Aberystwyth University

Probabilistic self-localisation on a qualitative map based on occlusions
Santos, Paulo E; Martins, Murilo F; Fenelon, Valquiria ; Cozman, Fabio G; Dee, Hannah

Probabilistic Self-Localisation on a Qualitative Map based on Occlusions

Paulo E. Santos[a]* and Murilo F. Martins[a] and Valquiria Fenelon[b] and
Fabio G. Cozman[b] and Hannah M. Dee[c]

[a]*Dep. of Electrical Engineering, Centro Universitário da FEI,
Av. Humberto de A.C. Branco, 3972, São Paulo, Brazil, CEP 09850-901*;

[b]*Escola Politécnica, Universidade de São Paulo,
Av. Prof. Luciano Gualberto, travessa 3 n. 380, São Paulo - SP, Brazil, CEP 05508-010*;

[c]*Department of Computer Science, Aberystwyth University,
Penglais, Aberystwyth, Ceredigion, SY23 3FL , UK.*

Spatial knowledge plays an essential role in human reasoning, permitting tasks such as locating objects in the world (including oneself), reasoning about everyday actions and describing perceptual information. This is also the case in the field of mobile robotics, where one of the most basic (and essential) tasks is the autonomous determination of the pose of a robot with respect to a map, given its perception of the environment. This is the problem of robot self-localisation (or simply the localisation problem). This paper presents a probabilistic algorithm for robot self-localisation that is based on a topological map constructed from the observation of spatial occlusion. Distinct locations on the map are defined by means of a classical formalism for qualitative spatial reasoning, whose base definitions are closer to the human categorisation of space than traditional, numerical, localisation procedures. The approach herein proposed was systematically evaluated through experiments using a mobile robot equipped with a RGB-D sensor. The results obtained show that the localisation algorithm is successful in locating the robot in qualitatively distinct regions.

Keywords: Qualitative Spatial Reasoning; Markov Localisation; Perception of Occlusion

---

*Corresponding author. Email: psantos@fei.edu.br

1.   Introduction

Currently the most successful algorithms for robot self-localisation are based on probabilistic methods that assume maps of the environment defined by means of metric information (Thrun *et al.*, 2005; Wolter *et al.*, 2008). One of the shortcomings of these methods, however, is that they largely ignore the knowledge that humans have (and use) about the environment (Thrun, 2003). Bridging the gap between probabilistic methods in robotics and knowledge about the robot's domain is not only of theoretical interest, but it is also an essential step towards equipping robots with the capability of interpreting sensor data using high-level knowledge (Falomir *et al.*, 2013; Tapus and Siegwart, 2006). The use of common-sense knowledge represented as spatio-temporal concepts is also of utmost importance in the development of robotic systems capable of interacting with humans in a natural way (Deits *et al.*, 2013; Moratz and Ragni, 2008).

The field of *Qualitative Spatial Reasoning* (QSR) (Cohn and Renz, 2008) (a subfield of Knowledge Representation in Artificial Intelligence) attempts the representation of commonsense spatial knowledge based on qualitative properties of the domain, aiming to achieve cognitively plausible theories about spatial information. For instance, QSR theories include a mereotopological theory based on the connectivity between spatial regions, the definition of occlusion and parallax, the formalisation of relative location and spatial vagueness, as well as the definition of qualitative theories about distance, boundaries, shapes and so forth (as overviewed in Ligozat (2011)).

Traditional QSR formalisms, however, are independent of observer viewpoints (apart from a few exceptions, e.g. (Randell *et al.*, 2001)), giving limited application to robotics research. In previous work (Fenelon *et al.*, 2013; Santos, 2007; Santos *et al.*, 2009; Souchanski and Santos, 2008), we have proposed a dynamic formalism about space in which qualitative changes observed by a mobile robot are the building blocks of the system, therefore including the observer in a QSR formalism. These approaches were developed in classical logic languages and, therefore, were not capable of handling sensor uncertainties.

In this context, the present paper proposes a *probabilistic* localisation algorithm based on a *qualitative* representation derived from a QSR formalism about object occlusion. We, therefore, bring together both qualitative and probabilistic (quantitative) reasoning techniques to bear on a problem which is inherently viewpoint-dependent. The contributions of this work are two-fold: first, it contributes to research on robot localisation by introducing an algorithm that accomplishes probabilistic localisation using non-metric information defined over a qualitative spatial reasoning formalism; second, this work contributes to QSR by presenting an experimental evaluation of a viewpoint-based QSR theory implemented within a probabilistic localisation algorithm in a real robotic domain.

The algorithm presented in this paper was tested on our ActiveMedia Pioneer Peoplebot mobile robot that was tele-operated through an indoor environment while an RGB-D sensor (that is, a Microsoft Kinect) captured a sequence of snapshots around reference objects.

This work builds upon various QSR formalisms, in particular the Region Connection Calculus (Randell *et al.*, 1992) and the Region Occlusion Calculus (Randell *et al.*, 2001). We outline the works which provide our foundation in Section 3, with our proposed extensions described in Section 4. For reasoning about space we define a qualitative map, upon which the robot can be localised: this stage is presented in Section 5.

The key technical contribution of this paper is a set of algorithms linking the qualitative map, the QSR formalism and robot perception via probabilistic reasoning. This can be found in Section 6, with our experimental setup and evaluation in Section 7.

The novelty of the work proposed in this paper is the investigation of a high-level

representation of space information, and how to combine it with probabilistic methods, in order to provide an intuitive way of handling robot localisation at a level of abstraction closer to the human categorisation of space.

2. Related Work

Research in robot mapping has over 30 years of history and its major results are summarised in the extensive literature surveys presented by Thrun (2003) and Boal *et al.* (2014). Boal *et al.* (2014) divides the maps currently investigated in the robotic literature into four classes: (1) *metric maps* that represent the robot domain by means of its geometrical properties (Filliat and Meyer, 2003; Leonard and Durrant-Whyte, 1991); (2) *topological maps* that encode the environment by a graph whose nodes represent possible locations and edges between nodes represent that two locations are connected (Liu and Siegwart, 2014; Ranganathan and Dellaert, 2011; Remolina and Kuipers, 2004); (3) *hybrid (or hierarchical) maps* which represent the environment by a combination of metrical and topological information (Konolige *et al.*, 2011; Zivkovic *et al.*, 2005); and, (4) *semantic (or cognitive) maps* that contain high-level information about the environment, including object types, functionalities, and their interrelations (Nüchter and Hertzberg, 2008; Posada *et al.*, 2014; Vasudevan *et al.*, 2007). According to Boal *et al.* (2014) metric representations of maps allow very accurate localisation algorithms, at the expense of an increasing computational complexity related to the high dimensionality of the entities involved. Topological maps, on the other hand, provide a more compact representation of the robot's domain than the metric approach, but require more complex processing of the sensory information and are more prone to *perceptual aliasing* (which is the problem of two distinct places being perceived as the same location). The development of hybrid maps aims to combine the precision of a metric representation with the more abstract representation of the robot's location given by the topological map. Semantic maps go one step upwards in terms of abstracting the robot's space, where relational information is used to allow high-level reasoning along with localisation procedures.

In this work we investigate a probabilistic localisation algorithm on topological maps defined over qualitative (or non-metrical) spatial relations. Therefore, the work proposed in this paper falls at the intersection between topological and semantic mapping.

Methodologically, our work has its roots on the localisation procedure presented by Levitt and Lawton (1990), where a topological map was built by regions bounded by sets of lines connecting pairs of point-wise landmarks. This idea inspired the development of several spatial representations. For instance, Schlieder (1996) proposes a spatial representation (called *panorama*) that constrains the location of a point with respect to the visual ordering of point-wise objects observed around the robot. The notion of qualitative navigation is defined in Schlieder (1996) and Stolzenburg (2010) in terms of changes in this ordering information, as the observer crosses each of the virtual lines defined by the reference objects.

An approach closely related to our work is that of Fogliaroni *et al.* (2009), which considers extended convex objects to be landmarks, and decomposes navigable space based on a model of object occlusion and visibility. More specifically, that model generates a tessellation of the navigable space in terms of portions of space that were *visible*, *partially visible* and *occluded*. The *ordering* in which these attributes are observed is also used to qualify the regions of space within the tessellation. Following similar ideas, Santos *et al.* (2009) propose a qualitative spatial theory based on a logical formalisation of occlusion and the observation of cast shadows. Preliminary results of applying this theory to a mobile robot domain for the task of localisation were presented in Fenelon *et al.* (2013). Also related to the approach proposed here is the method introduced in McClelland *et al.*

(2013) which builds a map using the relative positions of landmarks. These methods are usually based on some human-level conceptualisation of space, some of them grounded on recent findings in cognitive psychology (Wolter *et al.*, 2008). However, they are incapable of handling sensor noise, which limits their applicability to well-controlled, deterministic domains. A more complete overview of qualitative representations for robot localisation is presented in Wolter *et al.* (2008).

The most successful localisation methods in robotics are based on probabilistic algorithms. Probabilistic localisation algorithms are usually defined on top of Bayesian filtering (Thrun, 2003; Thrun *et al.*, 2005). In a nutshell, this is a recursive algorithm consisting of two parts: *prediction* (where the belief over the current state of a robot – usually its position and orientation – is calculated on the prior belief of a previous state); and *measurement update* (whereby the predicted current state is weighted by the probability of a measurement on this state). This basic definition summarises works on *Markov localisation* (Fox *et al.*, 1999) and localisation procedures that use *Kalman filters* (Chen, 2012). The latter is a special case of the former, where sensor and motion models are linear Gaussian functions. Methods known as *Monte Carlo localisation* (Dellaert *et al.*, 1999) extend the basic Markov localisation by applying a particle filter to represent the distribution of possible states. These particles are updated according to the prediction of the next state (given the robot's motion) and are resampled according to the agent's perception. These ideas have such a high acceptance in the robotics field that the seminal work of Dellaert *et al.* (1999) has over a thousand citations, most of which related to the direct application of the method. Therefore, a complete survey of recent works related to probabilistic localisation algorithms is a hard task. Surveys of early approaches to probabilistic robot localisation can be found in Thrun (2003); Thrun *et al.* (2005).

Preliminary results on the framework reported in the present paper were shown in Pereira *et al.* (2013), where a qualitative-probabilistic approach is developed combining the ideas of qualitative localisation using cast shadows proposed in Fenelon *et al.* (2013) with a Bayesian filter. This approach proved to be successful on the tasks of robot localisation and self-calibration of the robots vision system through experiments using a mobile robot in a real environment. The experiments reported in Pereira *et al.* (2013), however, assumed the localisation of a robot manoeuvring around a single object and its unique shadow. The present paper generalises this idea assuming a map defined over the occlusion between (any number of) pairs of objects. Thus, the ideas presented here can be applied to maps over any number of occluded bodies. To the best of our knowledge, this is the first work that presents empirical results on a probabilistic localisation algorithm based on qualitative spatial information.

This paper assumes the usual notation in logic languages where lower-case roman letters refer to variables and upper-case to constants (unless explicitly stated otherwise). Bold fonts in formulae will be reserved for sets and probability distributions. The theory of occlusion upon which this work is grounded is described in the next section.

3. Region Occlusion Calculus

The basic spatial theory used in this work is the Region Occlusion Calculus (ROC) (Randell *et al.*, 2001), which is an extension of the Region Connection Calculus (RCC) (Randell *et al.*, 1992). RCC is a first-order axiomatisation of spatial relations based on a reflexive, symmetric and non-transitive dyadic primitive relation of *connectivity* ($C/2$) between two regions. Informally, assuming two regions x and y, the relation $C(x, y)$, read as *"x is connected with y"*, is true if and only if the closures of x and y have at least one point in common.

Assuming the $C/2$ relation, and two spatial regions x and y, the following base relations

can be defined: *disconnected from* (*DC*), *part of* (*P*), *equal to* (*EQ*), *overlaps* (*O*); *partially overlaps* (*PO*); *externally connected* (*EC*); *tangential proper part* (*TPP*); *non-tangential proper part* (*NTPP*). RCC also includes the inverse relations of *P*, *TPP* and *NTPP*, which are represented by a capital 'I' appended to the relative relation: *PI*, *TPPI* and *NTPPI*.

The set constituted by the relations *DC*, *EQ*, *PO*, *EC*, *TPP*, *NTPP*, *TPPI*, and *NTPPI* is the jointly exhaustive and pairwise disjoint set (JEPD) usually referred to as RCC8. The continuous transitions between the RCC8 relations, for two regions x and y, are shown as a *conceptual neighbourhood diagram* (CND) in Figure 1. By continuous transitions we mean that in between adjacent vertices of the graph there can be no other possible relation qualifying the state of the two regions. That is, assuming that the objects move continuously on the plane, these are the only transitions that are possible.
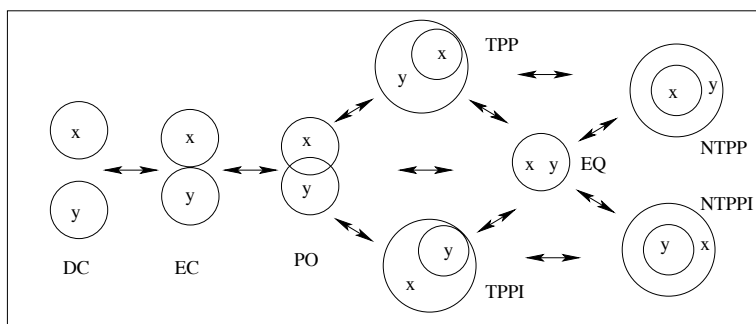

Figure 1. The RCC8 relations and their conceptual neighbourhood diagram (Randell *et al.*, 1992).

Using RCC8 relations, along with the primitive relation $TotallyOccludes(x, y, \nu)$ (which stands for "x totally occludes y with respect to the viewpoint ν"), the Region Occlusion Calculus (ROC) defines the 20 base JEPD relations. Figure 2 shows a graphical representation of the ROC relations between two objects, represented as a white and a shaded region. In this figure, the shaded region corresponds to the first argument, and the white region to the second argument of ROC relations. For instance, the relation $PartiallyOccludesTPP(x, y)$ is depicted with the shaded region x occluding the white region y, while the 2D projection of the shaded object is a tangential proper part (*TPP*) of the 2D projection of the white object. It is worth noting that the relations on *mutual occlusion* only occur if and only if at least one of the objects is non-convex. ROC also defines a conceptual neighbourhood diagram (introduced in Randell, Witkowski (2002)) that we do not present in this paper for brevity.

Region Occlusion Calculus makes a distinction between the occupancy regions of bodies and their images (or projections) from the viewpoint of an observer. This distinction is accomplished by assuming two functions: the function $region(x)$, which maps a body x to its 3D occupancy region, and the function $image(x, \nu)$ that maps a body x to the body's 2D projection, as seen from a viewpoint ν. The viewpoint in ROC is modelled as a pinhole camera whose parameters are not important for the qualitative theory.

It is worth pointing out also that the "I" in the relations $TotallyOccludesTPPI(x, y, \nu)$ and $TotallyOccludesNTPPI(x, y, \nu)$ represents the inverse of TPP and $NTPP$, respectively; so, for instance, $TotallyOccludesTPPI(x, y, \nu)$, means that the body x totally occludes the body y, but $image(y)$ is the tangential proper part of $image(x)$ (i.e., $TPPI(image(x, \nu), image(y, \nu))$). The superscript "-1" in some ROC relations represents the inverse of the occlusion part of the relation.
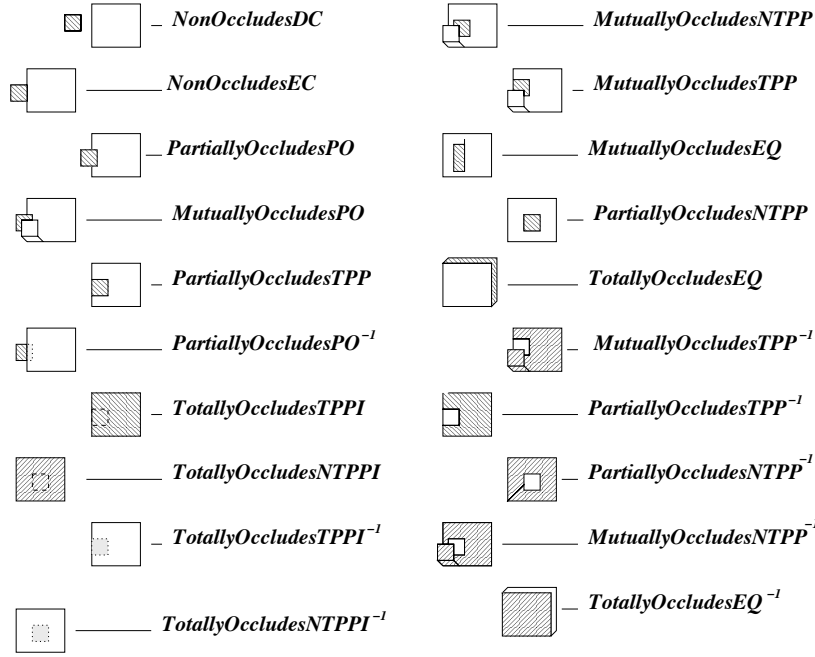
Figure 2. ROC relations between two objects (white and shaded regions).

4.   Relative Positions

As well as the 20 ROC relations, this work assumes observer-relative positions of pairs of objects by means of the relations $Left$, $Right$, $Closer$ and $Further$. Given two distinct bodies x and y and a viewpoint ν (and assuming that the observer's horizon is fixed and that the field of view is restricted) the relative positions are as follows:

- $Left(x, y, \nu)$, representing the fact that "x is to the left of y from viewpoint ν" (analogously, $Right(x, y, \nu)$);
- $Further(x, y, \nu)$, represents the fact that "x is further than y from viewpoint ν" (analogously, $Closer(x, y, \nu)$).

In this work, these relations are grounded on data from the vision system, given the depth information (e.g., provided by an RGB-D sensor) and the distance from the centroids of the objects' images to the left border of the camera field of view. More formally, let $dist\_from\_left(image(o, \nu))$ (read as *distance from the left border*) be a function that maps an image of an object o and a viewpoint ν to the distance of the image's centroid to the left border of ν's field of view; and, $depth(image(o, \nu))$ be a function that maps the image of an object o to the depth of o with respect to ν (i.e., the distance from the object to the observer). Thus, the relations $Left$, $Right$, $Closer$ and $Further$ can be defined by the formulae below.

$$Left(x, y, \nu) \leftrightarrow dist\_from\_left(image(x, \nu)) < dist\_from\_left(image(y, \nu)).$$

$$Right(x, y, \nu) \leftrightarrow dist\_from\_left(image(x, \nu)) > dist\_from\_left(image(y, \nu)).$$

$$Further(x, y, \nu) \leftrightarrow depth(image(x, \nu)) > depth(image(y, \nu)).$$

$$Closer(x, y, \nu) \leftrightarrow depth(image(x, \nu)) < depth(image(y, \nu)).$$

The relations on relative positions introduced above are transitive, irreflexive and asymmetric. For completion we also have to include the relations $NonLeftRight(x, y, \nu)$ (stating

that "x is neither left nor right of y") and $NonCloserFurther(x, y, \nu)$ (stating that "x is neither closer nor further than y"). However, these cases rarely occur in real vision data.

5.   Qualitative Map

Although ROC is defined on non-convex physical bodies, in this work we use the subset of ROC that is related to convex objects only. This constraint simplifies the base vocabulary of ROC without interfering on the generality of the localisation procedure, since in this paper the localisation is accomplished over the convex hulls of the objects' images.

Considering the ROC relations between (the convex hulls of) objects o_1 and o_2 from a viewpoint ν, only the following relations and their inverses have models since the *mutually occludes* relations do not hold with respect to convex hulls:

- $NonOccludesDC(o_1, o_2, \nu)$;
- $NonOccludesEC(o_1, o_2, \nu)$;
- $PartiallyOccludesPO(o_1, o_2, \nu)$ (and $PartiallyOccludesPO(o_1, o_2, \nu)^{-1}$);
- $PartiallyOccludesTPP(o_1, o_2, \nu)$ (and $PartiallyOccludesTPP(o_1, o_2, \nu)^{-1}$);
- $PartiallyOccludesNTPP(o_1, o_2, \nu)$ (and $PartiallyOccludesNTPP(o_1, o_2, \nu)^{-1}$);
- $TotallyOccludesTPPI(o_1, o_2, \nu)$ (and $TotallyOccludesTPPI(o_1, o_2, \nu)^{-1}$);
- $TotallyOccludesEQ(o_1, o_2, \nu)$ (and $TotallyOccludesEQ(o_1, o_2, \nu)^{-1}$);
- $TotallyOccludesNTPPI(o_1, o_2, \nu)$ (and $TotallyOccludesNTPPI(o_1, o_2, \nu)^{-1}$).

The definition of a qualitative map using occlusion relations depends on the notion of *lines of sight*, that is understood here as the virtual tangent lines that can be drawn on the borders between pairs of objects.

Considering the lines of sight and ROC relations, we define a discretisation of the space around pairs of objects into qualitatively distinct relations. This discretisation, exemplified in Figure 3, represents a qualitative map defined on occlusion relations between two objects ($O1$ and $O2$), whereby the regions marked with the numbers 1, 2, 3, 4, and 5 refer to regions where a viewpoint ν would observe $Right(O1, O2, \nu)$ and also:

- in Region 1 ν observes $NonOccludesDC(O1, O2, \nu)$;
- in Region 2: $NonOccludesEC(O1, O2, \nu)$;
- in Region 3: $PartiallyOccludesPO(O2, O1, \nu)$;
- in Region 4: $TotallyOccludesTPPI(O1, O2, \nu)$; and,
- in Region 5: $TotallyOccludesNTPPI(O1, O2, \nu)$.

An analogous qualification applies for the non-labelled regions in Figure 3. This idea can be easily extended to any number of referent objects: Figure 4 shows an analogous map related to three objects.
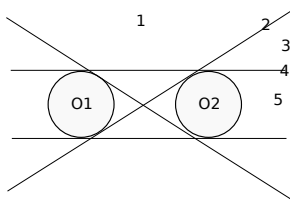

Figure 3. Distinct regions implied by the observation of occlusion relations between two objects (global view).

The qualitative map shown in Figure 4 divides the space around the three objects (red (r), blue (b) and green (g) boxes) into 28 regions defined by the lines of sight between every pair of these objects and their observer-relative positions. It is worth noting that
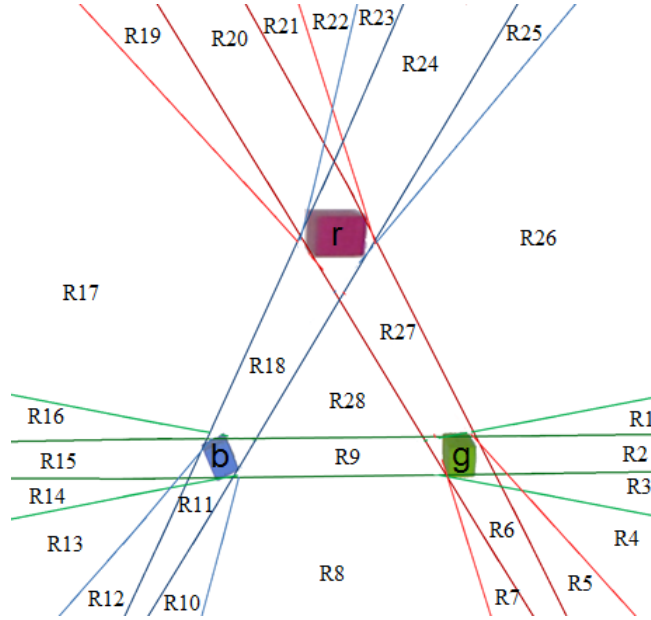
Figure 4.   A representation of the qualitative map with 28 distinct regions. The lines of sight between the objects define the boundaries between regions. The colours of the lines in this figure are only to help its visualisation, they do not have any special meaning.

there are regions around the target objects in Figure 4 that were not marked (and not considered in the map). These are regions that are near the target objects (or are results of multiple intersections of regions defined with respect to various pairs of objects) and whose sizes are negligible with respect to the size of the robot and, thus, were not considered as positions in the map. As we shall see later in this paper, if the robot happens to pass over these regions, it keeps the belief in the previous location inferred until it reaches the next region marked in the map.

The regions in the map on Figure 4 are defined by Formulae 1 to 28 below.

$$located(R1, \nu, b, g, r) \leftarrow ((PartiallyOccludesTPPI(b, g, \nu) \land \tag{1}$$
$$Right(b, g, \nu) \land Further(b, g, \nu)))$$
$$\land((NonOccludesDC(b, r, \nu) \land$$
$$Left(b, r, \nu)))$$
$$\land((NonOccludesDC(g, r, \nu) \land$$
$$Left(g, r, \nu) \land Closer(g, r, \nu))).$$
$$located(R2, \nu, b, g, r) \leftarrow TotallyOccludesNTPPI(b, g, \nu) \tag{2}$$
$$\land((NonOccludesDC(g, r, \nu) \land$$
$$Left(g, r, \nu) \land Closer(g, r, \nu))).$$

$$\vdots$$

$$located(R12, \nu, b, g, r) \leftarrow (NonOccludesDC(b, g, \nu) \tag{12}$$
$$\land Left(b, g, \nu) \land Front(b, g, \nu))$$
$$\land(PartiallyOccludesPO(b, r, \nu)$$
$$\land Right(b, r, \nu) \land Front(b, r, \nu))$$
$$\land(NonOccludesDC(g, r, \nu) \land Right(g, r, \nu)).$$

$$\vdots$$

$$located(R28, \nu, b, g, r) \leftarrow NonOccludesDC(b, r, \nu) \wedge \qquad (28)$$
$$NonOccludesDC(b, g, \nu) \wedge$$
$$NonOccludesDC(g, r, \nu).$$

According to Formula (12), a robot located in region R12 (on the map in Figure 4) will see objects b and g as $NonOccludesDC$; b will be to the *left* and *in front of* g; b and r will be $PartiallyOccludesPO$; while b will be to the *right* and *in front of* r; and, finally, g and r will be seen as $NonOccludesDC$ and g will be to the *right* of r. A snapshot taken by the robot located in R12 is shown in Figure 5.

Figure 5. A snapshot captured while the robot was on region R12 on the map (Figure 4).

In order to accomplish robot self-localisation, we extend the probabilistic algorithm proposed in Pereira *et al.* (2013) to take into account occlusion information from any number of objects in the scene. Although the localisation procedure can be applied over any finite number of objects, in the presentation below we always refer to the 28 regions defined on the occlusion relations between three distinct objects (as shown in Figure 4).

6.   Probabilistic Qualitative Self-Localisation

We refer to *belief distribution* as the posterior probabilities over state variables, conditioned on the available data. We use the notation $bel(s_t) := P(s_t|e_{0:t})$; that is, the belief $bel(s_t)$ is the probability of a state (s_t) given all the evidence ($e_{0:t}$) up to an instant (t). In this work, the state s_t indicates the region where the robot is located at instant t (given by qualitative maps such as that shown in Figure 4), and ($e_{0:t}$) represents the evidence, denoted in terms of occlusion relations between pairs of objects from instant 0 to instant t. This evidence is provided by the ROC relations as observed from a viewpoint (cf. defined in Formulae (1) to (28) for the diagram in Fig. 4). The goal of the algorithm is to find the value of s_t that maximises $bel(s_t)$.

In the remainder of this paper we use the letter N to represent the number of distinct locations on a map (e.g. N=28 on the map shown in Fig. 4).

In order to compute the beliefs we use:

$$
\begin{aligned}
bel(s_t) &= P(s_t|e_{0:t}) \\
&= P(s_t|e_{0:t-1}, e_t) \\
&= \eta P(e_t|s_t, e_{0:t-1}) P(s_t|e_{0:t-1}),
\end{aligned}
$$

where η is a normalisation constant. Using a Markov hypothesis, the next state is independent of earlier measurements $e_{0:t-1}$; hence:

$$
bel(s_t) = \eta P(e_t|s_t) P(s_t|e_{0:t-1}).
$$

This expression is the *measurement update* (Thrun *et al.*, 2005). Its first term $P(e_t|s_t)$ is, in fact, the image model. Its second term (known as the updated belief $P(s_t|e_{0:t-1}) = \overline{bel}(s_t)$) must be calculated before incorporating the evidence e_t.

The updated belief $\overline{bel}(s_t)$ represents a one-step prediction of the next state s_t, obtained by conditioning s_t on the previous state s_{t-1}, as follows:

$$
\begin{aligned}
\overline{bel}(s_t) &= \sum_{s_{t-1}} P(s_t|s_{t-1}, e_{0:t-1}) P(s_{t-1}|e_{0:t-1}) \\
&= \sum_{s_{t-1}} P(s_t|s_{t-1}) P(s_{t-1}|e_{0:t-1}).
\end{aligned}
$$

In the previous expression, the term $P(s_t|s_{t-1})$ is the state transition model. In this work the transition model $P(s_t|s_{t-1})$ refers to the qualitative motion across the regions defined in the qualitative map (such as those shown in Figure 4). The transition model encodes the probability of a state change, given a moving action and the ROC conceptual neighbourhood diagram (as explained in detail in Section 6.2 below).

The value of a state s_t is a region R_i where $i = 1, \ldots, N$. The term $P(s_{t-1}|e_{0\ldots t-1})$ is the belief calculated on the previous iteration:

$$
bel(s_{t-1}) = P(s_{t-1}|e_{0:t-1}).
$$

Therefore, the posterior belief (or prediction) is the combination of previous expressions:

$$
\overline{bel}(s_t) = \sum_{s_{t-1}} P(s_t|s_{t-1}) P(s_{t-1}|e_{0:t-1}) = \sum_{s_{t-1}} P(s_t|s_{t-1}) bel(s_{t-1}).
$$

Finally, the belief is given by:

$$
bel(s_t) = \eta P(e_t|s_t) \overline{bel}(s_t) = \eta P(e_t|s_t) \sum_{s_{t-1}} P(s_t|s_{t-1}) bel(s_{t-1}).
$$

6.1. *The localisation procedure*

The robot estimates its relative location using images captured from an RGB-D sensor. Given the robot's perceptions described as ROC relations between pairs of objects, a Bayesian filter is used to infer the robot's position.

This ROC-enhanced Bayesian filter we call the Probabilistic Qualitative Self-Localisation algorithm (PQS), shown in Algorithm 1. The algorithm is initialised with a

(a) Segmentation   (b) Contours   (c) Connectivity
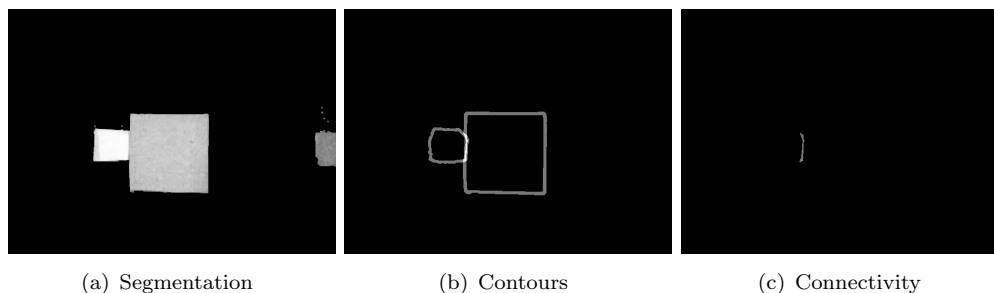
Figure 6. Example of the segmentation process applied on the original snapshot shown in Figure 5.

uniform distribution on the robot's position ($\mathbf{P}(\mathbf{S_0})$ in Alg. 1). Then PQS calls a Perceptual Procedure (called PP algorithm, shown in Algorithm 2) that returns the evidences with maximum probability from the set of all evidences (i.e., every region on the qualitative map). The PP algorithm is subsequently detailed further in this paper. Given the evidence set, the PQS algorithm first calculates probabilities for all the N regions of the qualitative map (here the set of regions is denoted by $\mathbf{S_t}$), then it returns the beliefs about the current region (s_t). Finally, PQS runs a Bayesian prediction step where the posterior belief $\overline{\mathbf{bel}}(\mathbf{S_{t+1}})$ is calculated, which takes into account the current beliefs for all states (given the evidence). The probability of the next states (given the current state) is given by the transition model.

Algorithm 1 $PQS(\langle S_1, \ldots, S_N \rangle, image, \nu)$

1: $\overline{\mathbf{bel}}(\mathbf{S_1}) = \mathbf{P}(\mathbf{S_0}) = [\frac{1}{N}, \ldots, \frac{1}{N}]$
2: **while** (1) **do**
3:     $\mathbf{E_t} \leftarrow \max PP(image, \nu)$
4:     $\mathbf{bel}(\mathbf{S_t}) = \eta \mathbf{P}(\mathbf{S_t}|\mathbf{E_{0:t}}) = \eta \mathbf{P}(\mathbf{E_t}|\mathbf{S_t}) * \overline{\mathbf{bel}}(\mathbf{S_t})$
5:     $s_t \longleftarrow \arg \max_{s_t} \mathbf{P}(\mathbf{S_t}|\mathbf{E_{0:t}})$
6:     $\overline{\mathbf{bel}}(\mathbf{S_{t+1}}) = \sum_{s_t} P(S_{t+1}|s_t) P(s_t|e_t)$
7: **end while**

In the Perceptual Procedure (PP), described in Algorithm 2, the occlusion and relative position relations between pairs of objects are evaluated by detected features. These features are the degree of connectivity between the nearest sides of the objects, the distance between the objects, the object's depth, the ratio between the area of their bounding boxes, and also the relative position of the object's bounding box.

All the necessary features are extracted from the images using off-the-shelf computer vision algorithms. In order to obtain the object's bounding boxes from the snapshots taken from the robot's viewpoint, we use a morphological operator along with saturation values on the images obtained to perform the region of interest segmentation. Figure 6 shows an output of this segmentation procedure. The relative positions between pairs of objects are obtained directly from the sensor data, as described in Section 4.

The algorithm analyses pairs of objects and qualifies the qualitative relations between them (i.e., the relations on occlusion and relative positions detailed in Section 3). This is accomplished by considering the connectivity between segmented objects, along with depth information, in a two-tier procedure: firstly, the contours of all segmented objects are normalised to a fixed value and summed up, pixel by pixel; subsequently, this fixed value is subtracted once from all contours. As a result, only the pixels where overlapping objects were detected remain in the image (as exemplified in Figure 6).

Cases where total occlusion occurs are handled by using a flag that is set if occlusion was detected in immediately previous frames. This flag is then cleared once a transition

is observed.

Algorithm 2 $PP(Scene, \nu)$

1: *Segment scene to obtain the region of interest*
2: *Qualitatively classify the spatial relations between pairs of objects*
3: **for all** regions R_i on the qualitative map **do**
4:     $P(e = R_i) \leftarrow \Phi(located(R_i, \nu, b, g, r))$
5: **end for**
6: **return** $\arg\max_e P(E)$;

In Algorithm 2, the function Φ maps every possibility of localisation to a real value in the interval $[0, 1]$. This value is defined by the number of predicates satisfied in the body of the *located* formula. For instance, a formula that has all of its body satisfied receives value 1, whereas a formula that has 3 out of 4 body predicates satisfied receives the value 0.75.

6.2. *Models*

The Bayesian filter presupposes the definition of two probability models: the sensor model and the transition model. In Algorithm 1, the sensor model is the image model ($\mathbf{P}(\mathbf{e}_t|\mathbf{s}_t)$), whilst the transition model ($\mathbf{P}(\mathbf{S}_{t+1}|\mathbf{s}_t)$) represents the transition between regions in the map and is used for calculating the posterior belief (line 6 of Algorithm 1). Next, we describe how the image and motion models were designed.

The image model

The image model is encoded by $\mathbf{P}(\mathbf{e}_t|\mathbf{s}_t)$, which indicates the probability of the sensor to perceive the correct evidence related to a given state. In this paper we assumed a discretised Gaussian model. This is a $N \times N$ matrix that shows high probability values on the main diagonal and a fast decrease in value the further the matrix entry is from the main diagonal. This represents the idea that the greater the number of regions that have to be traversed in order to go from a region R_i to a region R_j ($i \neq j$), the smaller the corresponding probability of transition.

The transition model

The transition model conveys the probability of a change in the robot's location given a moving action and the ROC conceptual neighbourhood diagram (CND). Informally, considering a map (such as that shown in Figure 4), and a possible location of the observer, the CND of the Region Occlusion Calculus gives the possible changes in the sensor data as the robot moves to the neighbouring regions on the map (i.e. any of the relations on the CND graph that is connected by an edge to the relation qualifying the current observation). Assuming a steady speed of robot motion around the target objects, the transition model assigns a weight on the possible next position of the robot representing the probability of changes between locations. This weight is inversely proportional to the regions' sizes and is dependent on the neighbouring relations on the CND with respect to the current perception (this dependence is represented as a function of the motion action).

More specifically, the transition model is represented by a $N \times N$ transition matrix relating every possibility of state transition in the map (in the case of the map shown in Figure 4 it is a 28×28 matrix). The entries in this matrix are composed of a multiplication

(a) Snapshot of the test domain.   (b) Robot trajectory around the objects.
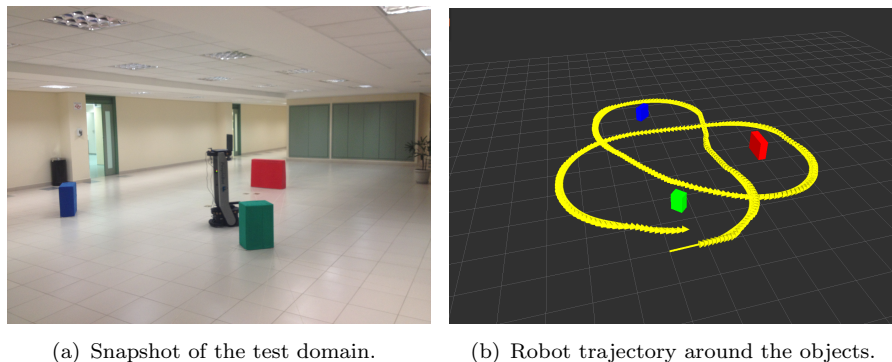
Figure 7. Experimental Scenario

by a real constant (representing the *weight*) and the function of the moving action. As mentioned above, the real constant is inversely related to the size of the region in the map, representing the probability of state transitions according to the size of the region: *the larger the region, the lower the probability of state transition.* The function represents perceived occlusion features that modulate the probability for a state change (i.e. the probability will be greater with respect to changes to locations related to neighbouring relations on the ROC CND, degrading gracefully to more distant regions). It is worth recalling that, in this work, the robot's motion is inferred from image changes with respect to visual features qualifying occlusions between pairs of objects. Thus, when the robot approximates a borderline region, the detection of certain occlusion features related to the next region increases the probability of a state change.

The entries in the $N \times N$ transition matrix representing the motion model are normalised by rows.

7.   Experimental Results

To verify the performance of the proposed PQS algorithm, experiments were conducted using our ActiveMedia Pioneer Peoplebot mobile robot, whose onboard computer runs the Robot Operating System (ROS) (Quigley *et al.*, 2009). Three boxes of distinct colours (red, green and blue) and varying sizes were carefully positioned in a triangular formation in an indoor environment forming a qualitative map akin to that illustrated in Figure 4. In addition, in order to capture both colour and depth images, the well-known Microsoft Kinect sensor was mounted on the base of the robot, whilst a joystick was used to tele-operate the robot through the environment. Figure 7(a) shows a snapshot of the robot in the environment during the experiments.

It is worth pointing out two key design issues of our experimental setup. Firstly, the Kinect sensor has an angular field of view of 57° horizontally and 43° vertically. Secondly, the distance between the boxes was, at most, 5 metres, due to space constraints. As a result, the Kinect sensor was conveniently mounted facing leftwards and the trajectory was carefully planned so that only left turns would be performed during the experiments, also ensuring that all regions of the qualitative map were visited at least once. The immediate area in which the robot was acted was in an open floor, where people were able to move around the room, thus the experiment involved considerable noise.

A detailed log was obtained from three subsequent runs that were saved for off-line processing. Figure 7(b) shows the trajectory followed during the first of three runs around the objects. The log is comprised of the robot's raw odometry data (logged at 10 Hz), as well as registered pairs of colour and depth images with resolution of 640×480 pixels (saved at 30 Hz) – a total of 9483 pairs of images. At each run the trajectory followed was

deliberately veered off-course at random – whilst keeping the same route – in order to *a*) increase variability of the collected samples, hence increasing the number of distinct viewpoints; and, as a result, *b*) prevent the recorded dataset from issues such as biasing and lack of generalisation.

Within the dataset, only 11% of the collected images contained all three target objects, whilst 45% contained two objects, 41% had only one object, and in 3% of the images no object was present. The ground truth of the dataset was manually annotated. In order to facilitate this annotation, we used markings on the floor (invisible to the robot) to represent the borders of the regions in the qualitative map.

Results from the PQS algorithm tested on the recorded dataset were compared with the answers from the PP-algorithm (described in Section 6.1), that provides the possible location of the robot by simply evaluating the qualitative relations (i.e. without using the Bayesian filter). From this comparison we verified whether the probabilistic algorithm contributed to the localisation procedure in map locations where not all target objects could be perceived.

Table 1 represents a confusion matrix with the results of PQS algorithm, in which the rows represent the actual location of the robot and the columns represent the algorithm output. The main diagonal of this matrix represents the percentage of true positives. We can see that in regions R2, R3, R4, R6, R10, R11, R12, R13, R15, R21, R22, R24, R25 and R28 the PQS algorithm had above 90% accuracy. Accuracy between 70% and 90% was obtained in regions R1, R5, R14, R16, R19, R20 and R23 and between 50% and 70% was obtained in regions R7, R17 and R26. The localisation procedure in regions R8, R9, R18 and R27 had accuracy below 50%.

Table 1 also shows that most false positives were cases where the algorithm located the robot in the immediate neighbourhood of its actual location. This is represented by the high density of false positives around the main diagonal of matrix 1, but also in some of the entries far from the main diagonal (whose cells also represent neighbouring regions of the correct robot location). These cases happened mainly at positions closer to the border limits of the regions, where the visual distinction between occlusion relations cannot be precisely captured by the vision system. Considering, however, that we want a system capable of locating a robot in *qualitatively distinct* regions of space, and that the border limits in the map cannot be distinguished from the bordering regions, it is conceivable to include the immediate neighbour regions to a robot location (that falls in the direction of the robot motion) as true positives. By doing that, the overall accuracy of our localisation algorithm improves considerably (as shown in Table 2).

Table 1. Confusion matrix with the results of PQS algorithm. The rows represent the actual location of the robot, whereas the columns represent the algorithm output.

	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20	r21	r22	r23	r24	r25	r26	r27	r28
r1	0.848	0.054	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.098	—	—
r2	—	0.991	0.009	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r3	—	0.010	0.971	0.019	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r4	—	0.024	0.003	0.969	0.003	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r5	—	0.208	—	—	0.703	0.089	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r6	—	—	—	—	—	1.000	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r7	—	0.218	—	—	—	0.080	0.701	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r8	—	0.090	—	—	—	—	0.002	0.382	0.003	0.053	—	—	—	—	—	—	—	0.013	—	—	—	—	—	—	—	0.052	0.021	0.384
r9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.000
r10	—	—	—	—	—	—	—	—	—	0.922	0.078	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r11	—	—	—	—	—	—	—	—	—	—	1.000	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r12	—	—	—	—	—	—	—	—	—	—	0.077	0.923	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r13	—	—	—	—	—	—	—	—	—	—	—	0.021	0.917	0.063	—	—	—	—	—	—	—	—	—	—	—	—	—	—
r14	—	—	—	—	—	—	—	—	—	—	—	—	—	0.833	0.167	—	—	—	—	—	—	—	—	—	—	—	—	—
r15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.000	—	—	—	—	—	—	—	—	—	—	—	—	—
r16	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.159	0.841	—	—	—	—	—	—	—	—	—	—	—	—
r17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.002	0.518	0.003	0.001	—	—	0.003	0.004	0.224	—	0.010	0.001	0.233
r18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.098	—	—	—	0.902
r19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.739	0.109	—	0.043	0.043	0.065	—	—	—	—
r20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.764	—	0.091	0.004	—	—	—	—	0.141
r21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.029	0.971	—	—	—	—	—	—	—
r22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.019	0.923	0.058	—	—	—	—	—
r23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.875	0.071	—	—	—	0.054
r24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.921	—	—	—	0.079
r25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.016	0.984	—	—	—
r26	0.003	—	—	—	—	0.002	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.002	0.625	0.005	0.365
r27	—	—	—	0.002	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.994
r28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1.000

Table 2 presents a more detailed description of the results obtained with PQS compared with the results obtained by running the localisation procedure using only the Perceptual Procedure (PP) algorithm (without the Bayesian filter). The first two columns of Table 2 represent, respectively, the actual location of the robot (column "Region") and the number of snapshots taken at this location (column "#im"); next, there are four columns under the label "#objects per image" representing the number of objects contained in the set of images of the related location. Accuracy results for the localisation procedure using only the PP algorithm are shown in column "PP", whereas accuracy results for the probabilistic algorithm introduced in this paper are under the label "PQS". The columns "PP-n" and "PQS-n" represent the results of PP and PQS including as true positives the inferred locations that were direct neighbours of the robot's actual location, taking into account the robot's direction of motion.

Comparing the results shown in columns "PP" and "PQS" (Table 2), we see that the Bayesian filter considerably improved the robot localisation provided by the PP algorithm in some regions where only one object could be seen (regions R6 and R28). In the remaining regions the accuracy results for both PP and PQS were similar.

When considering the direct neighbours of the correct robot location (shown in columns "PP-n" and "PQS-n"), we can see a considerable improvement in accuracy in both algorithms. However, PQS had a more expressive improvement, as we can see on regions R5, R9, R18 and R27. In particular R9, R18, R27 and R28 were regions where the robot always observed a single object and, therefore, the decision of the robot's location based solely on occlusion was under-constrained. In this case, the probabilistic algorithm provided the necessary support to infer the robot's location to the immediate neighbour of its actual position.

The regions where the accuracy results of PP and PQS had the lowest results (whether considering the neighbourhoods or not) were R8, R17 and R26. Two factors explain this poor performance: the size of these regions (as shown in the map on Figure 4) and the camera's limited field of view. Looking at Figure 4 we note that R8, R17 and R26 are the largest regions on the map and that they are defined by means of the observation of every pair of the three objects r, v and g as $NonOccludesDC$. However, the camera used to observe the environment had a limited field of view, implying that most images captured in these regions had only one or two objects. Therefore PP and PQS could not pin down the exact robot location, as there were other (competing) possibilities, given the observations of distinct objects in these regions. This problem could be solved by including an active vision system that searches for every object pair before making a decision. The development of such system within our framework is an issue left for future work.

8.   Discussion

The use of a formal representation of object occlusion in robotic localisation tasks was hypothesised in the paper that introduced the Region Occlusion Calculus (Randell *et al.*, 2001); Fogliaroni *et al.* (2009) explored a similar idea on a deterministic localisation algorithm that was further extended in Tassoni *et al.* (2011) to represent 3D maps of the environment. The maps used for localisation in the present paper resemble the maps defined in Fogliaroni *et al.* (2009) and Tassoni *et al.* (2011). However, to the best of our knowledge, the present work is the first working implementation of these ideas within a probabilistic algorithm and the first to evaluate the resulting localisation procedure on real robot data.

The qualitative maps based on occlusion (as proposed in this paper) are constructed with the lines of sight between pairs of landmarks. Taking that literally, we may have to

Table 2.   PP and PQS accuracy results considering the number of objects per image.

Region	#im	#objects per image				PP(%)	PQS(%)	PP-n(%)	PQS-n(%)
		0	1	2	3				
R1	112	0	3	78	31	85.7	84.8	99.1	100.0
R2	115	0	55	60	0	100.0	99.1	100.0	100.0
R3	415	0	3	346	66	96.1	97.1	100.0	100.0
R4	290	0	0	46	244	96.9	96.9	97.6	97.6
R5	101	0	9	87	5	70.3	70.3	76.2	79.2
R6	64	0	64	0	0	71.9	100.0	71.9	100.0
R7	87	0	7	80	0	70.1	70.1	75.9	78.2
R8	1452	0	661	698	93	38.4	38.2	43.9	44.1
R9	194	1	193	0	0	0.0	0.0	0.0	87.1
R10	102	0	4	71	27	92.2	92.2	100.0	100.0
R11	124	0	50	74	0	100.0	100.0	100.0	100.0
R12	91	0	3	30	58	92.3	92.3	100.0	100.0
R13	240	0	0	105	135	91.7	91.7	97.9	97.9
R14	54	0	7	31	16	83.3	83.3	100.0	100.0
R15	162	0	148	14	0	100.0	100.0	100.0	100.0
R16	44	0	7	37	0	84.1	84.1	100.0	100.0
R17	1651	0	577	986	88	52.0	51.8	52.4	52.5
R18	509	0	509	0	0	0.0	0.0	0.0	90.2
R19	46	0	1	15	30	73.9	73.9	84.8	84.8
R20	263	0	103	160	0	76.4	76.4	76.4	76,4
R21	34	0	0	4	30	85.3	97.1	100.0	100.0
R22	52	0	0	0	52	92.3	92.3	100.0	100.0
R23	56	0	3	5	48	87.5	87.5	92.9	94.6
R24	252	0	95	157	0	91.7	92.1	91.7	92.1
R25	64	0	2	32	30	95.3	98.4	98.4	100.0
R26	1970	0	725	1109	136	60.1	62.5	60.5	63.5
R27	503	3	497	3	0	0.0	0.0	0.0	99.4
R28	436	295	141	0	0	0.0	100.0	0.0	100.0

consider, for every pair of landmarks, the combination of diagrams such as that shown in Figure 3. The tessellation of space generated by this combination, assuming an increasing number landmarks, would impose a dense map to the algorithm, that may not have much relation to the high-level, qualitative, knowledge of the world that motivates the development of this work. Instead, we assume a strategy to *select* the most salient objects in the environment as target objects. Thus, not all objects in the environment should serve as landmarks, but only a selected few objects: those that are relevant to the task at hand (e.g. a service robot in a warehouse may have to use containers as landmarks), or that are distinctive features in the environment (e.g. most people navigate around the city of San Francisco using the relative positions of China Town, Fisherman's Wharf and the Ferry Tower). Currently, the map is given as input to the robot (i.e. the landmarks are selected by the system designer), automatic map building following the ideas presented here is an open issue.

In short, the localisation procedure developed in this paper is a Bayesian filter, with appropriate probabilities attached to sensor and transition models. The complexity of the Bayesian filter iteration is quadratic on the number of possible states (that is, in the number of distinct regions N). Consider n objects in our environment; for each pair of objects, we have to consider four lines of sight when we build the relevant regions in our qualitative map (see Figures 3 and 4 as examples). Now the number of distinct regions generated by an arrangement of m lines is $1+m+m(m-1)/2$ (Stanley, 2004, Section1.1); to these regions we should add a polynomial (in m) number of possible regions *on* the lines themselves. Hence we have a number of regions that is polynomial on the number

of objects (note that this is an upper bound on the number of regions, because many regions that are geometrically possible are not used in the qualitative map, as mentioned in Section 5). In other words, with n objects we have $m = O(n^2)$ lines and $N = O(m^2)$ regions; if the Bayesian filter algorithm has complexity in $O(N^2)$, it follows that it has complexity $O(n^8)$ with respect to the number n of objects (since $O(N^2) = O((m^2)^2) = O(m^4) = O((n^2)^4) = O(n^8)$).

The practical implications of this work reside in the fact that the qualitative representation of possible robot locations facilitates the definition of robot goals in terms of high-level relations. Such relations are both easily understandable by a non-expert user, and directly related to the map of the robot's environment. Additionally, various robot tasks do not require a fully defined metrical map, or the exact location of the agent, but only a relative localisation of the robot with respect to objects in the domain.

The qualitative-probabilistic self-localisation algorithm proposed in this work (called PQS) was evaluated on data from three distinct runs of a robot around coloured objects (red, blue and green boxes) disposed around an indoor corridor connecting various offices. Coloured boxes were chosen as target objects in order to simplify the vision processing task, whose development towards the perception of more complex target objects was outside the scope of this paper.

It is worth reiterating whilst the space around the boxes was free to allow manoeuvring, the robot environment was not completely empty, as there were doors, plants and other objects located around. Data collection was accomplished with humans moving normally within the area.

The experimental results presented in the previous section show that the localisation algorithm was capable of locating the robot in the correct region in most of the images collected. This is represented by the high accuracy values shown in the main diagonal of the confusion matrix (Figure 1). It is worth noting that in the majority of the cases where the algorithm was not capable of providing the exact location, it placed the robot in a neighbouring region of the actual robot's location (according to the map in Figure 4). As mentioned above, these cases were related to positions closer to the border limits of the regions, where the visual distinction between occlusion relations could not be precisely captured by the vision system.

We also compared PQS with a deterministic algorithm (PP algorithm, Algorithm 2), that works by simply comparing the observation with a set of rules relating the robot's location with the observation of object's occlusion. This comparison shows that the Bayesian filter, applied on qualitative information, was successful in improving the accuracy results of the localisation algorithm in regions of high uncertainty (i.e. in locations where the robot could not observe all the occlusion relations necessary to precisely define its position in the map).

The perceptual procedure (PP) method is a deterministic algorithm, that works fine provided that there is enough information to infer a fact (in the present case it infers the heads of Formulae 1 – 28 provided that the relations in their bodies hold). However, in the absence of information it fails. In contrast, the Bayesian filter, implemented in the PQS algorithm, allows the inference of the robot's position in these situations, since it takes into account past perceptions and a previous prediction of the current location, as well as the current perception. This fact is illustrated in Table 3 that relates the number of target objects observed in the scenes (column "#Objects") with the accuracy results for the Perceptual Procedure (column "PP") and the Probabilistic Qualitative Self-Localisation algorithm (Column "PQS") over all regions in the map. In this table we see that PQS greatly outperforms PP in situations where no object, or only one object, is present in the scene, whereas both algorithms have similar performances in situations where more than one object is observed.

Table 3. Algorithm performance with respect to the number of objects per scene.

#Objects	# Images	PP (%)	PQS (%)
0	299	0.0	70.9
1	3867	11.4	44.1
2	4228	86.8	90.2
3	1089	89.7	99.3

It is worth noting also, from Table 3, that PQS had a better performance in the situations where no object was observed, than in situations containing only one object. This is due to the fact that, in the latter situation, the predicted belief was competing with the hypothesis of "total occlusion", whereas there were no competing hypotheses with respect to the former case: the inference was guided mainly by the prediction step of the Bayesian filter.

9.   Conclusion and open issues

This paper developed a novel approach to robot localisation using a qualitative representation about spatial relations implemented within a Bayesian filter. Qualitative representations, in general, provide a higher-level (compact) abstraction than metric data that is closer to the human conceptualisation of the world. Therefore, the ultimate aim of this work is the development of intelligent robot systems capable of reasoning and interacting on the human environment using common-sense knowledge. This goal was highlighted in Thrun (2003) as one of the shortcomings of the pure probabilistic algorithms for robotics, and is on the research agenda of the cognitive robotics field (Levesque and Lakemeyer, 2008; Reiter, 2001). Within this context, the present paper concentrates on a particular aspect of qualitative space representation, which is one of the key aspects in robot navigation using vision sensors: the representation of (and reasoning about) object occlusion.

The main contribution of the work proposed in this paper is providing a higher-level of representation to complement current *Simultaneous Localisation And Map-Building* (SLAM) algorithms in order to leverage human-like knowledge representation and reasoning, and thus to provide an intuitive way of handling localisation at a level of abstraction closer to the human categorisation of space. This is a step toward a more effective interaction between humans and robots. We do not claim that the algorithm investigated in this paper outperforms traditional localisation techniques on metric, topological or semantic maps. Rather than comparing the algorithm proposed in this work with current, numerical, methods for localisation, we intend to combine both approaches in order to use the high-precision of the numerical methods with our localisation procedure based on a *qualitative* way of representing the world. This is a task for future investigations.

We are also currently investigating extensions of the ideas presented in this paper towards the representation of a 3D map that could be used by a flying agent (such as a quadrotor). To this end, there is the need to consider other modalities of qualitative relations in the underlying theory, such as relations about direction or shape. The development of a more robust object recognition module, so that general objects could be used as targets, is also in order. Although the environment in which the robot was immersed was not static (as there were people moving in the corridor) the target objects did not move. Relaxing this constraint is also an issue for future investigations.

We believe that the approach described in this paper could be used within a complete SLAM algorithm, so that the whole qualitative map could be learned from the observations of the robot around the environment. This is another issue for future work.

References

Boal, J., Sánchez-Miralles, A., and Arranz, A. (2014). Topological simultaneous localization and mapping: a survey. *Robotica*, **32**, 803–821.

Chen, S. (2012). Kalman filter for robot vision: A survey. *Industrial Electronics, IEEE Transactions on*, **59**(11), 4409–4420.

Cohn, A. G. and Renz, J. (2008). *Handbook of Knowledge Representation*, chapter Qualitative Spatial Representation and Reasoning, pages 551–596. Elsevier.

Deits, R., Tellex, S., Thaker, P., Simeonov, D., Kollar, T., and Roy, N. (2013). Clarifying commands with information-theoretic human-robot dialog. *Journal of the Human-Robot Interaction*, **2**(2), 58–79.

Dellaert, F., Fox, D., Burgard, W., and Thrun, S. (1999). Monte Carlo localization for mobile robots. In *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, volume 2, pages 1322–1328.

Falomir, Z., Museros, L., Castello, V., and Gonzalez-Abril, L. (2013). Qualitative distances and qualitative image descriptions for representing indoor scenes in robotics. *Pattern Recognition Letters*, **34**(7), 731 – 743. Scene Understanding and Behaviour Analysis.

Fenelon, V., Santos, P., Dee, H., and Cozman, F. (2013). Reasoning about shadows in a mobile robot environment. *Applied Intelligence*, **38**(4), 553–565.

Filliat, D. and Meyer, J.-A. (2003). Map-based navigation in mobile robots:: I. a review of localization strategies. *Cognitive Systems Research*, **4**(4), 243 – 282.

Fogliaroni, P., Wallgrün, J. O., Clementini, E., Tarquini, F., and Wolter, D. (2009). A qualitative approach to localization and navigation based on visibility information. In *COSIT'09: Proceedings of the 9th International Conference on Spatial Information Theory*, pages 312–329, Berlin, Heidelberg. Springer-Verlag.

Fox, D., Burgard, W., and Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, **11**.

Konolige, K., Marder-Eppstein, E., and Marthi, B. (2011). Navigation in hybrid metric-topological maps. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3041–3047.

Leonard, J. and Durrant-Whyte, H. (1991). Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, pages 1442–1447 vol.3.

Levesque, H. and Lakemeyer, G. (2008). *Handbook of Knowledge Representation*, chapter Cognitive Robotics, pages 869–882. Elsevier.

Levitt, T. S. and Lawton, D. T. (1990). Qualitative navigation for mobile robots. *Artificial Intelligence*, **44**, 305–360.

Ligozat, G. (2011). *Qualitative Spatial and Temporal Reasoning*. Wiley.

Liu, M. and Siegwart, R. (2014). Topological mapping and scene recognition with lightweight color descriptors for an omnidirectional camera. *Robotics, IEEE Transactions on*, **30**(2), 310–324.

McClelland, M., Campbell, M., and Estlin, T. (2013). Qualitative relational mapping for planetary rovers. In *Intelligent Robotic Systems: Papers from the AAAI 2013 Workshop*, pages 110 – 113.

Moratz, R. and Ragni, M. (2008). Qualitative spatial reasoning about relative point position.

Journal of Visual Languages & Computing, **19**(1), 75 – 98. Spatial and Image-based Information Systems.

Nüchter, A. and Hertzberg, J. (2008). Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, **56**(11), 915 – 926. Semantic Knowledge in Robotics.

Pereira, V., Santos, P., Martins, M., and Cozman, F. G. (2013). A qualitative-probabilistic approach to autonomous mobile robot self localisation and self vision calibration. In *Proc. of Brazilian Conference on Intelligent Systems*, pages 157 – 162.

Posada, L. F., Hoffmann, F., and Bertram, T. (2014). Visual semantic robot navigation in indoor environments. In *ISR/Robotik 2014; 41st International Symposium on Robotics; Proceedings of*, pages 1–7.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). ROS: an open-source robot operating system. volume 3.

Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. In *Proc. of KR*, pages 165–176, Cambridge, U.S.

Randell, D., Witkowski, M., and Shanahan, M. (2001). From images to bodies: Modeling and exploiting spatial occlusion and motion parallax. In *Proc. of IJCAI*, pages 57–63, Seattle, U.S.

Randell, D. and Witkowski, M. (2002). Building large composition tables via axiomatic theories In *Proc. of KR*, pages 26–35, Toulouse, France.

Ranganathan, A. and Dellaert, F. (2011). Online probabilistic topological mapping. *The International Journal of Robotics Research*, **30**(6), 755–771.

Reiter, R. (2001). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press, U.S.

Remolina, E. and Kuipers, B. (2004). Towards a general theory of topological maps. *Artificial Intelligence*, **152**(1), 47 – 104.

Santos, P. (2007). Reasoning about depth and motion from an observer's viewpoint. *Spatial Cognition and Computation*, **7**(2), 133–178.

Santos, P., Dee, H., and Fenelon, V. (2009). Qualitative robot localisation using information from cast shadows. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 220–225.

Schlieder, C. (1996). *Intelligent image database systems*, chapter Ordering information and symbolic projection, page 115–140. World Scientific Publishing.

Souchanski, M. and Santos, P. (2008). Reasoning about dynamic depth profiles. In *Proc. of the 18th European Conference on Artificial Intelligence (ECAI)*, pages 30–34, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Stanley, R. P. (2004). An introduction to hyperplane arrangements. In *Lecture notes, IAS/Park City Mathematics Institute*.

Stolzenburg, F. (2010). Localization, exploration, and navigation based on qualitative angle information. *Spatial Cognition and Computation*, **10**(1), 28–52.

Tapus, A. and Siegwart, R. (2006). A cognitive modeling of space using fingerprints of places for mobile robot navigation. In *IEEE International Conference on Robotics and Automation*, pages 1188–1193. IEEE.

Tassoni, S., Fogliaroni, P., Bhatt, M., and De Felice, G. (2011). Toward a qualitative model of 3D visibility. In *25th International Workshop on Qualitative Reasoning*.

Thrun, S. (2003). *Exploring Artificial Intelligence in the New Millennium*, chapter Robotic Mapping: A Survey, pages 1–35. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic Robotics*, chapter 8 and 9: Mobile Robot Localization, pages 189–276. Intelligent robotics and autonomous agents series. The MIT Press.

Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. (2007). Cognitive maps for mobile robots-an object based approach. *Robot. Auton. Syst.*, **55**(5), 359–371.

Wolter, D., Freksa, C., and Latecki, L. J. (2008). Towards a generalization of self-localization. In M. E. Jefferies and W.-K. Yeap, editors, *Robotics and Cognitive Approaches to Spatial Mapping*, volume 38 of *Springer Tracts in Advanced Robotics*, pages 105–138. Springer; Berlin.

Zivkovic, Z., Bakker, B., and Krose, B. (2005). Hierarchical map building using visual landmarks and geometric constraints. In *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 2480–2485.