (ms)	91 ± 3	88 ± 2	95 ± 4	98 ± 5
(860,60,0)		N_{iter}	382.6 ± 50.2	387.6 ± 42.9	362.4 ± 61.6	418.1 ± 40.6
Figure 3	100	time (ms)	51 ± 3	50 ± 4	53 ± 5	63 ± 9
(860,60,0)		N_{iter}	397.6 ± 61.0	376.1 ± 59.5	405.5 ± 74.6	378.9 ± 70.5
Figure 6	100	time (ms)	50 ± 4	48 ± 4	47 ± 3	50 ± 4
(60,60,0)		N_{iter}	189.9 ± 42.9	212.0 ± 57.1	200.3 ± 37.5	188.9 ± 20.1
Figure 6	50	time (ms)	26 ± 3	26 ± 3	27 ± 2	26 ± 2
(60,60,0)		N_{iter}	209.3 ± 42.5	216.3 ± 41.9	193.1 ± 32.05	192.0 ± 25.6
Figure 4	200	time (ms)	112 ± 6	111 ± 5	110 ± 2	111 ± 2
(350,120,0)		N_{iter}	371.9 ± 63.1	438.3 ± 45.7	424.5 ± 55.96	435.6 ± 40.37
Figure 4	100	time (ms)	61 ± 3	66 ± 5	65 ± 5	63 ± 3
(350,120,0)		N_{iter}	377.5 ± 74.5	388.3 ± 69.24	350.1 ± 59.0	433.6 ± 50.2

E. Tracking performance

The last experiment shows the robot's capabilities when it is executing a path. The typical scenario of a mobile robot involves the continuous reception of laser scans. The robot is navigating and, at the same time, the information provided by sensors is received. The previous experiments have to be considered as a more difficult problem to solve because a single laser scan is the only source of information. In addition, when a single scan is used, it is possible to find places where the laser readings are very similar. For example, this situation happens in environments with symmetric offices. In order to be successful in these cases, motion and more laser scans from different locations are required.

An advantage of the current approach is that some parameters can be relaxed after convergence if the robot's pose has been correctly estimated. The objective of this operation is to improve the computational cost. The population is limited to 20 candidates and the number of iterations is 100 in this experiment (only after convergence, these parameters are not reduced in the first execution). The computational cost in this conditions is in the interval [1.21, 1.32] s. In tracking, the algorithm can be focused on the obtention of a faster response because the GL problem has already been solved. However, the population size and the number of iterations can be modified if the objective is to increase the accuracy.

The path that is followed is displayed in Figure 11. The histograms of the position and orientation errors while the robot is navigating are plotted in Figure 12.

The mobile robot begins situated in $(100, 500, 0)$. The first part of the path covers the hallway that can be seen in the lower part of the image. Then, the aisle on the right side is traversed. The final location, with coordinates $(402, 206, 185)$, and the laser scan that is received from this point are shown in the figure.

In Figure 12, it can be appreciated that the filter obtains the correct location even after the first laser reading (all errors lower than 10 cm). This result has to be viewed as an advantage when compared to other techniques. For example, the MC-based approaches need more motion-perception cycles to return an accurate solution because the sensor information has to be integrated in the motion model.

If the position error is checked, it remains lower than 10 cm for all divergences. This error depends on different factors.

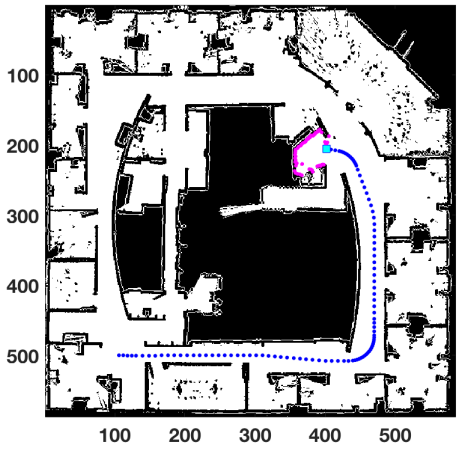

Fig. 11. Robot's path and laser scan from the final location. Starting point: bottom left corner. Units in cells.

Since the sensor noise is dependent on the distance to the obstacles, the error is larger when the robot is situated in larger rooms. The variation of the laser measurements is more significant when the robot is turning. Therefore, the error could be worse in sharp turns. As said before, an easy adjustment that can be easily applied to improve the accuracy (if necessary) is to increase the population size and the maximum number of iterations. The orientation errors are lower than 0.7 degrees for all divergences. If the cost functions are compared, satisfactory results are obtained with all of them and all the errors are in the same interval. Most errors are lower than 6 cm and 0.5 degrees. No significant differences are found between them.

If other methods are analyzed, a similar experiment is presented by Zhang *et al.* [23]. The authors have implemented an algorithm called SAMCL. One of the objectives of their tests is to make a comparison between their technique and MC. The idea is to measure the pose tracking error in an indoor corridor. The authors have reported that their errors are better than the errors of the MC basic version in pose tracking and GL. If their method is compared to the filter presented in the current paper, the main difference can be found when the error of the first motion-perception cycle is analyzed. First, their error is around 80 cm and, after that, it decreases to 20

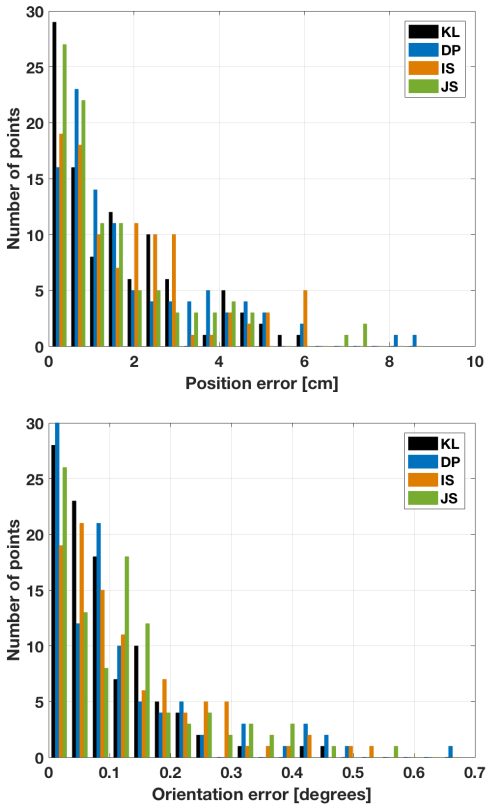

Fig. 12. Histograms of the errors for the path displayed in Figure 11. Top: position error. Bottom: orientation error. Number of points of the path with an error in the interval defined by each bin in the vertical axis.

cm when more scans are received. As can be observed, the performance is worse in the first execution.

VII. Conclusions

In this paper, three different cost functions have been implemented for an evolutionary-based GL filter. The main objectives are to validate that the non-symmetric probabilistic approach presented in our previous work is a suitable technique to be applied to multiple metrics, and to examine the advantages when compared to the classic symmetric quadratic fitness functions. In particular, the most important improvement is the exceptional behavior in environments with different types of occlusions.

Although many different metrics could be selected, the chosen ones present interesting properties. The JS divergence can be viewed as a symmetric version of the KL divergence. The DP divergence is promising because it can be equivalent to a quadratic error version when using probability distributions. The IS distance is similar to the KL divergence when its formula is analyzed. These divergences belong to different families of metrics with interesting properties and connections between them.

The non-symmetric function that is enabled by the probability profiles approach makes it possible to give different weights to the components of the cost function depending on different situations. This property is exploited to increase the robustness to occlusions.

Multiple trials have been carried out in different maps. The GL task is efficiently accomplished by the DE-based algorithm with all metrics. It does not show any particular shortcoming with respect to previous versions with quadratic or KL-based cost values.

The accuracy and the sensor noise have been analyzed in the first tests. The accuracy is similar to the accuracy obtained by our previous method with a quadratic cost function. The level of sensor noise that still allows an optimum percentage of success is worse than the typical noise that is presented by the laser range finders. When comparing metrics, a slightly better behavior is found with the KL divergence when the sensor noise is increased.

The next experiments study the algorithm behavior when there are occlusions originated by different factors. The localization results are strongly improved when there is a contaminated noise. This situation occurs when there are dynamic objects, people, or outliers. It has to be remarked that the authors in [12] highlight that the power functions (used to define the families of divergences) allow to increase the robustness with respect to outliers. Although the best results are obtained when the IS divergence is applied, no significant differences are observed between metrics. The GL performance is also improved when there are different types of unmodeled obstacles, such as small obstacles or a big obstacle in front of the robot. In all cases, the results are improved when there are occlusions. The IS and the JS divergences are the best choices for this situation. It can be concluded that the probability profiles that are defined to deal with occlusions produce a great improvement of the method capabilities in these situations.

The GL algorithm has also been examined when the robot is moving. Analyzing the results, the localization filter presents flexibility (variable number of iterations per perception cycle) and promising results. A clear advantage of our technique when is applied to the GL and tracking problem is that the robot can be successfully localized even with a single scan. All metrics are successfully applied to this task and no significant differences are found between them.

An interesting work to be accomplished in the future is to make an exhaustive study of the convergence properties of each divergence. In addition, experiments with different types of noises could be useful to model other phenomena that can influence the localization process.

Acknowledgment

The research leading to these results has received funding from the *Tuñel* project (Research for the Competitive Improvement of Drilling and Blasting Cycle in Mining and Underground-Works through New Techniques of Engineering, Explosives, Prototypes, and Advanced Tools), which is a R&D project undertaken by the following companies: Obras Subterráneas, MaxamCorp Holding, Putzmeister Ibérica, Subterra Ingeniería, Expace On Boards Systems, Dacartec Servicios Informáticos, and Cepasa Ensayos Geotécnicos.

REFERENCES

[1] R. Storn and K. Price, "Differential Evolution A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of Global Optimization*, vol. 11, p. 341359, December 1997.

[2] L. Moreno, S. Garrido, and M. L. Muñoz, "Evolutionary Filter for Robust Mobile Robot Localization," *Robotics and Autonomous Systems*, vol. 54, no. 7, pp. 590–600, 2006.

[3] F. Martín, L. Moreno, S. Garrido, and D. Blanco, "High-Accuracy Global Localization Filter for three-dimensional Environments," *Robotica*, vol. 30, pp. 363–378, 2011.

[4] F. Martín, L. Moreno, D. Blanco, and M. L. Muñoz, "Kullback-Leibler divergence-based global localization for mobile robots," *Robotics and Autonomous Systems*, vol. 62, no. 2, p. 120130, February 2014.

[5] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951.

[6] F. Martín, L. Moreno, S. Garrido, and D. Blanco, "Kullback-Leibler Divergence-Based Differential Evolution Markov Chain Filter for Global Localization of Mobile Robots," *Sensors*, vol. 15, no. 9, pp. 23 431–23 458, 2015.

[7] I. Csiszár, "Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten," *Publ. Math. Inst. Hungar. Acad.*, vol. 8, pp. 95–108, 1964.

[8] L. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. Math. Phys.*, vol. 7, no. 10, p. 200217, 1967.

[9] S. Amari and H. Nagaoka, *Methods of information geometry*. American Mathematical Soc., 2007, vol. 191.

[10] H. Zhu and R. Rohwer, "Measurements of generalisation based on information geometry," in *Mathematics of Neural Networks*. Springer, 1997, pp. 394–398.

[11] A. Cichocki, R. Zdunek, and S. Amari, "Csiszars divergences for non-negative matrix factorization: Family of new algorithms," in *International Conference on Independent Component Analysis and Signal Separation*, Springer, Ed., 2006, pp. 32–39,.

[12] A. Cichocki and S. Amari, "Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities," *Entropy*, vol. 12, no. 6, pp. 1532–1568, 2010.

[13] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[14] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[15] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proceedings of the 6th International Congress on Acoustics*. Los Alamitos, CA, USA: IEEE, 1968, p. C17C20.

[16] W. Burgard, D. Fox, D. Henning, and T. Schmidt, "Estimating the absolute position of a mobile robot using position probability grids," in *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI'96)*, 1996.

[17] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello, "Bayesian Filters for Location Estimation," *Pervasive Computing*, vol. 2, pp. 24–33, 2003.

[18] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo Localization for Mobile Robots," *Artificial Intelligence*, vol. 128, pp. 99–141, 2001.

[19] C. Gamallo, C. V. Regueiro, P. Quintía, and M. Mucientes, "Omnivision-based KLD-Monte Carlo Localization," *Robotics and Autonomous Systems*, vol. 58, pp. 295–305, 2010.

[20] L. Zhang, R. Zapata, and P. Lepinay, "Self-adaptive Monte-Carlo localization for mobile robots using range sensors," in *Proceedings of the lEEW/RSJ International Conference on Intelligent Robots and System (IROS'09)*, 2009.

[21] J. Biswas, B. Coltin, and M. Veloso, "Corrective Gradient Refinement for Mobile Robot Localization," in *Proceedings of the IEEE/RSJ International Conferenceon Intelligent Robots and Systems (IROS'11)*, 2011.

[22] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, and W. Burgard, "Efficient Estimation of Accurate Maximum Likelihood Maps in 3D," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'07)*, 2007.

[23] L. Zhang, R. Zapata, and P. Lépinay, "Self-adaptive Monte Carlo localization for mobile robots using range finders," *Robotica*, vol. 30, pp. 229–244, 2011.

[24] S. Lenser and M. Veloso, "Sensor resetting localization for poorly modelled mobile robots," in *International Conference on Robotics and Automation (ICRA-2000), San Francisco*, 2000.

[25] S. Ito, F. Endres, M. Kuderer, G. D. Tipaldi, C. Stachniss, and W. Burgard, "W-RGB-D: Floor-plan-based indoor global localization using a depth camera and wifi," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 417–422.

[26] Y. Wang, W. Chen, and J. Wang, "Map-based localization for mobile robots in high-occluded and dynamic environments," *Industrial Robot: An International Journal*, vol. 41, no. 3, pp. 241–252, 2014.

[27] A. R. Vahdat, N. N. Ashrafoddin, and S. S. Ghidary, "Mobile Robot Global Localization using Differential Evolution and Particle Swarm Optimization," in *Proceedings of the Congress on Evolutionary Computation (CEC'07)*, 2007.

[28] M. Lisowski, "Differential Evoution approach to the Localization Problem for Mobile Robots," Master's thesis, Technical University of Denmark, 2009.

[29] Z. Geem, J. Kim, and G. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76(2), pp. 60–78, 2001.

[30] M. Mirkhania, R. Forsatib, M. Shahric, and A. Moayedikiad, "A novel efficient algorithm for mobile robot localization," *Robotics and Autonomous Systems*, 2013.

[31] L. Ronghua and H. Bingrong, "Coevolution Based Adaptive Monte Carlo Localization (CEAMCL)," *International Journal of Advanced Robotic Systems*, vol. 1, no. 3, pp. 183–190, 2004.

[32] C. H. Chien, C. C. Hsu, W. Y. Wang, W. C. Kao, and C.-J. Chien, "Global localization of Monte Carlo localization based on multi-objective particle swarm optimization," in *Consumer Electronics-Berlin (ICCE-Berlin), 2016 IEEE 6th International Conference on*. IEEE, 2016, pp. 96–97.

[33] I. J. Cox and J. J. Leonard, "Modeling a dynamic environment using a Bayesian multi hypothesis approach," *Artificial Intelligence*, vol. 66, pp. 311–44, 1994.

[34] P. Jensfelt and S. Kristensen, "Active Global Localization for a Mobile Robot Using Multiple Hypothesis Tracking," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 748–760, 2001.

[35] P. Pfaff, C. Plagemann, and W. Burgard, "Gaussian Mixture Models for Probabilistic Localization," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA'08)*, Pasadena, CA, USA, May 2008.

[36] G. Jochmann, S. Kerner, S. Tasse, and O. Urbann, "Efficient multi-hypotheses unscented kalman filtering for robust localization," *RoboCup 2011: Robot Soccer World Cup XV (pp. 222-233)*, pp. 222–233, 2012.

[37] L. Moreno, D. Blanco, M. L. Muñoz, and S. Garrido, "L1L2-norm comparison in global localization of mobile robots," *Robotics and Autonomous Systems*, vol. 59, pp. 597–610, 2011.

[38] F. Donoso-Aguirre, J. P. Bustos-Salas, M. Torres-Torriti, and A. Guesalaga, "Mobile robot localization using the Hausdorff distance," *Robotica*, vol. 26, p. 129141, 2008.

[39] D. Fox and W. Burgard, "Active Markov Localization for Mobile Robots," *Robotics and Autonomous Systems*, vol. 25, pp. 195–207, 1998.

[40] K. O. Arras, J. A. Castellanos, and R. Siegwart, "Feature-based multi-hypothesis localization and tracking for mobile robots using geeometric constraints," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02)*, Washington DC, USA, 2002, pp. 1371–1377.

[41] F. Martín, L. Moreno, M. L. Muñoz, and D. Blanco, "Initial population size estimation for a Differential-Evolution-based global localization filter," *International Journal of Robotics and Automation*, vol. 29, no. 3, pp. 245–258, 2014.

[42] D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*. Addison Wesley Publishing Company, 1989.

[43] S. Markon, D. V. Arnold, T. Back, T. Beielstein, and H.-G. Beyer, "Thresholding-a selection operator for noisy ES," in *Proceedings of the Congress on Evolutionary Computation (CEC'01)*, 2001.

[44] T. de Laet, J. Schutter, and H. Bruyninckx, "Rigorously Bayesian Range Finder Sensor Model for Dynamic Environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'08)*, Pasadena, USA, 2008.

[45] D. Fox, W. Burgard, and S. Thrun, "Markov localization for mobile robots in dynamic environments," *Journal of Artificial Intelligence Research*, vol. 11, no. 11, pp. 391–427, 1999.

[46] H. P. Moravec, "Sensor fusion in certainty grids for mobile robots," *AI Magazine*, vol. 9, pp. 61–74, 1988.

[47] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, M. Press, Ed. Cambridge, 2005.

[48] H. Choset, K. M. Lynch, S. Hutchinson, G. A. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*, M. Press, Ed. Cambridge, 2005.

[49] M. Hein and O. Bousquet, "Hilbertian metrics and positive definite kernels on probability measures," in *AISTATS*, 2005, pp. 136–143.

[50] J. Zhang, "Divergence function, duality, and convex analysis," *Neural Computation*, vol. 16, no. 1, pp. 159–195, 2004.

[51] S. Eguchi and S. Kato, "Entropy and divergence associated with power function and the statistical application," *Entropy*, vol. 12, no. 2, pp. 262–274, 2010.

[52] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 131–142, 1966.

[53] M. Mihoko and S. Eguchi, "Robust blind source separation by beta divergence," *Neural computation*, vol. 14, no. 8, pp. 1859–1886, 2002.

[54] H. Jeffreys, "An invariant form for the prior probability in estimation problems," in *Proceedings of the Royal Society of London a: mathematical, physical and engineering sciences*, vol. 186, no. 1007. The Royal Society, 1946, pp. 453–461.

[55] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*. MIT Press, 1999.

[56] I. Dagan, L. Lee, and F. Pereira, "Similarity-based methods for word sense disambiguation," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997, pp. 56–63.

[57] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions,," *IEEE Transactions on Information theory*, 2003.

[58] J. L. W. V. Jensen, "Sur les fonctions convexes et les ingalits entre les valeurs moyennes," *Acta Mathematica*, vol. 30, no. 1, pp. 175–193, 1906.

[59] A. P. Majtey, P. W. Lamberti and D. P. Prato, "Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states," *Physical Review A*, vol. 72, no. 5, p. 052310, 2005.

[60] J. A. Aslam and V. Pavlu, "Query hardness estimation using Jensen-Shannon divergence among multiple scoring functions," in *European Conference on Information Retrieval*. Springer, 2007, pp. 198–209.

[61] P. Gajer, R. M. Brotman, G. Bai, J. Sakamoto, U. M. Schütte, X. Zhong, S. S. Koenig, L. Fu, Z. S. Ma, X. Zhou, Z. Abdo, L. J. Forney, and J. Ravel, "Temporal dynamics of the human vaginal microbiota," *Science translational medicine*, vol. 4, no. 132, pp. 132ra52–132ra52, 2012.

[62] B. Jian and B. C. Vemuri, "A robust algorithm for point set registration using mixture of Gaussians," in *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 2. IEEE, 2005, pp. 1246–1251.

[63] M. Kim and S. Lee, "Estimation of a tail index based on minimum density power divergence," *Journal of multivariate Analysis*, vol. 99, no. 10, pp. 2453–2471, 2008.

[64] Y. Sogawa, T. Ueno, Y. Kawahara, and T. Washio, "Active learning for noisy oracle via density power divergence," *Neural Networks*, vol. 46, pp. 133–143, 2013.

[65] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[66] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *Journal of machine learning research*, vol. 6, no. Oct, pp. 1705–1749, 2005.

[67] S. M. Tabish, M. Z. Shafiq, and M. Farooq, "Malware detection using statistical analysis of byte-level file content," in *Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics*. ACM, 2009, pp. 23–31.

[68] D. L. Olson and D. Delen, *Advanced data mining techniques*, 2008, Ed. Springer Science & Business Media.

[69] S. Se, D. G. Lowe, and J. J. Little, "Vision-Based Global Localization and Mapping for Mobile Robots," *IEEE Transactions on Robotics*, vol. 21, p. 3, 2005.

[70] T. Y. Tsou and S. H. Wu, "A Robust Feature Matching Method for Robot Localization in a Dynamic Indoor Environment," in *Technologies and Applications of Artificial Intelligence*. Springer, 2014, pp. 354–365.

[71] L. Moreno, F. Martín, M. L. Muñoz, and S. Garrido, "Differential Evolution Markov Chain filter for global localization," *Journal of Intelligent and Robotic Systems*, 2015.