Probabilistic Qualitative Localization and Mapping

Roee Mor[1] and Vadim Indelman[2]
[1]Department of Computer Science [2]Department of Aerospace Engineering
Technion - Israel Institute of Technology, Haifa 32000, Israel
roeeki.mor@gmail.com, vadim.indelman@technion.ac.il

***Abstract*— Simultaneous localization and mapping (SLAM) is essential in numerous robotics applications such as autonomous navigation. Traditional SLAM approaches infer the metric state of the robot along with a metric map of the environment. While existing algorithms exhibit good results, they are still sensitive to measurement noise, sensors quality, data association and are still computationally expensive. Alternatively, we note that some navigation and mapping missions can be achieved using only qualitative geometric information, an approach known as qualitative spatial reasoning (QSR). In this work we contribute a novel probabilistic qualitative localization and mapping approach, which extends the state of the art by inferring also the qualitative state of the camera poses (localization), as well as incorporating probabilistic connections between views (in time and in space). Our method is in particular appealing in scenarios with a small number of salient landmarks and sparse landmark tracks. We evaluate our approach in simulation and in a real-world dataset, and show its superior performance and low complexity compared to state of the art.**

I. Introduction

Robotic and autonomous navigation has a lot of impact on state of the art applications in various domains. Image based navigation and specifically simultaneous localization and mapping (SLAM) have a key role in this field. The SLAM problem has been extensively investigated in the past three decades (see [1] for a recent survey of state of the art approaches and challenges). In particular, highly-efficient open source SLAM software packages [2]–[5] have been developed and are gradually incorporated into real-world applications. Lately, the problem of planning under uncertainty and active SLAM, also received considerable research attention (see, e.g. [6], [7]). Here actions are planned to realize a given task while accounting for different sources of uncertainty and considering a SLAM setup.

Some challenges, however, still remain. Firstly, while online performance for passive SLAM can often be achieved, real-time performance for low cost platforms is more challenging. In active planning, complexity is still an obstacle. Secondly, state of the art approaches are mainly based on linearization of the non-linear geometric problem in order to use fast solvers [3], [8]–[10]. Usually a large number of landmarks is used, to enable noise filtering, and outlier removal algorithms. This is another factor in both high complexity, and error accumulation. These approaches usually require an accurate initial guess for the estimated variables, usually achieved using good GPS or IMU sensors, or via accurate image based camera re-sectioning techniques. An overview of basic methods can be found in [11], [12]. Some advanced robust graph optimization techniques that attempt to be resilient, or less sensitive, to outliers are [13]–[18].

A different approach is topological mapping, where the environment is described by a graph that holds distinguishable places in its vertices, and relative attributes in the edges (mostly reachability between places). This problem is also called ”Visual Place Recognition” as the estimation usually handles a discrete set of places in a graph. These approaches usually contain minimal or no geometrical data, and do not fully integrate geometric inference. The advantage is that no continuous geometric estimation is performed, and thus errors are not accumulated, and results are less noise dependent. On the other hand, the lack of geometric constraints in the estimation usually makes the approaches computationally expensive. A survey of topological mapping and place detection approaches can be found in [19].

Qualitative spatial reasoning (QSR) is yet another approach, where relative geometrical qualitative constraints are used for understanding the environment. Recent works include [20], [21] and [22]. In QSR approaches, estimation of the map and robot states is qualitative, and hence, less metrically accurate, but also less noise dependent. This potentially requires less noise filtering and a smaller number of landmarks which can lead to computationally light inference algorithms. Using a small number of most salient landmarks can also improve recognition. These properties, in addition to the insight that many robotic tasks do not require accurate metric navigation, motivate the research reported herein.

In this work we contribute a probabilistic QSR based mapping and localization framework designed to be used for large scale navigation with simple sensors and with low complexity. Our approach enables using a small number of high-quality landmarks unlike metric SLAM which is noise-sensitive and typically exploits many tracked features to average noise. Before stating the specific contributions of this work, we discuss the most relevant QSR approaches.

A. Related Work

Application of QSR to robotic navigation and mapping started in the 90’s. The early work by [23] suggests qualitative localization of a robot in reference to landmarks, given their azimuth ordering in a single view. This approach has been extended in [24], [25] and [26] to include multi-view inference, and some aspects of data association and place re-identification, but not a full SLAM problem. Other methods address qualitative representation of relative ori-

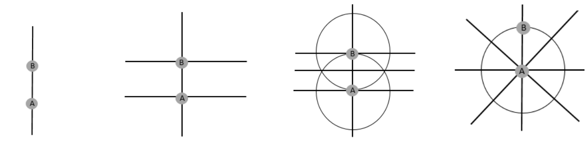


Fig. 1: Different partitions of the metric space: (a) left right [31]; (b) Freska's double cross [29]; (c) Extended double cross [21]; (d) TPCC [30].

entation between two oriented landmarks such as "bipole orientation" [27], and "OPRAM" [20]. Another approach for representing spatial location was suggested by Freksa [28]. A relative frame is built on two landmarks. Instead of considering metric location in this frame, it is represented by a partitioning of the space into a discrete set of qualitative states known as "Freska's single cross". The "Freska's double cross" was suggested in [29], and was recently extended by McClelland et al. [21] to the "extended double cross" (EDC) for a more detailed representation. Other qualitative partitions were also proposed over the years, including a "close and far pie" (TPCC) by [30], and "Left Right" by [31]. These different partitions are illustrated in Figure 1.

The first full QSR based mapping and navigation framework was proposed by McClelland et al. [21] for NASA's planetary rover. They proposed a 2D large scale qualitative algorithm to be used with low quality monocular camera, and no GPS or ego-motion sensing. For each triplet of landmarks observed together, one landmark location is estimated in a local frame defined by the other two using azimuth measurements and range ordering of the landmarks. Instead of considering the metric location of the landmark in this local frame, it is represented by a discrete set of EDC partitioned qualitative states. Geometric estimation classifies each qualitative state to be "feasible" or "non-feasible" in a binary manner. Mapping goal is to associate feasible qualitative states to each landmark triplet. Measurements from different viewpoints can reduce ambiguity. An additional "composition" stage is taken where qualitative data is propagated through triplets with common landmarks, including triplets that were never viewed together. Further work [22] extends this method to probabilistic estimation. Instead of assigning true or false labels, a probability is calculated for each qualitative state of each landmark triplet. [32] takes another step, and addresses qualitative active planning.

However, [21] and [22] do not model any spatial connection between different camera views of the same triplet (such as geometric triangulation). Furthermore, no probabilistic "composition" is addressed. Finally, in both papers the focus is on mapping, and camera pose is not inferred.

B. Contributions

In this work we develop a probabilistic qualitative approach for localization and mapping. We aim for an algorithm used with simple sensors (such as low quality monocular camera, with no GPS or significant IMU), low complexity, and large scale navigation. Low complexity and robustness to error potentially enables a global non-linear solution that is better suited to large scale navigation and does not require prior knowledge. The properties of QSR estimation enable to achieve this at the expense of accuracy. These properties, in addition to the key insight that many robotic tasks do not require accurate metric navigation, drive us to further extend existing QSR approaches. Moreover, we envision the developed approach to be a step towards an active QSR planning framework, leveraging a belief space planning formulation (e.g. [7], [18], [33]).

Our approach uses a different and innovative formulation of the problem to address some of the key limitations and introduces a number of improvements with respect to the state of the art, most notably [21] and [22]. In particular, our *main contributions* are as follows: (a) While existing QSR approaches handle mainly the environment mapping, we develop a full probabilistic QSR approach for concurrent qualitative localization and mapping; (b) We incorporate motion model and line of sight triangulation to improve both results, and complexity; (c) We further develop an approximated algorithm that uses inherent robustness of the qualitative inference to small errors, and is much faster but with similar performance; (d) We develop probabilistic composition for propagating information between different landmark triplets, thereby extending the deterministic approach developed in the seminal work [21], and further improving results; (e) We evaluate the performance of our approach in simulation and in a real-world dataset, and compare it to the state of the art.

The paper is organized as follows. Section II introduces notations and provides problem formulation. Section III describes in detail our approach. It presents a probabilistic formulation of the qualitative localization and mapping problem. Then it presents in detail our proposed algorithm, while also addressing run-time aspects. Lastly it also includes the derivation of our probabilistic composition technique. Section IV provides performance evaluation, while Section V concludes the discussion and suggests several avenues for future research. Additional details and result analysis can be found in [34].

II. Notations and Problem Formulation

We consider a robot navigating in and mapping an unknown 2D environment. As the robot moves, it tracks a few high quality landmarks across different image frames. Our goal is to *qualitatively* describe the environment, and camera trajectory. Motivated by the approach in [21], [22], we consider multiple landmark-centric frames, and use qualitative states to describe the geometry of landmark triplets, and corresponding camera poses.

For each three landmarks A, B, C observed together, we set a landmark-centric local frame so that landmark A is located in $L^A = (0, 0)$ and landmark B in $L^B = (0, 1)$. The metric location of landmark C in this frame is $L^{AB:C}$. Alternatively, we use a qualitative partition of the space (see Figure 1), and associate C to a discrete set of m qualitative states. The qualitative state $S^{AB:C}$ is a vector of dimension $m \times 1$ that contains hypothesis for C being in each qualitative state. The event that a qualitative state

$S^{AB:C} = i; \in \{1, ..., m\}$ is denoted as $s_i^{AB:C}$. Environment map is then described as the collection of all observed landmark triplets (i, j, k), and their qualitative states $\{S^{i,j:k}\}$ (see Figures 2a and 2b).

Similarly, the metric location of the camera in the AB frame at time step n is denoted by X_n^{AB} and the corresponding camera qualitative state is denoted by $S_n^{AB:X}$. The camera qualitative location at time step n is then described as the collection of states in all local frames: $S_n^X = \{S_n^{i,j:X}\}$.

The partition of the space into a set of qualitative states can be chosen freely to fit different tasks, platforms or scenarios. In our implementation and tests we use the EDC partition that was used in previous works [22], [21].

The true mapping and localization qualitative states ($S^{ij:k}$ and $S_{1:n}^{i,j:X}$) are unknown; In this work we infer them within a Bayesian framework using a set of measurements $Z_{1:n}^{ABC} = \{Z_j^{ABC}; j \in \{1, ..., n\}\}$ from time steps $1, ..., n$. Z_j^{ABC} denotes an observation of landmarks A, B, C from time instant j. We note that while, for simplicity, this notation suggests observations are assumed to exist for all time instances $[1, n]$, in practice our method does not require this assumption, as further elaborated in Section III-E.

Herein, we also introduce an innovative usage of a motion model within a qualitative formulation, which was not incorporated in previous work. As will be seen, this enables us to use stronger geometric constraints to improve estimation. Thus, we assume some action a_{n-1} is given for moving the camera between time instances $n-1$ and n.

Our goal at time instant n is to estimate the posterior probabilities of landmark C and camera qualitative states, given history $H_n^{ABC} \doteq \{Z_{1:n}^{ABC}, a_{1:n-1}\}$:

$$\mathbb{P}(S^{AB:C}|H_n^{ABC}) \quad , \quad \mathbb{P}(S_{1:n}^{AB:X}|H_n^{ABC}). \tag{1}$$

The formulation in this paper is general, and any measurement and motion model can be taken (although we use the Markov assumption). In our implementation and results we consider 2D coordinate systems, and a monocular camera setup. Therefore $L^{AB:C} = (x^{AB:C}, y^{AB:C})$ and $X_n^{AB} = (x_n^{AB:X}, y_n^{AB:X}, \alpha_n^{AB:X})$, where α is the camera orientation angle. Measurements are bearing angles to landmarks A, B and C, i.e. $Z_n^{ABC} = \{\phi_n^A, \phi_n^B, \phi_n^C\}$.

For this setup we assume a Gaussian measurement model, which is given for $i \in \{A, B, C\}$ by

$$\mathbb{P}(\phi_n^i|L^i, X_n) \propto \exp\left\{-\frac{1}{2}\|\phi_n^i - f(L^i, X_n)\|_{\Sigma_v}^2\right\}, \tag{2}$$

where Σ_v is the measurement noise covariance and

$$f(L^i, X_n) \doteq \text{atan}_2(y^i - y_n^X, x^i - x_n^X). \tag{3}$$

Considering bearing measurements to different landmarks statistically independent, the joint likelihood for Z_n^{ABC} is readily obtained as a product of individual likelihood terms (2) for each bearing measurement $\phi_n^i \in Z_n^{ABC}$.

In our setup, we also consider a simple motion model that can be used with a monocular camera, and basic azimuth keeping control. Specifically, we consider the robot is moving along a specific heading $a_n = \psi_n$ relative to the previous camera frame and assume Gaussian noise as in

$$\mathbb{P}(X_n^{AB}|X_{n-1}^{AB}, a_{n-1}) \propto \exp\left\{-\frac{1}{2}\|a_{n-1} - g(X_n^{AB}, X_{n-1}^{AB})\|_{\Sigma_w}^2\right\}, \tag{4}$$

where Σ_w is the motion (process) noise covariance and $g(.)$ is defined, similarly to Eq. (3), as $g(X_n, X_{n-1}) \doteq \text{atan}_2(y_n^X - y_{n-1}^X, x_n^X - x_{n-1}^X)$. We note the motion model (4) does not constrain the camera orientation. As will be seen, incorporating a motion model, even as simple as this, leads to a much better qualitative state estimation.

III. Approach

In this section we present our probabilistic qualitative localization and mapping approach. First, we formulate the probabilistic inference of camera and landmark triplet qualitative states with multiple views, incorporating a motion model. Next, we present our implementation and how it uses the advantages of our approach. Finally, we derive a novel probabilistic composition algorithm for propagating information between different landmark triplets.

As specified in Section II, our approach considers multiple landmark-centric triplet frames. For each local frame, the underlying metric problem is actually a SLAM problem determining camera poses and landmark C location, given a set of noisy measurements $\mathbb{P}(X_{1:n}^{AB}, L^{AB:C}|Z_{1:n}^{ABC})$. It is well known that incorporating a motion model, makes the problem easier to solve, and requieres less measurements. This insight is also valid for the qualitative problem. Therefore we expect incorporating a motion model will enable solving the qualitative problem better and with less measurements and prior knowledge (demonstrated by results in section IV-A).

A. Probabilistic Formulation

In this section we focus on a single triplet of landmarks A, B and C viewed together at time steps $1, \ldots, n$. Given the motion model (4) for each camera transition, we want to infer the posterior probabilities (1) of landmark C qualitative state, and camera qualitative trajectory, both in the AB frame. To reduce clutter, in this section we drop the superscript AB notation, as long as everything is in the AB frame. Also, for easier explanation, instead of looking at $\mathbb{P}(S^C|H_n)$ we look at the separate components of this random vector: $\mathbb{P}(s_i^C|H_n)$ (i.e. the probabilities of landmark C to be in each qualitative state separately). Generalizing to $\mathbb{P}(S^C|H_n)$ is trivial.

We start by marginalizing over the metric camera poses and landmark locations, writing the belief over s_i^C as:

$$\mathbb{P}(s_i^C|H_n) = \iint\limits_{X_{1:n}, L^C} \mathbb{P}(s_i^C, X_{1:n}, L^C|H_n) dL^C dX_{1:n}. \tag{5}$$

We shall now apply chain rule. Note that L^C uniquely determines s_i^C so that $\mathbb{P}(s_i^C|L^C) = 1$ for $L^C \in s_i^C$ and 0 else. Since $\mathbb{P}(s_i^C|L^C)$ is independent of any other history and can also be replaced by the corresponding integration range, we get

$$\mathbb{P}(s_i^C|H_n) = \iint\limits_{L^C \in s_i^C, X_{1:n}} \mathbb{P}(X_{1:n}, L^C|H_n) dL^C dX_{1:n}. \tag{6}$$

The camera qualitative state can be similarly inferred:

$$\mathbb{P}(s_i^{X_i}|H_n) = \iint_{X_i \in s_i, X_{1:n/i}, L^C} \mathbb{P}(X_{1:n}, L^C|H_n) dL^C dX_{1:n}. \quad (7)$$

We get an intuitive result: solve the corresponding SLAM problem, $\mathbb{P}(X_{1:n}, L^C|H_n)$, in the AB frame and marginalize over camera trajectories, and over L^C from the relevant qualitative state.

This approach is very different from previous works [22], [21], and has several advantages: (i) Summing over qualitative states also can be trivially adjusted to use any qualitative space partition (see Figure 1); (ii) While we choose a specific method, solving the *small* metric SLAM problem can be done using any existing method or code to fit different applications or scenarios

Using a standard SLAM formulation, the problem can be broken down into simpler factors. We use the assumptions that measurement model $\mathbb{P}(Z_n|X_n, L^C)$ is independent of history, and that the motion model has Markov property. The result of such decomposition is:

$$\mathbb{P}(X_{1:n}, L^C|H_n) = \frac{\mathbb{P}(Z_1|X_1, L^C)\mathbb{P}(X_1, L^C)}{\mathbb{P}(Z_1)} \quad (8)$$
$$\prod_{i=2}^{n} \frac{1}{\zeta_i} \mathbb{P}(Z_i|X_i, L^C)\mathbb{P}(X_i|X_{i-1}, a_{i-1}),$$

where $\zeta_i \doteq \mathbb{P}(Z_i|a_{1:i-1}, Z_{1:i-1})$. Using this decomposition, we can see how both measurement and motion models can be used to solve the underlying SLAM problem.

Recalling Eqs. (6) and (7), we can see that estimating required qualitative states can be done by integrating over camera trajectories, or landmark locations.

B. Algorithm Design Considerations

Standard SLAM approaches usually address big bundle adjustment problems with many landmarks and camera poses. It is hard to achieve a global solution for such problems. These approaches therefore model the estimated variables as multi-variate Gaussian, resort to linearization and find only locally optimal solutions. Moreover, these approaches require enough measurements for initial geometric estimation, or prior knowledge.

Alternatively, we solve many small landmark-centric SLAM problems with 3 landmarks, and a few camera poses at a time. We want to be able to work with a small number of measurements, and no prior knowledge. Under these conditions, we choose a sampling based approach to solve the *global* non-linear *small* SLAM problem. This approach avoids linearization, and gets a global robust solution. Also, integration over coarse resolution qualitative states compensates for some of the sampling error.

For a large SLAM problem this approach can be computationally expensive, but for many small problems, it is feasible. Furthermore, the fact that we use a motion model allows us to test consistency of samples for motion model and geometric camera line of sight triangulation constraints. Practically, this drastically reduces valid samples for two or more views, making this method fast.

C. Detailed Algorithm

We now describe the specific algorithm we use for estimating each single landmark triplet. Usually, a camera-centric global frame is used for solving SLAM problems. We use a different, landmark-centric frame by fixing landmarks $A = (0, 0)$ and $B = (0, 1)$, as we believe it is more intuitive to infer qualitative landmark-related states. [35] suggests a full and efficient solution to this problem geometry.

First, we introduce a few notations. While the random variable X_n represents camera pose at time n, we denote *samples* of this camera pose as $X_n^{(k_n)}$, where n and k_n are, respectively, time and sample indices. We also define a "trajectory hypothesis" $th_n^j \doteq \{X_1^{(k_1)}, \ldots, X_n^{(k_n)}\}$ as a set of specific camera pose samples - one for each time step. Index $j \mapsto \{k_1, \ldots, k_n\}$ is a simplification for the collection of specific sample indices for all n time steps. With a slight abuse of notation, we denote X_l^j as the camera pose sample from time l in th_n^j. The set of all trajectory hypotheses at time n is $TH_n \doteq \{th_n^j\}$.

To estimate a landmark triplet qualitative state with measurements from n time steps we iterativly apply a three-stage algorithm for each time step:

Sampling step: Generate camera pose samples $X_n^{(k_n)}$, with $k_n \in [1, m_n]$, from the distribution $\mathbb{P}(X_n|\phi_n^A, \phi_n^B)$ using bearing measurements ϕ_n^A and ϕ_n^B to landmarks A and B. Given these measurements, the 2D camera pose is determined on a specific part of a *circle* that goes through the two landmarks A,B (see Figure 3a). The *locus circle* parameters can be calculated as specified in [35] and [34]. Camera poses are sampled in the vicinity of this locus circle considering the noisy nature of ϕ_n^A and ϕ_n^B.

Motion step: For each trajectory hypothesis $th_{n-1}^j \in TH_{n-1}$, we can use X_{n-1}^j and motion azimuth ψ_{n-1} to intersect the locus circle from time n (see Figure 3a). Camera pose samples $X_n^{(k_n)}$ that are consistent with this intersection are found, also considering the noisy nature of the motion model (4). Using these matches we generate multiple new, extended trajectory hypotheses th_n^j. For each th_n^j, we also calculate and keep motion model consistency wight: $wm_n^j = \mathbb{P}(X_n^j|X_{n-1}^j, \psi_{n-1})$.

Resection step: For each valid trajectory hypothesis th_n^j, we test consistency of bearing measurements from all cameras to landmark C. First, we use camera poses $X_{1:n}^j$ and bearing measurements $\phi_{1:n}^C$ to triangulate landmark C location, denoted by $L^{C,j}$. It is estimated to be the centroid of all line of sight pairs intersection points (see Figure 3a). Then we estimate the probability for this configuration as: $wr_i^j = \mathbb{P}(\phi_i^C|L^{C,j}, X_i^j), \forall i \in \{1...n\}$ (assuming independent measurement noise). We disqualify trajectories that do not intersect, or have low probability.

Remark: Using a single $L^{C,j}$ for each th_n^j is a heuristic for $n > 1$. We can sample over the area of line of sight intersections to cover all probable L^C locations. We choose the simpler way for run-time considerations.

If we have measurements from only one time step, we sample landmark C location L^C from $\mathbb{P}(L^C|X_1^{(k1)}, \phi_1^C)$ for each sampled camera pose $X_1^{(k1)}$.

Thus far, the results of our algorithm so far are the set of valid trajectory hypotheses $th_n^j \in TH_n$. For each , th_n^j we keep consistency weights wm_i^j and wr_i^j $\forall i \in \{1...n\}$, and a single landmark location $L^{C,j}$.

Finally, we approximate Eqs. (6) and (7) by summing over each qualitative state to get state probability distribution (see the similarity to (8)):

$$\mathbb{P}(s_k^C|H_n) \approx \eta^C \sum_j \mathbb{1}(L^{C,j} \in s_k^C) wr_1 \prod_{i=2}^{n} wm_i^j wr_i^j, \quad (9)$$

$$\mathbb{P}(s_k^{X_i}|H_n) \approx \eta^{X_i} \sum_j \mathbb{1}(X_i^j \in s_k^{X_i}) wr_1 \prod_{i=2}^{n} wm_i^j wr_i^j, \quad (10)$$

where η^C and η^{X_i} are normalization constants, and the sum is over all trajectory hypotheses $th_n^j \in TH_n$. Note the term $wr_1 \prod_{i=2}^{n} wm_i^j wr_i^j$ is a sampled approximation of the joint pdf $\mathbb{P}(X_{1:n}, L_C|H_n)$.

This algorithm seems to handle a large number of trajectory hypotheses that grows exponentially in time. Practically, the geometric constraints enforced in the "motion" step and the "resection" step dramatically decrease the number of hypotheses for two views, and even more for three views or more. This effect of incorporating a motion model to the qualitative estimation makes our algorithm much faster. It also means that using this algorithm incrementally, requires to save only a small number of hypotheses, and therefore is not memory intensive. A pseudo-code for a simplified version of this approach is given in Algorithm 1.

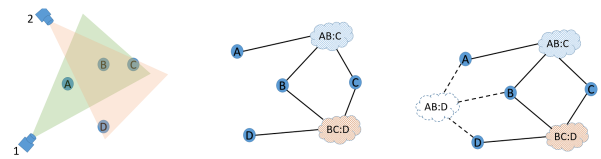
Algorithm 1 Single triplet qualitative state estimation

```
sample camera poses X_1^(k1) ~ P(X_1 | φ_1^A, φ_1^B) with k_1 ∈ [1, m_1]
initialize trajectory hypothesis set: TH = {X_1^(k1)}
for i = 1, ..., n do
    // sampling step:
    sample camera pose X_i^(ki) ~ P(X_i | φ_i^A, φ_i^B) with k_i = [1, m_i]
    // motion model step:
    - ∀th_{i-1}^j find X_i^j that are consistent with motion azimuth ψ_{i-1}
    - extend TH_i with new matches
    - keep consistency weight: wm_i^j = P(X_i^(ki) | X_{i-1}^j, ψ_{i-1})
    // resection step:
    - ∀th_i^j find L^{C,j} by line of sight triangulation
    - keep consistency weight: wr_t^j = P(φ_t^C | L^{C,j}, X_t^j), ∀t ∈ {1...i}
    - dismiss low probability hypotheses
end for
estimate landmark qualitative state probability via Eq. (9)
estimate camera qualitative state probability via Eq. (10)
```

Remark: This algorithm is simplified for the sake of explanation. Our actual implementation is more efficient.

D. Faster Variant

We seek to further utilize the ability of the qualitative coarse resolution to absorb errors, aiming to achieve a faster algorithm. With this motivation in mind, we introduce an approach similar to the full sampling based approach described above, but sampling is done only to represent the prior term in Eq. (8). Motion and measurement noise are neglected. In Section IV-A we evaluate this algorithm variant to see how much noise-related errors it can handle.



(a) **(b)** **(c)**

Fig. 2: qualitative map: (a) Landmarks A,B,C observed from camera pose 1. Landmarks B,C,D observed from camera pose 2. (b)qualitative map is represented as landmark triplet graph (c) Composition: try to estimate triplet AB:D given only AB:C and BC:D estimations (for unobserved triplets, or improving existing estimations).

As we shall see, it can handle reasonable levels of noise with accuracy similar to the full algorithm and with a significant speedup. This is a demonstration of how we can use the coarse qualitative resolution to cover for fast approximations.

E. Spatial Data Propagation - Probabilistic Composition

Until now, we used inference with information from different times. In [21], it is noted that if two landmark triplets share two common landmarks, information can be propagated from one to the other. This "composition" operation enables to enhance estimation for existing overlapping triplets, and to infer triplets that were not viewed together at all (see Figure 2). However, in [21] the estimation only refers to binary evaluation of EDC states.

In this section, we develop a probabilistic composition operator. We infer the posterior probability $\mathbb{P}(S^{AB:D}|H_n^{AB:C}, H_n^{BC:D})$ of the qualitative state $S^{AB:D}$, using history only for triplets $AB:C$ and $BC:D$.

For the sake of readability, we simplify notations in this section. We drop the time index n from notations. All history is taken into account. We also use shortened notations for triplets. $AB:C, BC:D, AB:D$ are denoted as $p1, p2$ and t, respectively (i.e. prior 1,2 and target). So for example: qualitative states notations $s_i^{AB:C} \equiv s_i^{p1}$.

First, we marginalize over the known qualitative states $s_i^{p1}, s_j^{p2}; \forall i, j \in \{1, ..., m\}$:

$$\mathbb{P}(S^t|H^{p1}, H^{p2}) = \sum_{s_i^{p1}} \sum_{s_j^t} p(S^t, s_i^{p1}, s_j^{p2}|H^{p1}, H^{p2}).$$

Using the formula of total probability, and considering proper dependencies we get:

$$\begin{aligned} &\mathbb{P}(S^t, s_i^{p1}, s_j^{p2}|H^{p1}, H^{p2}) = \\ &= \mathbb{P}(S^t|s_i^{p1}, s_i^{p2}, H^{p1}, H^{p2})\mathbb{P}(s_i^{p1}|H^{p1})\mathbb{P}(s_j^{p2}|H^{p2}). \end{aligned} \quad (11)$$

The factors $\mathbb{P}(s_i^{p1}|H^{p1}), \mathbb{P}(s_j^{p2}|H^{p2})$ are the current estimations. The full solution for $\mathbb{P}(S^t|s_i^{p1}, s_j^{p2}, H^{p1}, H^{p2})$ can be obtained through marginalization over the metric locations of landmarks C, D in the relevant triplet frames:

$$\begin{aligned} &\mathbb{P}(S^t|s_i^{p1}, s_j^{p2}, H^{p1}, H^{p2}) = \\ &= \iint_{L^{p1}, L^{p2}} \mathbb{P}(S^t, L^{p1}, L^{p2}|s_i^{p1}, s_j^{p2}, H^{p1}, H^{p2}) dL^{p1} dL^{p2}. \end{aligned}$$

Now, using the formula of total probability, and considering dependencies we get:

$$\mathbb{P}(S^t, L^{p1}, L^{p2} | s_i^{p1}, s_j^{p2}, H^{p1}, H^{p2}) \\ = \mathbb{P}(S^t | L^{p1}, L^{p2}) \mathbb{P}(L^{p1} | s_i^{p1}, H^{p1}) \mathbb{P}(L^{p2} | s_j^{p2}, H^{p2}). \tag{12}$$

Note that $\mathbb{P}(L^{p1}|s_i^{p1}, H^{p1})$ and $\mathbb{P}(L^{p2}|s_j^{p2}, H^{p2})$ are again, metric problems with a prior on landmark location, speficied by the conditioned qualitative state.

approximation: In order to avoid solving these complex problems multiple times in map update, we introduce an approximation. We reduce the full history to the estimated qualitative states: $\mathbb{P}(L|s_i, H) \approx \mathbb{P}(L|s_i)$. These probabilities are uniform for all landmark locations in the relevant qualitative state, and 0 otherwise. We can express them using equivalent integration range. Finally we get:

$$\mathbb{P}(S^t|H^{p1}, H^{p2}) \approx \sum_{s_i^{p1}} \sum_{s_j^t} \mathbb{P}(s_i^{p1}|H^{p1}) \mathbb{P}(s_j^{p2}|H^{p2}) \cdot \\ \iint_{L^{p1} \in s_i^{p1}, L^{p2} \in s_j^{p2}} \mathbb{P}(S^t|L^{p1}, L^{p2}) dL^{p1} dL^{p2}. \tag{13}$$

Note that $L^{AB:D} = f(L^{AB:C}, L^{BC:D})$ is a simple geometric function, and then $\mathbb{P}(S^t|L^{p1}, L^{p2}) = \mathbb{1}(L^{AB:D} \in s_k^{AB:D})$. Furthermore, the double integral is *independent* of $\mathbb{P}(s_i^{p1}|H^{p1}), \mathbb{P}(s_j^{p2}|H^{p2})$, and can be calculated *offline* for all combinations of $S^{p1} = i, S^{p2} = j; \forall \{i, j\}$. This gives us a very low complexity probabilistic composition algorithm.

F. Single View Estimation

The baseline for evaluating our performance is the previous work of [22] (mapping only). We do not directly implement [22], but an equivalent algorithm; we use our formulation from Section III-A, but estimate each view separately, and assume the views are independent. Thus, for the nth view we get $\mathbb{P}(s_i^C|Z_n) = \frac{1}{\mathbb{P}(Z_n)} \iint_{X_n, L_C \in s_i^C} \mathbb{P}(Z_n|X_n, L_C) dL_C dX_n$. The approach in [22] estimates multiple views assuming independence, and not using motion model or triangulation:

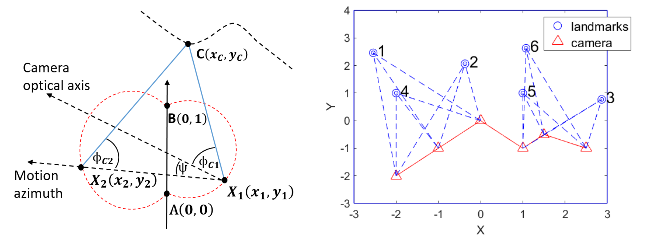
$$\mathbb{P}(S^C|Z_{1:n}) = \frac{\mathbb{P}(S^C|Z_n)\mathbb{P}(S^C|Z_{1:n-1})}{\sum_{s=S^C} \mathbb{P}(S^C = s|Z_n)\mathbb{P}(S^C = s|Z_{1:n-1})},$$

where $\zeta = \sum_{s=S^C} \mathbb{P}(S^C = s|Z_n)\mathbb{P}(S^C = s|Z_{1:n-1})$ is a normalization factor.

IV. Results

We evaluate our approach in a number of steps. We first use simulation, to compare both our full algorithm, and our fast approximation to previous work [22]. Then we also use MCRLAM dataset [36] for testing our approach in a more realistic scenario. Finally, we again use simulation to estimate basic attributes of probabilistic qualitative composition.

Our simulation considers a 2D scenario with point-landmarks and a mobile camera. All 2D locations are limited to $x \in [-3, 3]$ and $y \in [-3, 4]$. To keep the analysis general, we consider locations and poses as uniform as possible: Landmark locations are sampled uniformly; Camera trajectory is a set of uniformly sampled camera poses. For each pose, only 3 landmarks are observed. This simulation meant to avoid biasing results with a specific assumptions on camera trajectory or landmark visibility. A simple example for a simulation scenario is given in Figure 3b.



(a) **(b)**

Fig. 3: (a) Camera resection and landmark triangulation with two measurements. (b) Example simulation scenario.

We randomize numerous scenarios, with multiple noise levels. Measurements and motion commands are generated according to Gaussian models described in Section II, with $\sigma_v \in [0°, 10°]$, and $\sigma_w \in [0°, 20°]$ respectively.

We also note that landmark recognition is ideal, and does not model false identification. Identification error outliers can be modeled in our approach, but here we want to evaluate the plain qualitative geometry, and therefore do not consider these aspects in this study.

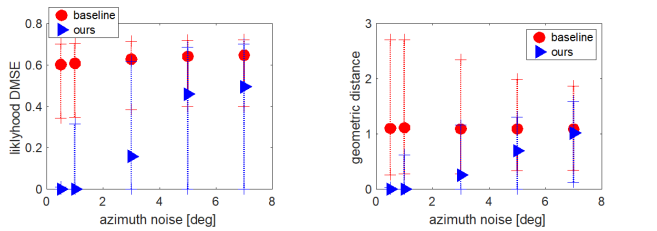
To measure performance we use several metrics: (i) Probability DMSE: mean square error for the difference between estimated qualitative state probability vector and ground truth (GT) $DMSE = \sqrt{\sum_{i=1}^{m} (\mathbb{P}(s_i) - \mathbb{P}(s_{GT}))^2}$. It tests estimation accuracy. (ii) *GT rating*: Position of the GT qualitative state when ordering qualitative states by their estimated likelihood (1 is most likely) (iii) geometric distance: Mean distance from qualitative state centroids to GT state centroind weighted by their estimated probability, $gmd = \sum_{i=1}^{m} \mathbb{P}(s_i)||c_i - c_{GT}||_2$. It measures if estimated states are spatially close to GT (1 is the distance between A,B). (iv) Estimation entropy $e = -\sum_{i=1}^{m} \mathbb{P}(s_i) log(\mathbb{P}(s_i))$. Using these various metrics (along with others that are not shown here) helps getting a clearer and deeper understanding of the algorithm value than was done so far.

A. EDC landmark Triplet Estimation with motion model

To evaluate the effect of our motion model in-cooperated inference, we compare three different algorithms: (a) Baseline algorithm: An equivalent to [22] without motion model, as described in Section III-F. (b) Our full multi-view estimation described in Section III-A. (c) Our faster approximated inference as described in Section III-A.

For this evaluation we randomize 300 scenarios $\times 36$ different noise levels. In Figure 4 we show a set of scenarios with 5 different levels of measurement noise. Motion model noise was taken to be $\times 2$ correspondingly. In this test, we use landmark triplet estimations with 3 time steps, as shorter trajectories will not use the advantage of our contribution (and therefore hinder the comparison), while longer trajectories typically do not improve results significantly. More results can be found in supplementary material [34].

We observe that using a motion model greatly improves performance. Up to measurement noise of 2° (which is

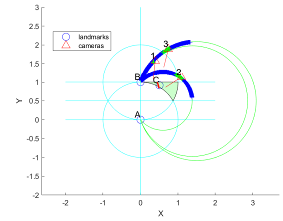


(a) **(b)**

Fig. 4: EDC estimation with multi-view motion model and triangulation. The plot shows median and percentiles 25 and 75 for each algorithm. (a) DMSE Vs measurement noise. (b) mean geometric distance vs. measurement noise.

single triplet EDC estimation results	baseline	ours	ours-fast
DMSE	0.39, 0.63, 0.71	**0, 0.16, 0.63**	0, 0.21, 0.62
geometric distance	0.28, 1.10, 2.30	**0, 0.25, 1.15**	0, 0.27, 1.16
Entropy	0.28, 0.66, 0.87	**0, 5e-3, 0.58**	0, 0.07, 0.64
time[sec]	26	18	**0.05**

TABLE I: Single triplet qualitative state results for bearing measurement noise $\mu_v = 3°$ and motion model noise $\mu_w = 7°$. For each algorithm, and metric we show: 25 percentile, median, 75 percentile. For run time, only median time is shown.

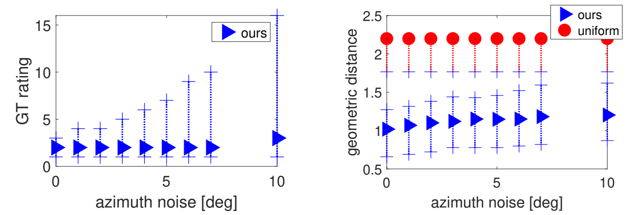


(a)

MRCLAM dataset EDC estimation	ours-fast	uniform
DMSE	0.03, 0.45, 0.69	0.97
gmd	5e-3, 0.27, 0.71	2.2
Entropy	4e-3, 0.38, 0.69	3
GT rating	1, 1, 2	-

(b)

Fig. 5: MRCLAM dataset: (a) example EDC scenario result. We can see camera locus circle with estimated camera poses, and landmark C GT state with estimated location hypothesis. All estimations are close to GT, and error is still inside GT qualitative state. (b) Results summary - our approximated fast algorithm is compared to an uninformative uniform estimation. For each metric we show: 25 percentile, median, 75 percentile.



(a) **(b)**

Fig. 6: EDC probabilistic composition. The plot shows median and percentiles 25 and 75 for each algorithm. (a) DMSE vs. measurement noise. (b) mean geometric distance vs. measurement noise. We comare our composition to uninformative estimation.

reasonable to camera based platforms), results are much better than state of the art, and almost perfect. Up to 7° (which is a large error for camera based systems), our estimated probability *DMSE* is still better than state of the art. We can also see that for 7° *gmd* < 1 which means that all likely qualitative states are close to GT state (demonstrated in Figure 4 and for a specific noise level in Table I).

Another important result is that our fast approximated algorithm achieves performance very close to the full one. While our full algorithm is about two time faster than the baseline, approximation is about 100 times faster than both (Demonstrated for a specific noise level in Table I). This is a successful usage of the ability of coarse qualitative estimation to absorb errors and enable fast approximations. While we use a simple MATLAB implementation, absolute run times are irrelevant, but good for comparing algorithms.

B. MRCLAM dataset qualitative estimation

To evaluate our method in a realistic scenario, we use the *MRCLAM* dataset [36], which comprises several scenarios of multiple robots moving around pre-set landmarks. We choose this dataset since it has robot pose GT as well as landmark identification and location GT. This real-data dataset allows not only to evaluate how informative our qualitative estimation framework is (e.g. in terms of entropy), but also to quantify performance with respect to GT, which has not been tackled in previous work [22].

We use 5 scenarios, each with 5 robots, and 15 landmarks. We get 16-230 landmark triplets that are observed more than 3 times in each scenario. Our qualitative approach thus generates rich mapping for these scenarios.

A typical estimation result for a single landmark triplet observed by three cameras is shown in Figure 5a. As seen, the usage of motion model reduces the trajectory hypotheses TH (blue color), which is further refined by the landmark C triangulation step to only a small subset (green color). The resulting landmark C location hypotheses are shown in red. One can observe the valid camera and landmark hypothesis are very close to GT. As a result, in this example run, the likelihood of the GT qualitative state is 1.

Results for all the observed landmark triplets in the above 5 scenarios are summarized in Table 5b. We compare results to an uninformative uniform distribution. *Entropy* and *DMSE* show that estimation is informative, and not too far from the truth. *GT rating* metric shows that most of the time the GT state is the most likely, and mostly it is within the top 2. Also, looking at *gmd* we can see that all likely qualitative states are close to GT. In conclusion, we note that results show meaningful and informative estimation.

C. Composition

We now evaluate the composition operator on a previously un-estimated triplet, to test how informative this process is. We don't discuss composition effect on existing estimation. Specifically, we simulate 1000 scenarios comprising three inter-connected landmark triplets (AB:C, BC:D, AB:D) and camera trajectories. Measurements Z^{ABC} and Z^{BCD} are available but not Z^{ABD}. We estimate $S^{AB:C}, S^{BC:D}$ using our approximated method, and then infer the $S^{AB:D}$ only through composition. We also test multiple noise levels.

Figure 6 summarizes the performance evaluation. Results are far from perfect when looking at DMSE (not shown). But other more refined metrics, tell us that composition still holds a lot of information. Looking at GT rating tells us that for reasonable measurement noise of up to 2°, GT state is mostly estimated to be among the 4 most probable EDC states (and half of the time among the 2 leaders). The gmd plot shows that the False EDC states are mostly close to the GT state. Considering that these are triplets that were never measured together, and could not be estimated otherwise, this is a considerable value. Looking towards active planning, composition may enable planning for using unseen triplets.

V. Conclusions

In this paper we present a new approach for localization and mapping based on qualitative spatial reasoning. We demonstrate how incooperating a motion model in qualitative estimation improves results. Our method is also more general in the way that it can easily use various types qualitative space partitions, and various underlying SLAM based methods. We also successfully used qualitative inference inherent ability to accommodate small errors, and generated a low compute approximated algorithm. In addition to improving both complexity and performance compared to the state of the art [21] and [22], we also show that this approach is a practical alternative for low cost robotic systems or for active planning, in cases where exact metric location is not important. Another building block in that direction is the ability to use inference for landmark triplets that were not seen together through composition.

Future research may focus on using qualitative data for handling recognition errors, addressing complex volumed landmarks, and qualitative active planning.

VI. Acknowledgments

This work was partially supported by the Israel Ministry of Science & Technology (MOST).

References

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, "Simultaneous localization and mapping: Present, future, and the robust-perception age," *arXiv preprint arXiv:1606.05830*, 2016.

[2] F. Dellaert, "Factor graphs and gtsam: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012, gTSAM.

[3] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, Shanghai, China, May 2011.

[4] S. A. et al, "Ceres solver," Google, Tech. Rep., 2016, http://ceres-solver.org,.

[5] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. Kelly, and A. J. Davison, "Slam++: Simultaneous localisation and mapping at the level of objects," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1352–1359.

[6] A. Kim and R. M. Eustice, "Active visual SLAM for robotic area coverage: Theory and experiment," *Intl. J. of Robotics Research*, vol. 34, no. 4-5, pp. 457–475, 2014.

[7] V. Indelman, L. Carlone, and F. Dellaert, "Planning in the continuous domain: a generalized belief space approach for autonomous navigation in unknown environments," *Intl. J. of Robotics Research*, vol. 34, no. 7, pp. 849–882, 2015.

[8] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research*, vol. 31, pp. 217–236, Feb 2012.

[9] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec 2008.

[10] L. Polok, V. Ila, M. Solony, P. Smrz, and P. Zemcik, "Incremental block cholesky factorization for nonlinear least squares in robotics," in *Robotics: Science and Systems (RSS)*, 2013.

[11] D. Scaramuzza and F. Fraundorfer, "Visual odometry: Part i the first 30 years and fundamentals," 2011.

[12] F. Fraundorfer and D. Scaramuzza, "Visual odometry. Part II: Matching, robustness, optimization, and applications," 2012.

[13] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via l1 relaxation: A convex approach for robust estimation over graphs," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2014, pp. 2667–2674.

[14] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 556–563.

[15] V. Indelman, E. Nelson, J. Dong, N. Michael, and F. Dellaert, "Incremental distributed inference from arbitrary poses and unknown data association: Using collaborating robots to establish a common reference," *IEEE Control Systems Magazine (CSM), Special Issue on Distributed Control and Estimation for Robotic Vehicle Networks*, vol. 36, no. 2, pp. 41–74, 2016.

[16] N. Sünderhauf and P. Protzel, "Switchable constraints vs. max-mixture models vs. RRR - a comparison of three approaches to robust pose graph SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013.

[17] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Intl. J. of Robotics Research*, vol. 32, no. 7, pp. 826–840, 2013.

[18] S. Pathak, A. Thomas, and V. Indelman, "A unified framework for data association aware robust belief space planning and perception," *Intl. J. of Robotics Research*, vol. 32, no. 2-3, pp. 287–315, 2018.

[19] S. Lowry, S. Niko, N. Paul, J. L. John, C. David, C. Peter, and J. M. Michael, "Visual place recognition: A survey," *IEEE Transactions on Robotics 32.1 (2016): 1-19.*, 2016.

[20] T. Mossakowski and M. Reinhard, "Qualitative reasoning about relative direction of oriented points," 2012.

[21] M. McClelland, M. Campbell, and T. Estlin, "Qualitative relational mapping for mobile robots with minimal sensing," *Journal of Aerospace Information Systems*, vol. 11, no. 8, pp. 497–511, 2014.

[22] J. Padgett and M. Campbell, "Probabilistic qualitative mapping for robots," *Robotics and Autonomous Systems*, vol. 98, pp. 292–306, 2017.

[23] T. S. Levitt and D. T. Lawton, "Qualitative navigation for mobile robots," *Artificial intelligence*, vol. 44, no. 3, pp. 305–360, 1990.

[24] J. Y. Zheng and S. Tsuji, "Panoramic representation for route recognition by a mobile robot," *International Journal of Computer Vision*, vol. 9, no. 1, pp. 55–76, 1992.

[25] C. Schlieder, "Representing visible locations for qualitative navigation," *Qualitative reasoning and decision technologies*, vol. 523532, 1993.

[26] T. Wagner, U. Visser, and O. Herzog, "Egocentric qualitative spatial knowledge representation for physical robots," *Robotics and Autonomous Systems*, vol. 49, no. 1, pp. 25–42, 2004.

[27] R. Moratz, L. Dominik, and M. Till, "A condensed semantics for qualitative spatial reasoning about oriented straight line segments," 2011.

[28] C. Freksa, "Using orientation information for qualitative spatial reasoning," 1992.

[29] C. Freksa and K. Zimmermann, "On the utilization of spatial structures for cognitively plausible and efficient reasoning," in *IEEE International Conference on Systems, Man and Cybernetics*, 1992, pp. 261–266.

[30] R. Moratz and R. Marco, "Qualitative spatial reasoning about relative point position," 2008.

[31] A. Scivos and N. Bernhard, "The finest of its class: The natural point-based ternary calculus lr for qualitative spatial reasoning." *International Conference on Spatial Cognition. Springer, Berlin, Heidelberg.*, 2004.

[32] J. Padgett and M. Campbell, "Q-link: A general planning architecture for navigation with qualitative relational information," *Robotics and Autonomous Systems*, vol. 108, pp. 51–65, 2018.

[33] D. Kopitkov and V. Indelman, "No belief propagation required: Belief space planning in high-dimensional state spaces via factor graphs, matrix determinant lemma and re-use of calculation," *Intl. J. of Robotics Research*, vol. 36, no. 10, pp. 1088–1130, August 2017.

[34] R. Mor and V. Indelman, "Probabilistic qualitative localization and mapping - supplementary material," Tech. Rep., 2020. [Online]. Available: https://tinyurl.com/qqqcn58

[35] V. Pierlot and M. Van Droogenbroeck, "A new three object triangulation algorithm for mobile robot positioning," *IEEE Transactions on Robotics*, vol. 30, no. 3, pp. 566–577, 2014.

[36] K. Y. Leung, Y. Halpern, T. D. Barfoot, and H. H. Liu, "The utias multi-robot cooperative localization and mapping dataset," *Intl. J. of Robotics Research*, vol. 30, no. 8, pp. 969–974, 2011.