Reasoning about shadows in a mobile robot environment [1]

Author(s):
Paulo E. Santos
Valquiria Fenelon
Hannah M. Dee

Reasoning about shadows in a mobile robot environment

Paulo E. Santos · **Valquiria Fenelon** · **Hannah M. Dee**

Abstract This paper describes a logic-based formalism for qualitative spatial reasoning with cast shadows and presents results of a mobile robot qualitative self-localisation experiment using this formalism. In practice, shadow detection was accomplished by mapping the images from the robot's monocular colour camera into a HSV colour space and then thresholding on V. We present results of self-localisation using two methods for obtaining the threshold automatically: in one method the images are segmented according to their grey-scale histograms, in the other, the threshold is set according to a prediction about the robot's location, given a shadow-based map defined upon a qualitative spatial reasoning theory. This map-related threshold search is the main contribution of the present work, and to the best of our knowledge this is the first work that uses qualitative spatial representations both to perform egolocation and to calibrate a robot's interpretation of its perceptual input.

1 Introduction

Cast shadows as cues for depth perception have been used to enhance depictions of natural scenes since the Renaissance [11]. Recent research within psychology suggests that the human perceptual system gives preferential treatment to information from shadows when inferring motion in depth and perceiving 3D scene layout. These studies suggest that information coming from shadows can override such basic notions as conservation of object size [22,7,24]. Casati in [4] points out that cast shadows also contain information that is not used during passive perception, for instance, information about the presence and location of the

Paulo E. Santos
Centro Universitário da FEI, S. Paulo, Brazil
Tel.: +55-11-4353 2910
Fax: +55-11-4353 2910
E-mail: psantos@fei.edu.br

Valquiria Fenelon
Escola Politécnica
Universidade de S. Paulo, Brazil

Hannah M. Dee
Department of Computer Science
Aberystwyth University, UK

light source and the caster; the intensity of the source; the caster's shape; the screen texture; and the distance between the caster and the screen.

Whilst psychologists have demonstrated the centrality of shadows to our own perception of depth, size and motion, much work in computer vision and robotics starts from the premise that shadows are sources of noise rather than information. The present work falls within the small but growing area of research which aims to use shadows not as sources of noise, but as sources of information. This requires not only a model of the kinds of information that shadows can purvey, but also a robust and accurate shadow detection system. Researchers within both computer vision and robotics have been working in this area – many engaged in shadow suppression in videos from fixed cameras, but some engaged in the more challenging task of shadow identification, localisation and use [14].

The contribution of this paper is the investigation of a qualitative self-localisation method using information from cast shadows. We discuss the experimental evaluation of this method using two techniques for automatically obtaining the threshold for segmenting each image from a robot's camera: in one method the images are segmented according to its grey-scale histogram, in the other method the threshold is searched according to a prediction about the robot's location, given a shadow-based qualitative map.

This paper is organised as follows. Section 2 outlines related research from within both computer vision and robotics. Section 3 describes the theory upon which this work is based: the Perceptual Qualitative Relations about Shadows (PQRS), which formalises the problem of reasoning about shadows within a qualitative spatial reasoning context. The adaptive thresholding methods considered in this work are presented in Section 4, and results are described in Section 5. Discussions are drawn on Section 6 and Section 7 concludes this paper.

Throughout this paper, constants are written in upper-case letters and variables in lower case, unless explicitly stated otherwise.

2 Related research

When considering the task of segmenting moving objects from a static background, shadows are a frequent source of false positives [12,25] and therefore shadow *suppression* is a major research area. In this context, shadow detection in computer vision almost always involves some model of the colour of the *screen* (or, in computer vision terminology: *background*) and detection is performed using a model of shadows characterising them as *'roughly the same colour as the background, but darker'*. Perhaps the simplest shadow detection method proposed is that of [40], in which a grey-scale image is simply thresholded and the darker pixels are labelled *shadow*; however this approach fails on complex images and in situations where lighting changes due to either environmental effects or egomotion. Prati in [31] provides an overview and a taxonomy of early shadow-detection techniques, dividing them into *model-based* and *non-model-based*, however this categorisation does not apply well to more recent works, many of which can be thought of as ensemble methods [25,30].

Cucchiara *et al.* in [12] take as their starting point detected moving objects (and a background model). The pixel values of moving objects are converted to the HSV (Hue, Saturation and Value) colour space, and then these objects are investigated to determine whether they are real moving objects or merely shadow pixels. This is accomplished by considering observed and background values of all three HSV components, considering the difference between foreground and background values for H and S, and the ratio of the two V values. This captures the intuitive observations that shadows are about the same hue as the same

part of the scene unshadowed, slightly more saturated, and darker. A similar approach based upon the observation of colour changes in cast shadows is presented in [35]. Stauder *et al.* in [38] use assumptions about the background (*it will dominate the scene*), the nature of shadows and luminance (*shadows are darker and tend to have uniform shading*) and the presence of moving and static edges. Other methods for shadow filtering use a model of the shadow's caster, either assuming it is rectangular (like a car) as in [42], or upright (from a moving person for instance) as in [17, 34]. These assumptions improve filtering considerably, but break down when shadows from arbitrary objects are considered.

There are a few systems within computer vision that use cast shadows as sources of information rather than noise. The work reported in [3] uses known 3D locations and their cast shadows to perform camera calibration and light location (using known casters and screen to tell about the light source); whereas, Caspi and Werman in [5] use the moving shadows cast by known vertical objects (e.g., flagpoles, or the side of buildings) to determine the 3D shape of objects on the ground (using the shadow to tell about the shape of the screen). Balan *et al.* [1] use shadows as a source of information for detailed human pose recognition: they show that using a single shadow from a fixed light source can provide a similar disambiguation effect as using additional cameras.

In robotics, the story is similar. Fitzpatrick and Torres-Jara in [15], inspired by work suggesting that humans use the shadows of their own limbs when judging limb location [6], track the position of a robotic arm and its shadow cast on a table to derive an estimate of the time of contact between the arm and the table. Shadows are detected in this work using a combination of two methods: in the first, a background model of the workspace is built without the arm and then used to determine light changes when the arm is within the camera view. The second method compares subsequent frames in order to detect moving regions of light change. The authors motivate their work pointing out that depth from shadows and stereopsis may work as complementary cues for robot perception, while the latter is limited to surfaces rich in textures, the former works well in smooth (or even reflective) surfaces. Cheah *et al.* [8] present a novel controller for a robot manipulator, providing a solution to the problem of trajectory control in the presence of kinematic and dynamic uncertainty. In order to evaluate their results, an industrial robot arm was controlled using the visual observation of the trajectory of its own shadow. Lee and colleagues [20] use cast shadows inside pipes to detect landmarks: by fitting bright lights to the front of their pipe inspection robot, they can determine when a pipe bends by detecting cast shadows.

Information from shadows are also considered in unmanned autonomous planetary exploration. Tompkins *et al.* [39] describe an autonomous path planning system that takes into account various conditions of the robot's state, including particularities of the terrain and lighting. In this context, the information about shadows cast by terrain irregularities allows the rover to plan a trajectory that maximises the trade-off between the exposure of the solar cells to sun light and the limited resources in planetary missions. Kunii and Gotoh [19] propose a *Shadow Range Finder* system that uses the shadow cast by a robot arm on the surface of a terrain in order to obtain depth information around target objects, thus providing low-cost, energy-saving, sensors for the analysis of the terrain surrounding rock samples of interest.

More recently, we developed an initial representation of cast shadows in terms of a spatial formalism based on occlusion relations (presented in [36]). This representation, called Perceptual Qualitative relations about Shadows (PQRS), is used in a qualitative self-localisation procedure for a mobile robot in an office-like environment. The present paper builds upon this idea and, therefore, a more complete and accurate description of the PQRS formalism is presented in the next section.

The idea of qualitative self-localisation, however, first appeared in [21] whereby a tessellation of a mobile robot's environment is obtained from the set of lines connecting pairs of pointwise landmarks. The space bounded by these lines define regions, which can then be treated as vertices of a topological map. The spatial representation behind this idea was further developed in [41, 37, 16]. In particular, [16] considers extended convex (instead of pointwise) objects as landmarks and the decomposition of space is based on the notions of occlusion and visibility, which (as we shall see) has much in common to the PQRS formalism investigated in this paper. However, in contrast to [16], the present work shows empirical results on applying these ideas.

3 Perceptual qualitative relations about shadows (PQRS)

Perceptual Qualitative Relations about Shadows (PQRS) [36] is a theory inspired by the idea that shadows provide the observer with the viewpoint of the light source, as shadows are projections of casters from the light source's location. Equivalently, we can say that every point in the shadow region is totally occluded by the caster from the viewpoint of the light source. This idea is developed by representing relations of occlusion and shadows within the scope of the Qualitative Spatial Reasoning (QSR) field of research, which is part of the artificial intelligence sub-area known as Knowledge Representation and Reasoning [13].

The goal of QSR is to provide appropriate formalisms for representing and reasoning about spatial entities, such as part-whole relations, connectivity, orientation, line segments, size and distance, and so on [10, 9]. In practice, QSR formalisms are based on a number of constraints that reflect the structure of space, which are represented as a set of qualitative (i.e., non numerical) relations. With these relations, QSR methods also facilitate the representation of high-level domain knowledge, adding a more abstract (conceptual) level to the systems in which they are applied, including robotics and vision systems. Therefore, research on qualitative spatial reasoning for robotics does not preclude the use of more traditional quantitative methods, but complements them.

The Perceptual Qualitative Relations about Shadows (PQRS) theory shares the same goals of QSR theories, making explicit some of the information content in cast shadows. PQRS assumes the existence of a major, static, light source denoted by L, situated above the observer (in agreement to recent research on the psychophysics of perception [23]). It is also assumed that the scenes are observed from a viewpoint v, and that shadows are cast on a single screen Scr, that is assumed to be much larger than the shadow and is not necessarily planar.

The foundation of PQRS is the QSR theory named the Region Occlusion Calculus (ROC) [33], which is itself built upon one of the best known QSR approaches: the Region Connection Calculus (RCC) [32]. 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, some mereotopological relations between two spatial regions can be defined, such as:

- *disconnected from* (DC): $DC(x, y) \equiv \neg C(x, y)$,
- *part of* (P): $P(x, y) \equiv \forall z(C(z, x) \rightarrow C(z, y))$;
- *equal to* (EQ): $EQ(x, y) \equiv P(x, y) \wedge P(y, x)$,

- *overlaps* (*O*): $O(x, y) \equiv \exists z(P(z, x) \wedge P(z, y))$;
- *partially overlaps* (*PO*): $PO(x, y) \equiv O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$;
- *proper part of* (*PP*): $PP(x, y) \equiv P(x, y) \wedge \neg P(y, x)$;
- *externally connected* (*EC*): $EC(x, y) \equiv C(x, y) \wedge \neg O(x, y)$;
- *tangential proper part* (*TPP*): $TPP \equiv PP(x, y) \wedge \exists z(EC(z, x) \wedge EC(z, y))$;
- *non-tangential proper part* (*NTPP*): $NTPP \equiv PP(x, y) \wedge \neg \exists z(EC(z, x) \wedge EC(z, y))$.

RCC also includes the inverse relations of P, PP, TPP and $NTPP$, which are represented by a capital 'I' in front of the relative relation: $PI, PPI, TPPI$ and $NTPPI$.
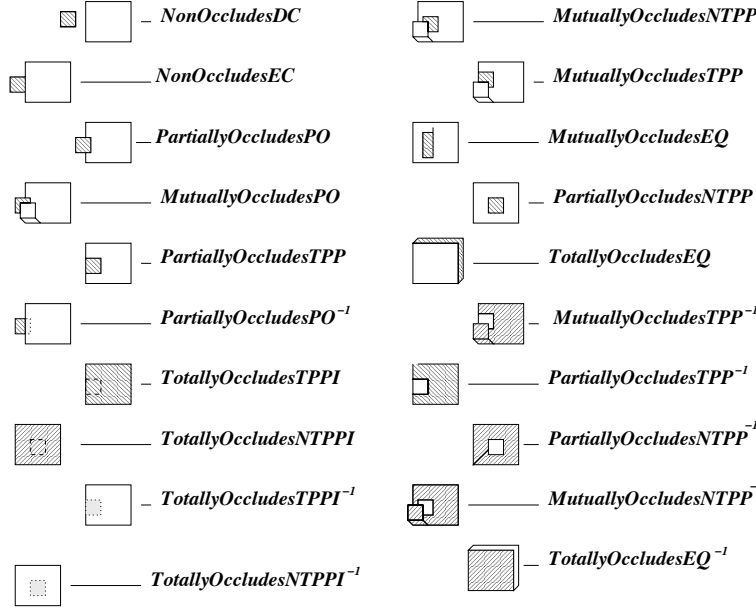

Fig. 1 The 20 ROC relations.

Using RCC relations, along with the primitive relation $TotallyOccludes(x, y, v)$ (which stands for "x totally occludes y with respect to the viewpoint v"), the Region Occlusion Calculus (ROC) represents the various possibilities of interposition relations between two arbitrarily-shaped objects. In particular, with RCC and the primitive $TotallyOccludes/3$, it is possible to define occlusion relations for *non occlusion* ($NonOccludes/3$), *partial occlusion* ($PartiallyOccludes/3$) and *mutual occlusion* ($MutuallyOccludes/3$). In fact, [33] defines 20 such relations (shown in Figure 1).

The 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 made by assuming two functions: the function $region(x)$, which maps a body x to its occupancy region, and the function $image(x, v)$ that maps a body x (and the viewpoint v) to the body's 2D image, as seen from viewpoint v. It is worth pointing out that the viewpoint in ROC is modelled as a pinhole camera whose parameters, however, are not important for the qualitative theory.

Using these definitions, ROC is constrained by a number of axioms, of which we only cite two ($A1$ and $A2$), since these are relevant for proving the PQRS theorems proposed below.

Formula ($A1$) is the ROC axiom that states that "if x totally occludes y from a viewpoint v, x totally occludes any part of y":

$$(A1) \quad \forall x \, y \, z \, v \;\; [[TotallyOccludes(x, y, v) \land P(region(z), region(y))] \rightarrow$$
$$TotallyOccludes(x, z, v)].$$

Formula ($A2$) below states that "if x totally occludes y from v, no part of y totally occludes any part of x":

$$(A2) \quad \forall x \, y \, z \, v \;\; [TotallyOccludes(x, y, v) \rightarrow$$
$$\forall z \, u[[P(region(z), region(x)) \land P(region(u), region(y))] \rightarrow$$
$$\neg TotallyOccludes(u, z, v)]]$$

In order to simplify notation we use the following abbreviation [33]:

$$Occludes(x, y, v) \equiv \qquad\qquad\qquad\qquad\qquad\qquad\qquad (1)$$
$$\exists z \, \exists u \; P(region(z), region(x)) \land$$
$$P(region(u), region(y)) \land$$
$$TotallyOccludes(z, u, v)$$

Considering the ROC relations between a caster o and its shadow s, from a viewpoint v (and the fact that a shadow never occludes its caster, as proved in theorem (T1) below) only the following ROC relations have models in PQRS:

- $NonOccludesDC(o, s, v)$, meaning that o and s do not occlude each other from the viewpoint v, and $DC(image(o, v), image(s, v))$ (the other relations below have analogous descriptions);
- $NonOccludesEC(o, s, v)$,
- $PartiallyOccludesPO(o, s, v)$,
- $PartiallyOccludesTPP(o, s, v)$,
- $PartiallyOccludesNTPP(o, s, v)$,
- $TotallyOccludesTPPI(o, s, v)$,
- $TotallyOccludesEQ(o, s, v)$, and
- $TotallyOccludesNTPPI(o, s, v)$.

Figure 2 represents these relations, where the dashed object is the caster and the blank is its shadow.

Apart from the ROC relations inherited by PQRS, this theory also assumes the primitive $Shadow(s, o, Scr, L)$ that represents that a shadow s is cast by a caster o, from the light source L, on the screen Scr. The axiom constraining the $Shadow/4$ relation is represented by Formula ($A3$) below.

$$(A3) \quad Shadow(s, o, Scr, L) \leftrightarrow \neg \exists o'(o \neq o')$$
$$\land Occludes(o', o, L) \land PO(region(s), region(Scr)) \land$$
$$TotallyOccludes(o, s, L).$$
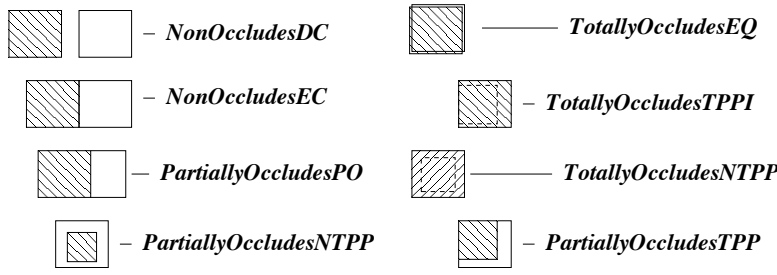
Fig. 2 The ROC relations that are part of PQRS, where the dark object is the caster and the white its cast shadow.

Axiom($A3$) states that the shadow of a caster o is the region in a screen Scr that is totally occluded by o from the light source viewpoint L (if there is no other object o' occluding o from L).

With this formalism it is possible to prove a number of theorems about commonsense facts. For instance, it follows directly from Axioms ($A2$) and ($A3$) that no shadow occludes its own caster[1], as denoted by Theorem ($T1$) below.

$$(T1) \quad Shadow(s, o, Scr, L) \rightarrow \neg TotallyOccludes(s, o, L).$$

It is also a consequence of Axiom ($A3$) and the ROC axioms that no shadow casts a shadow itself (cf. Theorem ($T2$)):

$$(T2) \quad Shadow(s, o, Scr, L) \rightarrow \neg \exists s' Shadow(s', s, Scr, L).$$

Proof: Let's assume, reasoning by contraposition, that both (a) $Shadow(s, o, Scr, L)$ and (b) $Shadow(s', s, Scr, L)$ are true, for an object o and shadows s' and s. From (b) and axiom ($A3$) we have that $TotallyOccludes(s, s', L)$, similarly from (a) and ($A3$), we have that $TotallyOccludes(o, s, L)$. From the transitivity of $TotallyOccludes/3$, we have: $TotallyOccludes(o, s', L)$. Therefore, from ($A3$) we conclude that $Shadow(s', s, Scr, L)$ is false, since the condition

$$\neg \exists o'(o \neq o') \wedge Occludes(o', o, L)$$

does not hold.

$\square$
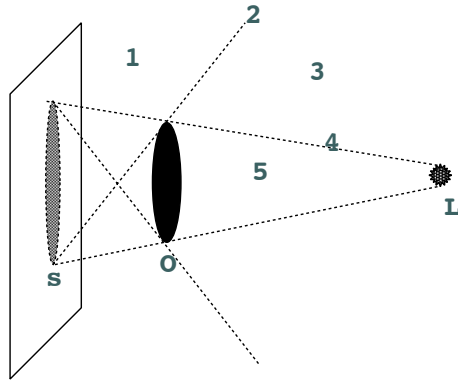
We can also prove that if two shadows of distinct objects partially overlap, then the objects will be in a relation of occlusion with respect to the light source, as expressed in Theorem ($T3$).

$$(T3) \quad Shadow(s, o, Scr, L) \wedge Shadow(s', o', Scr, L) \wedge$$
$$\wedge O(region(s'), region(s))$$
$$\rightarrow Occludes(o, o', L) \vee Occludes(o', o, L).$$

---

[1] Note that we are only dealing with cast, and not self attached, shadows.

Fig. 3 Distinct regions implied by the observation of a shadow and its caster. It is worth noting that, in this figure, regions 2 and 4 have zero-width boundaries.

Proof:
From $Shadow(s, o, Scr, L)$ and axiom $(A3)$ we have (a) $TotallyOccludes(o, s, L)$, and similarly from $Shadow(s', o', Scr, L)$ we have (b) $TotallyOccludes(o', s', L)$. From $O(region(s'), region(s))$ and the definition of $O/2$ we have (c) $\exists z P(z, s) \land P(z, s')$.
From $(A1)$, (a) and (b) it is true that o totally occludes any part of s (and analogously to o' and s'), thus from (c) it is true that:
$\exists z P(z, s) \land P(z, s') \land TotallyOccludes(o, z, L) \land TotallyOccludes(o', z, L)$.
By construction, o and o' are disconnected, therefore the previous formula implies that either $occludes(o, o', L)$ or $occludes(o', o, L)$.

$\square$

Theorem $(T3)$ is an example of an inference that presupposes a relational theory, i.e., it is a result that cannot be achieved with the traditional numerical methods in robotics.

In contrast with our earlier work (presented in [36]), axiom $A3$ above has a clearer definition, as it is constrained to situations where no interference of shadows from other objects occurs. Additionally, this paper is the first to present complete proofs of the theorems which make up PQRS.

3.1 Relative location

The formalism given above can be used to reason about shadows from arbitrary viewpoints: relating shadows with occlusion suggests the distinction of five regions defined from the lines of sight between the light source, the caster and its shadow as represented in Figure 3. This idea can also be applied to the usual case of a shadow connected to the caster (e.g. Figures 5(b) and 5(c)), in this case, we only use the portion of the shadow farthest from the caster (which we shall call *top-half* for convenience) to build the diagram. In order to formally state this distinction between whole shadows and their top-half parts, we introduce a function $top\_h(s)$ that maps a shadow (s) to its top half part[2] (this is analogous to the $region$ and $image$ functions used in ROC, as mentioned above).

[2] This distinction was missing in [36].

Therefore, considering the diagram in Figure 3, any viewpoint v located on Region 1 will observe the (top half of) shadow s and the object o as $NonOccludesDC(o, top\_h(s), v)$; similarly, if v observes o and $top\_h(s)$ from Region 3 it should see that $PartiallyOccludes$-$PO(o, top\_h(s), v)$ and from Region 5 that $TotallyOccludesNTPPI(o, top\_h(s), v)$. Region 4 is the surface defined by the lines of sight from l tangential to o and $top\_h(s)$, from where v would observe $TotallyOccludesTPPI(o, top\_h(s), v)$. In Region 2, v perceives object and shadow as $NonOccludesEC(o, top\_h(s), v)$. These facts are included in PQRS by the axioms $(A4)$ to $(A8)$ below, where the predicate $located(o, r, v, s)$ represents that object o is located at region r according to the viewpoint v and shadow s.

$(A4)$ $located(o, Region\ 1, v, s) \leftarrow \exists scr \exists l\ Shadow(s, o, scr, l) \wedge$
$$NonOccludesDC(o, s, v) \wedge v \neq o;$$

$(A5)$ $located(o, Region\ 2, v, s) \leftarrow \exists scr \exists l\ Shadow(s, o, scr, l) \wedge$
$$NonOccludesEC(o, s, v) \wedge v \neq o;$$

$(A6)$ $located(o, Region\ 3, v, s) \leftarrow \exists scr \exists l\ Shadow(s, o, scr, l) \wedge$
$$PartiallyOccludesPO(o, s, v) \wedge v \neq o;$$

$(A7)$ $located(o, Region\ 4, v, s) \leftarrow \exists scr \exists l\ Shadow(s, o, scr, l) \wedge$
$$TotallyOccludesTPPI(o, s, v) \wedge v \neq o;$$

$(A8)$ $located(o, Region\ 5, v, s) \leftarrow \exists scr \exists l\ Shadow(s, o, scr, l) \wedge$
$$TotallyOccludesNTPPI(o, s, v) \wedge v \neq o.$$

Regions 2 and 4 in Figure 3 are in fact *boundaries* separating regions 1 and 3, and between 3 and 5 respectively. Therefore, it is virtually impossible for a robot to locate itself on them. In the real robot environment, however, regions 2 and 4 are extended assuming an interval of uncertainty around these boundaries. Figure 4 represents the diagram obtained assuming the extended versions of regions 2 and 4.
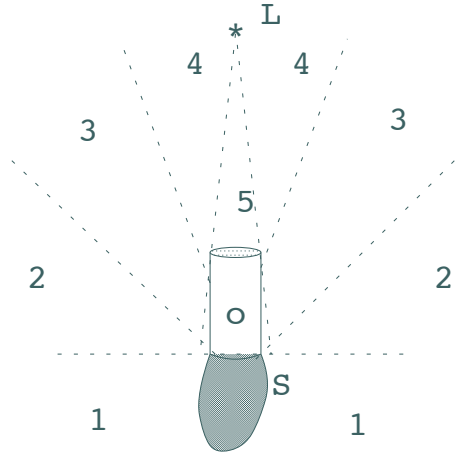
Fig. 4 Regions implied by the observation of a shadow and its caster, where L is the light source, O is the object (caster) and S is its shadow. .

(a) robot environment   (b) Two shadows   (c) Example segmented image
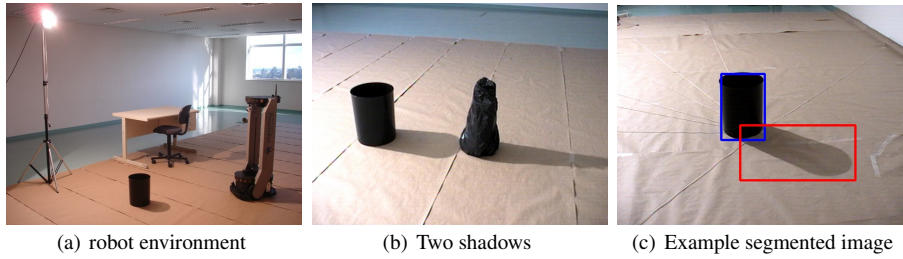
Fig. 5 Images depicting environment, sample input (with ambiguous shadow-base relations) and segmented object/shadow.

Using ROC it is also possible to relate shadows with the relative distance of objects from a viewpoint. The region occlusion calculus relates occlusion with relative distance using the relation $nearness$ ($N(x,y,z)$), read as "x is nearer to y than x is to z", that was first defined in [2]. *Nearness* is incorporated into ROC by the following axiom [33]:

$$\forall x\ y\ v\ [PartiallyOccludes(x,y,v) \rightarrow N(v,x,y)]$$

representing that "if a body x partially occludes a body y with respect to some viewpoint v then x is nearer to v than y is to v".

Within PQRS this implies the commonsense fact that $N(L,o,s)$ (for a light source L, an object o and its shadow s) and consequently that $N(L,o,Scr)$: "the light source is nearer the caster than it is to its shadow". This fact, allied with the qualitative egolocation (according to the diagram 4), facilitates inferences about relative depth from monocular views in some cases.

4 Robot environment and methods

The idea for qualitative robot self-localisation using cast shadows, presented in the previous section, was implemented on our Pioneer PeopleBot mobile robot using its monocular colour camera to obtain snapshots of objects and their shadows in an office-like environment, containing a major (but not necessarily single) light source, as shown in Figure 5(a). Shadow detection was accomplished by first mapping the images captured by the camera into a HSV colour space. These images were then segmented by thresholding on V, whereby high values (light objects) are filtered out and low values (dark objects) are casters. Shadows are located within a value range in between light and dark objects. Morphological operators and the saturation value were used to filter noise (such as reflections of the light source on the object or background shadows). The robot was set to navigate through the room, stopping after a certain time interval to analyse its position with respect to object-shadow locations according to the diagram shown in Figure 4. An example of the snapshots used in this work is shown in Figure 5(c). Shadow correspondence, which is the problem of matching each shadow to its caster [22,24], is solved in this work by assuming a simple heuristic: the shadow that is connected to an object's base is the shadow of this object. When there are various shadows connected to the object's base, the caster is associated with the shadow that is further away from the light source (Fig. 5(b) shows an example of such situation).

Given a threshold Th, a $Scene$ and a viewpoint v, Algorithm 1 summarises the method for self-localisation described in this section.

Algorithm 1 PERCEPTION_ACTION(Th, Scene, v)

1: segment $Scene$ using the threshold Th to obtain a caster O and the top half of its shadow $top\_h(S)$
2: **if** $NonOccludesDC(O, top\_h(S), v)$ **then**
3:   robot is on region 1
4: **else if** $NonOccludesEC(O, top\_h(S), v)$ **then**
5:   robot is on region 2
6: **else if** $PartiallyOccludesPO(O, top\_h(S), v)$ **then**
7:   robot is on region 3
8: **else if** $TotallyOccludesTPPI(O, top\_h(S), v)$ **then**
9:   robot is on region 4
10: **else if** $TotallyOccludesNTPPI(O, top\_h(S), v)$ **then**
11:   robot is on region 5
12: **else**
13:   FAIL
14: **end if**

In Algorithm 1 the ROC relations between a caster O and its shadow S are evaluated according to a threshold on the distance between the (top part of) the shadow's bounding box when *Non Occlusion* holds. If the shadow is in some degree occluded by its caster, from the observer's viewpoint, the ROC relation is evaluated according to a percentage of the shadow's bounding box that can be observed from behind the caster: $PartiallyOccludes$-$PO(O, top\_h(S), v)$ is interpreted when more than (or equal to) 10% of the bounding box is observed, but not all of it; $TotallyOccludesTPPI(O, top\_h(S), v)$ is assumed when less than 10% is still observed; and, when no part of the shadow is seen from behind the caster, $TotallyOccludesNTPPI(O, top\_h(S), v)$ is concluded.

4.1 Adaptive thresholds for foreground/background segmentation

In this work we investigate the use of two distinct methods for automatically finding the best threshold for each given image: the traditional Otsu's method [29] and a threshold search related to the robot's prediction. The latter is the main contribution of the present paper.

Otsu's method [29] works by finding the threshold (t) that maximises the inter-class variance σ between two groups of pixels. Formula 2 expresses σ in terms of the threshold-dependent class probabilities ($\omega_1(t)$ and $\omega_2(t)$) and class means ($\mu_1(t)$ and $\mu_2(t)$) of groups 1 and 2.

$$\sigma^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{2}$$

The second method for finding the best threshold uses the knowledge about the robot's previous location in order to make a prediction about its current location. This procedure works as follows. The robot has to start in a known region. From this position the robot moves to another region (according to the diagram in Figure 4) in a moving action that is currently preprogrammed. In this new position the robot captures a snapshot of the target object and uses it to decide on its location. If the location interpreted matches the prediction of its current position, then the robot moves on. If not, the robot varies the threshold until it finds a match between its predicted and interpreted positions, or fails otherwise. This method is summarised in Algorithm 2 below.

In the pseudocode THRESHOLD AND POSITION (Algorithm 2), the function MOV-ING_ACTION(s_0, v, dir, I) gives the prediction of the robot's position s_i after its motion from the position s_0, with speed v, direction dir and for a time interval I; the function

PERCEPTION_ACTION($th, Scene, v$) outputs the perceived robot's position according to the observed PQRS relation, for a threshold th, a $Scene$ and a viewpoint v (as represented in Algorithm 1), th_{aux} is an auxiliary variable for threshold and s_0, s_i and sj are variables for the robot's position. The constant $Step$ is used to update the threshold from its minimum (Th_{min}) to its maximum (Th_{max}) values (in this work these constants were set at $Step = 5, Th_{min} = 40$ and $Th_{max} = 230$).

Algorithm 2 THRESHOLD AND POSITION($th_0, t_0, scene$)

1: $s_i = $ MOVING_ACTION(s_0, v, dir, I)
2: $s_j = $ PERCEPTION_ACTION($th_0, Scene, v$)
3: **if** ($s_i == s_j$) **then**
4:      return (th_0, s_i)
5: **else**
6:      $th_{aux} = Th_{min}$
7:      **while** (($si \neq sj$) and ($th_{aux} < Th_{max}$)) **do**
8:          $th_{aux} = th_{aux} + Step$
9:          $s_j = $PERCEPTION_ACTION($th_{aux}, Scene, v$)
10:      **end while**
11:      **if** ($th_{aux} >= Th_{max}$) **then**
12:          return FAIL
13:      **else**
14:          return (th_{aux}, s_j)
15:      **end if**
16: **end if**

To the best of our knowledge this is the first paper where the segmentation threshold is obtained as a result of the robot's prediction of its location according to a qualitative map. In this way, we use the PQRS theory not only for robot self-localisation based upon shadow perception, but also for the refinement of the shadow perception itself. The next section presents an empirical evaluation of this technique.

5 Experiments

This section describes the results of the experiments on robot localisation with respect to the map in Figure 4. In these experiments, the robot collected 1361 snapshots around the target object, which provides the frame of reference (e.g. the black bucket in Figure 5(c)). This target was always within camera view, but not necessarily at its centre. We allowed up to three objects within the robot's field of view. Due to the narrow view of the robot, and the use of a single dominant light source, localisation estimates with respect to each object do not contradict one another.

The baseline experiment uses fixed thresholds for image analysis chosen by experimentation within one of the camera views.

The results of the baseline experiment are represented on the third column of Table 1, which shows a poor global performance of the system (47%) on localising the robot in every region. A high accuracy was obtained in the specific region used to calibrate the threshold (above 70% with respect to region 1), but within other regions the results where lower than or equal to 50%. The poor performance outside of region 1 is because the foreground/background segmentation is not optimal for images obtained under other light conditions (i.e., the distinct position configurations between robot, caster and light produced

by the agent's motion). In fact, by tweaking the thresholds, the system improved its performance in locating the robot on other regions, however this improvement came at the expense of losing accuracy on region 1.

Table 1 Percentage of correct answers from using fixed thresholds, Otsu's method and the knowledge-based adaptive thresholding, where "# images" is the number of snapshots taken at each region.

Region	# images	Fixed threshold (%)	Otsu (%)	Knowledge based (%)
1	320	73	59	93
2	436	27	45	88
3	438	50	36	94
4	111	40	24	92
5	56	38	25	86
Global	1361	47	43	91

The obvious approach for improving the poor results obtained by fixed-thresholding is to *adjust the thresholds for each snapshot taken*. The technique we have used to perform this adjustment is the Otsu method [29] (cf. Section 4.1). This should be able to automatically find the threshold for segmenting objects of interest (i.e. casters and their shadows) from background. The results obtained are represented on the fourth column of Table 1.

The results obtained with a variable threshold method (Otsu's method), surprisingly, were slightly worse than those obtained with a fixed threshold. For global localisation, the method answered correctly on 43% of the total 1361 snapshots. The localisation at region 1 was correct in 59% of the trials (decreasing from the 70% obtained with a fixed threshold), and the localisation accuracy on the other regions was below 50%. Investigation of the pixel value distributions indicated that the problem is that these distributions are not in general bi-modal, which increases the difficulty of searching for an appropriate threshold from the image histogram.

In our third set of results, the robot was set to vary the threshold until the interpretation of the target object and its shadow matches a robot's prediction of its location (using Algorithm 2, as explained in Section 4.1). The results obtained are represented in the fifth column of Table 1 ("*knowledge based*"), which shows that the system achieved an accuracy of around 90% in all regions. Thus the incorporation of knowledge about shadow appearance, and reasoning based upon past location, can greatly assist in the refinement of a simple shadow-detection algorithm, outperforming also a traditional algorithm for adaptive thresholding.

6 Discussion and open issues

In this work we investigated robot self-localisation using qualitatively distinct regions defined from a visual observation of cast shadows. These regions were defined on a formal framework within a qualitative spatial reasoning standpoint, which provides a principled approach to reasoning about space.

Central to the problem of qualitative self-localisation using shadows is the segmentation of casters and shadows from the background, which was accomplished here by thresholding on value (V) on the HSV colour space. In the present paper we proposed a new strategy for calibrating this threshold, where the prediction about the robot's location is used to search for a match between the interpreted position (as given from visual observation) and its predicted location. In order to evaluate this method, we have presented three sets of experiments in

which different ways of defining the threshold were tried. In the first set of experiments a hand-coded fixed threshold was used, in the second set of experiments we used Otsu's method [29] in order to find the threshold values from the image histogram. Finally, the third set of experiments presents the results of applying our proposed method for matching the prediction with the observation.

The intuition behind the experiments with fixed thresholds was to provide a lower-bound for the evaluation of our idea, since (as we hypothesised) nothing could perform worse than a hand-coded threshold. Experiments with Otsu's method were then to set the standard, as this is one of the most traditional methods for adaptive thresholding. However, it turned out that Otsu's performance was in fact approximately as accurate as that of using the fixed threshold. This is due to the fact that it was chosen for the first set of experiments the best threshold that could be found, after a number of trials where the value was changed by hand. Otsu's method, however, had to deal with arbitrary images, where it had to maximise a value that is dependent on an *a priori* hypothesis of bi-modal pixel distribution. This was not the case in some of the snapshots taken by the robot: a great number of them suffered from the effect of reflections of the light source on the caster. Moreover, from some angles, there was a negative gradient of luminosity just behind the object. These problems caused Otsu's method to perform worse than using a fixed threshold.

In contrast, the method for calibrating the threshold using the prediction about the robot's location performed as well as could be expected, obtaining an accuracy of around 90% with respect to our dataset containing 1361 snapshots of the robot's environment. However, this method is totally dependent on the capability of the robot's actuators on generating accurate predictions for the robot's future location, given a moving action. In this paper, the robot's motion was completely pre-programmed in order to minimise actuator noise, this gave us the guarantee that we were only evaluating the localisation procedures. We leave for a future work applying this framework on a system that has a path planning module, so that we can determine how the qualitative localisation procedure proposed in this paper is affected by errors in the planning-acting-sensing cycle. Small errors in the path planning procedure could then be solved by allowing a search for thresholds agreeing with the neighbouring regions from the predicted location. Evaluating the ideas put forward in this paper on a more complex scenario is also a future goal.

Apart from the investigation of qualitative robot localisation from cast shadows, contrasting methods for adaptive thresholding, this paper also presented in Section 3 an improved version of the theory first proposed in [36]. The main improvements are: the inclusion of theorems' proofs, the explicit representation of a shadow's top-half as a function and the introduction of a stronger version of axiom ($A3$). The improved version of axiom ($A3$) includes a constraint ruling out the interference of shadows from other objects on the target shadow. Without this, however, the theory presented in [36] leads to inconsistent answers with respect to the intended inferences.

Subject of our long term investigation is the complete exploration of the knowledge content of shadows, as described in [4], in order to create a robotic system that is capable of perceiving (and interpreting) shadows in a similar fashion to humans. The reason for pursuing this goal resides in our hypothesis that the human perceptual system, by preferring shadow information over other depth cues (even when these cues contradict each other [24]), is in fact saving processing time. Investigating how this could be accomplished in robotic systems is a major motivation of this work.

Although this work explores only a qualitative theory about space, this choice does not preclude the use of quantitative or statistical methods. Rather, we believe that qualitative methods in robotics should complement the traditional numerical algorithms, providing an

additional processing level in situations where it is possible to extract qualitative information from such methods.

7 Conclusion

This paper has demonstrated how the incorporation of qualitative spatial representation and *a priori* knowledge about shadow regions can be combined to enhance a simple shadow-detection algorithm based upon thresholding. Future work will consider the incorporation of more sophisticated shadow detection algorithms, the extension of the current snapshot-based system to one which incorporates continuous video, and the inclusion of shadow reasoning within the perception-planning-action loop.

A number of questions have been raised by this work, and we consider these questions in themselves to be a useful contribution. For example, how can shadows improve object localisation when contrasted to object-based methods? Under what conditions can shadows be effectively exploited? How can we combine predictive shadow-based localisation with predictive localisation based upon object's pose? These are all questions which we hope to consider in more depth in future work.

8 Acknowledgements

References

1. Balan, A.O., Black, M.J., Haussecker, H., Sigal, L.: Shining a light on human pose: On shadows, shading and the estimation of pose and shape. Proc. International Conference on Computer Vision (ICCV) (2007)
2. van Benthem, J.: The logic of time: a model theoretic investigation into the varieties of temporal ontology and temporal discourse. Kluwer Academic, Dordrecht, The Netherlands (1991)
3. Cao, X., Foroosh, H.: Camera calibration and light source orientation from solar shadows. Proc. Computer Vision and Pattern Recognition (CVPR) **105**(1), 60–72 (2007)
4. Casati, R.: The shadow knows: a primer on the informational structure of cast shadows. Perception **33**(11), 1385–1396 (2004)
5. Caspi, Y., Werman, M.: Vertical parallax from moving shadows. In: Proc. Computer Vision and Pattern Recognition (CVPR). New York, USA (2006)
6. Castiello, U., Lusher, D., Burton, C., Disler, P.: Shadows in the brain. Journal of Cognitive Neuroscience **15:6**, 862–872 (2003)
7. Cavanagh, P.: The artist as neuroscientist. Nature **434**, 301–307 (2005)
8. Cheah, C.C., Liu, C., Slotine, J.J.E.: Adaptive tracking control for robots with unknown kinematic and dynamic properties. I. J. Robotic Res. **25**(3), 283–296 (2006)
9. Cohn, A.G., Hazarika, S.M.: Qualitative spatial representation and reasoning: An overview. Fundamenta Informaticae **46**(1-2), 1–29 (2001)
10. Cohn, A.G., Renz, J.: Qualitative spatial representation and reasoning. In: Handbook of Knowledge Representation, pp. 551–596. Elsevier (2008)
11. da Costa Kauffmann, T.: The perspective of shadows: The history of the theory of shadow projection. Journal of the Warburg and Courtauld Institutes **38**, 258–287 (1979)
12. Cucchiara, R., Grana, C., Neri, G., Piccardi, M., Prati, A.: The sakbot system for moving object detection and tracking. In: Video-based Surveillance Systems: Computer Vision and Distributed Processing (Part II - Detection and Tracking), pp. 145–158. Kluwer Academic Publishers (2001)
13. Davis, R., Shrobe, H., Szolovits, P.: What is a knowledge representation? AI Magazine **14**(1), 17–33 (1993)

14. Dee, H., Santos, P.: The perception and content of cast shadows: an interdisciplinary review. Spatial Cognition and Computation (2011). To appear

15. Fitzpatrick, P., Torres-Jara, E.: The power of the dark side: using cast shadows for visually-guided touching. In: Proc. of the 4th IEEE/RAS International Conference on Humanoid Robots, pp. 437– 449 (2004)

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

17. Hsieh, J.W., Hu, W.F., Chang, C.J., Chen, J.S.: Shadow elimination for effective moving object detection by Gaussian shadow modeling. Image and Vision Computing **21(6)**, 505–516 (2003)

18. Huang, J.B., Chen, C.S.: Moving cast shadow detection using physics-based features. In: Proc. Computer Vision and Pattern Recognition (2009)

19. Kunii, Y., Gotoh, T.: Evaluation of Shadow Range Finder: SRF for Planetary Surface Exploration. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp. 2573–2578 (2003)

20. Lee, J.S., Roh, S.G., Kim, D.W., Moon, H., Choi, H.R.: In-pipe robot navigation based upon the landmark recognition system using shadow images. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp. 1857–1862 (2009)

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

22. Mamassian, P.: Impossible shadows and the shadow correspondence problem. Perception **33**, 1279–1290 (2004)

23. Mamassian, P., Goutcher, R.: Prior knowledge on the illumination position. Cognition **81**(1), B1–B9 (2001)

24. Mamassian, P., Knill, D.C., Kersten, D.: The perception of cast shadows. Trends in cognitive sciences **2(8)**, 288–295 (1998)

25. Martel-Brisson, N., Zaccarin, A.: Learning and removing cast shadows through a multidistribution approach. IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI) **29 (7)**, 1134–1146 (2007)

26. Martel-Brisson, N., Zaccarin, A.: Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation. In: Proc. Computer Vision and Pattern Recognition (2008)

27. Maxwell, B., Friedhoff, R., Smith, C.: A bi-illuminant dichromatic reflection model for understanding images. In: Proc. Computer Vision and Pattern Recognition (2008)

28. Nadimi, S., Bhanu, B.: Physical models for moving shadow and object detection in video. IEEE transactions on Pattern Analysis and Machine Intelligence **26(8)**, 1079– 1087 (2004)

29. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. (9), 62–66 (1979)

30. Porikli, F., Thornton, J.: Shadow flow: A recursive method to learn moving cast shadows. In: Proc. International Conference on Computer Vision (ICCV) (2005)

31. Prati, A., Mikic, I., Trivedi, M., Cucchiara, R.: Detecting moving shadows: algorithms and evaluation. IEEE transactions on Pattern Analysis and Machine Intelligence **25(7)**, 918–923 (2003)

32. Randell, D., Cui, Z., Cohn, A.: A spatial logic based on regions and connection. In: International Conference on Knowledge Representation and Reasoning, pp. 165–176. Cambridge, U.S. (1992)

33. Randell, D., Witkowski, M., Shanahan, M.: From images to bodies: Modeling and exploiting spatial occlusion and motion parallax. In: International Joint Conference on Artificial Intelligence, pp. 57–63. Seattle, U.S. (2001)

34. Renno, J.R.R., Orwell, J., Thirde, D.J., Jones, G.A.: Shadow classification and evaluation for soccer player detection. In: Proc. British Machine Vision Conference, pp. 839–848. Kingston upon Thames, UK (2004)

35. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using invariant color features. Computer Vision and Image Understanding (CVIU) **95**(2), 238–259 (2004)

36. Santos, P., Dee, H., Fenelon, V.: Qualitative robot localisation using information from cast shadows. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA), pp. 220–225 (2009)

37. Schlieder, C.: Representing visible locations for qualitative navigation. In: Qualitative Reasoning and Decision Technologies, pp. 523–532. CIMNE (1993)

38. Stauder, J., Mech, R., Ostermann, J.: Detection of moving cast shadows for object segmentation. IEEE Transactions on multimedia **1(1)**, 65–76 (1999)

39. Tompkins, P., Stentz, A., Whittaker, W.L.: Automated surface mission planning considering terrain, shadows, resources and time. In: Proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS '01), Montreal, Canada. (2001)

40. Troccoli, A., Allen, P.K.: A shadow-based method for model registration. In: Computer Vision and Pattern Recognition Workshops (2004)

41. Wagner, T., Huebner, K.: Egocentric qualitative spatial knowledge representation for physical robots. In: in 8th International Workshop on RoboCup 2004 (Robot World Cup Soccer Games and Conferences), Lecture Notes in Artificial Intelligence, pp. 134–149. Springer (2004)
42. Wang, Y., Loe, K.F., Tan, T., Wu, J.K.: A dynamic Hidden Markov Random Field Model for foreground and shadow segmentation. In: Proc. IEEE Workshop on Applications of Computer Vision (2005)