

Describing Images using Qualitative Models and Description Logics

Zoe Falomir, Ernesto Jiménez-Ruiz
M. Teresa Escrig, Lledó Museros
University Jaume I, Castellón, Spain

Our approach describes any digital image qualitatively by detecting regions/objects inside it and describing their visual characteristics (shape and colour) and their spatial characteristics (orientation and topology) by means of qualitative models. The description obtained is translated into a description logic (DL) based ontology, which gives a formal and explicit meaning to the qualitative tags representing the visual features of the objects in the image and the spatial relations between them. For any image, our approach obtains a set of individuals that are classified using a DL reasoner according to the descriptions of our ontology.

Keywords: qualitative shape, qualitative colours, qualitative orientation, topology, ontologies, computer vision

1 Introduction

Using computers to extract visual information from space and interpreting it in a meaningful way as human beings can do remains a challenge. As digital images represent visual data numerically, most image processing has been carried out by applying mathematical techniques to obtain and describe image content.

From a cognitive point of view, however, visual knowledge about space is qualitative in nature (Freksa, 1991). The retinal image of a visual object is a quantitative image in the sense that specific locations on the retina are stimulated by light of a specific spectrum of wavelengths and intensity. However, the knowledge about a retinal image that can be retrieved from memory is qualitative. We cannot re-

trieve absolute locations, wavelengths and intensities from memory. We can only recover certain qualitative relationships between features within the image or between image features and memory features. Qualitative representations of this kind are similar in many ways to “mental images” (Kosslyn, 1994; Kosslyn et al., 2006) that people report on when they describe what they have seen from memory or when they attempt to answer questions on the basis of visual memories.

Extracting semantic information from images as human beings can do is still an unsolved problem in computer vision. The approach presented here can describe any digital image qualitatively and then store the results of the description as facts according to an ontology, from which new knowledge in the application domain can be inferred by the reasoners. The association of meaning with the representations obtained by robotic systems, also known as the *symbol-grounding problem*, is still a prominent issue within the field of Artificial Intelligence (AI) (Kuhn et al., 2007; Williams, 2008; Williams et al., 2009). Therefore, in order to contribute in this field, the first application on which our approach has been tested is extracting semantic information from images captured by a robot camera in indoor environments. This semantic information will support robot self-localization and navigation in the future.

As Palmer (1999) points out, in an image, different colours usually indicate different objects/regions of interest. Cognitively, this is the way people process images. Therefore, in our approach, a graph-based region segmentation method based on intensity differences (Felzenszwalb and Huttenlocher, 2004)¹ has been used in order to identify the relevant regions in an image. Then the visual and spatial features of the regions are computed. The visual features of each region are described by qualitative models of shape (Falomir et al., 2008) and colour, while the spatial features of each region are described by qualitative models of topology (Egenhofer and Al-Taha, 1992) and orientation (Hernández, 1991; Freksa, 1992).

We have adopted description logics (DL) (Baader et al., 2003) as the formalism for representing the low-level information from image analysis and we have chosen OWL 2² (Horrocks et al., 2003; Cuenca Grau et al., 2008), which is based on the description logic *SR_QIQ* (Horrocks et al., 2006), as the ontology language. This logic-based representation enables us to formally describe the qualitative features of our images. Our system also includes a DL reasoner, enabling objects from the images and the images themselves to be categorized according to the definitions incorporated into the ontology schema, which enhances the qualitative description of the images with new inferred knowledge.

Description logics are fragments of first order logic, therefore they work under the open world assumption (OWA) (Hustadt, 1994), that is, unlike databases, they work under the assumption that knowledge of the world is incomplete. In this paper, the suitability of the OWA for our domain is analyzed and the cases where

¹More details in: <http://people.cs.uchicago.edu/~pff/segment/>

²Ontology Web Language: <http://www.w3.org/TR/owl2-syntax/>

additional reasoning services or the closed world assumption (CWA) would be necessary are detected. Moreover, a partial solution for our setting is proposed.

The remainder of this paper is organized as follows. Section 2 describes the related work. Section 3 summarizes our approach for qualitative description of images. Section 4 presents the proposed ontology-based representation of qualitative description of images. Section 5 shows the tests carried out by our approach in a scenario where a robot navigates and then the results obtained are analysed. Finally, Section 6 explains our conclusions and future work.

2 Related Work

Related studies have been published that extract qualitative or semantic information from images representing scenes (Socher et al., 1997; Lovett et al., 2006; Qayyum and Cohn, 2007; Oliva and Torralba, 2001; Quattoni and Torralba, 2009).

Socher et al. (1997) provide a verbal description of an image to a robotic manipulator system so it can identify and pick up an object that has been previously modelled geometrically and then categorized qualitatively by its type, colour, size and shape. The spatial relations between the predefined objects detected in the image are also described qualitatively. Lovett et al. (2006) propose a qualitative description for sketch image recognition, which describes lines, arcs and ellipses as basic elements and also the relative position, length and orientation of their edges. Qayyum and Cohn (2007) divide landscape images using a grid for their description so that semantic categories (grass, water, etc.) can be identified and qualitative relations of relative size, time and topology can be used for image description and retrieval in databases. Oliva and Torralba (2001) obtain the *spatial envelope* of complex environmental scenes by analysing the discrete Fourier transform of each image and extracting perceptual properties of the images (naturalness, openness, roughness, ruggedness and expansion) which enable classification of images in the following semantic categories: coast, countryside, forest, mountain, highway, street, close-up and tall building. Quattoni and Torralba (2009) propose an approach for classifying images of indoor scenes in semantic categories such as bookstore, clothing store, kitchen, bathroom, restaurant, office, classroom, etc. This approach combines global spatial properties and local discriminative information (i.e. information about objects contained in the places) and uses learning distance functions for visual recognition.

We believe that all the studies described above provide evidence for the effectiveness of using qualitative/semantic information to describe images. However, in the approach developed by Socher et al. (1997), a previous object recognition process is needed before qualitatively describing the image of the scene the robot manipulator has to manage, whereas our approach is able to describe the image of the scene in front of the robot without this prior recognition process because the object characterization is done afterwards using the definitions of our ontology.

The approach of Lovett et al. (2006) is applied to sketches, while our approach is applied to digital images captured from the real robot environment. Qayyum and Cohn (2007) use a grid to divide the image and describe what is inside each grid square (grass, water, etc.), which is adequate for their application but the objects are divided into an artificial number of parts that depend on the size of the cell, while our approach extracts complete objects, which could be considered more cognitive. The approach of Oliva and Torralba (2001) is useful for distinguishing between outdoor environments. However, as this approach does not take into account local object information, it will obtain similar *spatial envelopes* for similar images corresponding to the indoor environments where our robot navigates, such as corridors in buildings. The approach of Quattoni and Torralba (2009) performs well in recognizing indoor scenes, however it uses a learning distance function and, therefore, it must be trained on a dataset, while our approach does not require training.

There are related studies in the literature that examine the possible benefits and challenges of using description logics (DL) as knowledge representation and reasoning systems for high-level scene interpretation (Neumann and Möller, 2008; Dasiopoulou and Kompatsiaris, 2010). Neumann and Möller (2008) also present the limitations of current DL reasoning services in a complete scene interpretation. In addition, they give some useful guidelines for future extensions of current DL systems. Nevertheless, the use of DLs in image interpretation is still presented as an open issue (Dasiopoulou and Kompatsiaris, 2010) because of their inherent open world semantics.

Only a few approaches, using DL-based ontologies to enhance high-level image interpretation, can be found in the literature (Maillot and Thonnat, 2008; Johnston et al., 2008; Schill et al., 2009; Bohlken and Neumann, 2009). Maillot and Thonnat (2008) describe images using an ontology that contains qualitative features of shape, colour, texture, size and topology and apply this description to the classification of pollen grains. In the work by Maillot and Thonnat (2008), the regions to describe inside an image are segmented manually using intelligent scissors within the knowledge acquisition tool, while in our approach they are extracted automatically. For the ontology-backend, Maillot and Thonnat (2008) perform, as in our approach, a good differentiation of three levels of knowledge; however, they do not tackle the open world problem of image interpretation. Johnston et al. (2008) present an ontology-based approach to categorize objects and communicate among agents. This approach was innovatively tested at the RoboCup tournament where it was used to enable Sony AIBO robots to recognize the ball and the goal. Similarly to our approach, the authors adopt description logics to represent the domain entities and they use a reasoner to infer new knowledge. In contrast to our approach, the lighting conditions are controlled in the RoboCup tournament and the shape and colour of the objects to search for (ball and goal) are known a priori and are easy to locate using colour segmentation techniques. Moreover, this

work does not address the problems related to the open world assumption. Schill et al. (2009) describe an interesting scene interpretation approach that combines a belief theory with an OWL-like ontology based on DOLCE (Gangemi et al., 2002). Identified objects are classified into the ontology concepts with a degree of belief or uncertainty. This approach could be considered as complementary to ours, and future extensions may consider the introduction of uncertainty. Bohlken and Neumann (2009) present a novel approach in which a DL ontology is combined with the use of rules to improve the definition of constraints for scene interpretation. The use of rules enables them to combine the open world semantics of DLs with closed world constraint validation. However, the use of rules may lead to undecidability and so their use should be restricted (Motik et al., 2005; Krötzsch et al., 2008). Our approach implements a simpler solution, although it would be interesting to analyze extensions involving the use of rules. Finally, it should be noted that our DL-ontology is not designed for a particular type of robot or scenario. It is based on a general approach for describing any kind of image detected by a digital camera.

Other interesting approaches are those that relate qualitative spatial calculus with ontologies (Bhatt and Dylla, 2009; Katz and Cuenca Grau, 2005). Bhatt and Dylla (2009) modelled spatial scenes using an ontology that represents the topological calculus $\mathcal{RCC-8}$ and the relative orientation calculus \mathcal{OPRA} . In contrast to our approach, they do not address the problem of extracting and describing objects contained in digital images and their ontology is not based on DL. Katz and Cuenca Grau (2005) exploit the correspondences among DL, modal logics and the Region Connection Calculus $\mathcal{RCC-8}$ in order to propose a translation of the $\mathcal{RCC-8}$ into DL.

Despite all the previous studies combining the extraction of qualitative/semantic information and its representation using ontologies, the problem of bringing together low-level sensory input and high-level symbolic representations is still a big challenge in robotics. Our approach is a small contribution to meeting this challenge.

3 Our Approach for Qualitative Description of Images

The approach presented in this paper describes any image qualitatively by describing the visual and spatial features of the regions or objects within it.

For each region or object in the image, the visual features (Section 3.1) and the spatial features (Section 3.2) are described qualitatively. As visual features, our approach describes the shape and colour of the region, which are absolute properties that only depend on the region itself. As spatial features, our approach describes the topology and orientation of the regions, which are properties defined with respect to other regions (i.e. containers and neighbours of the regions).

A diagram of the qualitative description obtained by our approach is shown in Figure 1.

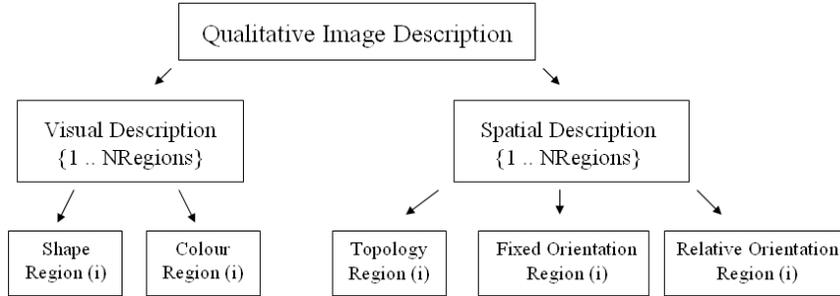


Figure 1: Structure of the qualitative image description obtained by our approach.

3.1 Describing Visual Features of the Regions in the Image

In order to describe the visual features of the regions or objects in an image, we use the qualitative model of shape description formally defined in (Falomir et al., 2008), which is summarized in Section 3.1.1, and a new model for qualitative colour naming based on the Hue, Saturation and Lightness (HSL) colour space, which is defined in Section 3.1.2.

3.1.1 Qualitative Shape Description

Given a digital image containing an object, our approach for Qualitative Shape Description (QSD) first automatically extracts the boundary of this object using colour segmentation (Felzenszwalb and Huttenlocher, 2004).

From all the points that define the boundary of each of the closed objects extracted, a set of consecutive points are compared by calculating the slope between them. If the slope between a point P_i and its consecutive point P_{i+1} , denoted by s_1 , and the slope between P_i and P_{i+2} , termed s_2 , are equal, then P_i , P_{i+1} and P_{i+2} belong to the same straight segment. If s_1 and s_2 are not equal, P_i , P_{i+1} and P_{i+2} belong to a curved segment. This process is repeated for a new point P_{i+3} , calculating the slope between P_i and P_{i+3} (s_3), and comparing that slope with s_1 and so on. The process stops when all the consecutive points of the boundary are visited.

P is considered a relevant point if: (1) it belongs to a straight segment and it is the point at which the slope stops being constant; or (2) it belongs to a curved segment and it is the point at which the slope changes its sign.

Note that the points of a boundary that are considered consecutive are those separated by a pre-established *granularity step*. For example, if the *granularity step* is k , the first point considered (P_i) will be point 1 of the set of boundary points ($PSet(1)$), P_{i+1} will be point k of the set of boundary points $PSet(k)$, P_{i+2} will be $PSet(2k)$, P_{i+3} will be $PSet(3k)$, etc. This *granularity step* is set by experimentation as a function of the edge length of the described object: if the edges are long, the *granularity step* will have a larger value; if they are short, the *granularity step* will have a smaller value.

Finally, a set of relevant points, denoted by $\{P_0, P_1, \dots, P_N\}$, determines the shape of the object. Each of those relevant points P is described by a set of four features $\langle KEC_P, A_P \text{ or } TC_P, L_P, C_P \rangle$, which are defined below.

The first feature is the Kind of Edges Connected (denoted by **KEC**) and it indicates the connection occurring at the relevant point P . This feature is described by the following tags:

- *line-line*, if the point P connects two straight lines;
- *line-curve*, if P connects a line and a curve;
- *curve-line*, if P connects a curve and a line;
- *curve-curve*, if P connects two curves; or
- *curvature-point*, if P is a point of curvature of a curve.

If **KEC** is a *line-line*, *line-curve*, *curve-line* or *curve-curve*, the second feature to consider is the Angle (denoted by **A**) at the relevant point. The angle is a quantitative feature that is discretized by using the Angle Reference System or ARS = $\{\circ, A_{LAB}, A_{INT}\}$ where, degrees (\circ) indicates the unit of measurement of the angles; A_{LAB} refers to the set of labels for the angles; and A_{INT} refers to the values of degrees (\circ) related to each label. In our approach the A_{LAB} and A_{INT} used are:

$$A_{LAB} = \{very_acute, acute, right, obtuse, very_obtuse\}$$

$$A_{INT} = \{(0, 40], (40, 85], (85, 95], (95, 140], (140, 180]\}$$

On the other hand, if **KEC** is a *curvature-point*, the second feature is the Type of Curvature (denoted by **TC**) at P which is defined by the Type of Curvature Reference System or TCRS = $\{\circ, TC_{LAB}, TC_{INT}\}$, where \circ refers to the amplitude in degrees of the angle given by the relation between the distances da and db (see Figure 2(a) where the type of curvature of the relevant point P_j is shown with respect to the relevant points P_{j-1} and P_{j+1}), that is, $Angle(P_j) = 2 \arctg(da/db)$, TC_{LAB} refers to the set of labels for curvature; and TC_{INT} refers to the values of degrees (\circ) related to each label. In our approach the TC_{LAB} and TC_{INT} are:

$$TC_{LAB} = \{very_acute, acute, semicircular, plane, very_plane\}$$

$$TC_{INT} = \{(0, 40], (40, 85], (85, 95], (95, 140], (140, 180]\}$$

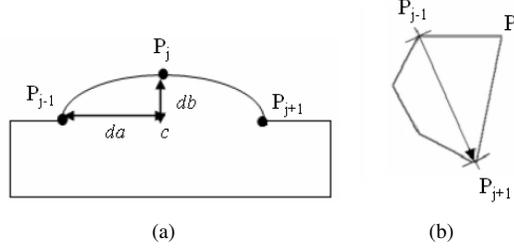


Figure 2: Characterization of P_j as: (a) a point of curvature, and (b) a point connecting two straight segments.

The third feature considered is the compared length (denoted by \mathbf{L}) which is defined by the Length Reference System or $\text{LRS} = \{\text{UL}, \text{L}_{LAB}, \text{L}_{INT}\}$, where UL or Unit of compared Length refers to the relation between the length of the first edge and the length of the second edge connected by P , that is, $ul = (\text{length of } 1^{\text{st}} \text{ edge})/(\text{length of } 2^{\text{nd}} \text{ edge})$; L_{LAB} refers to the set of labels for compared length; and L_{INT} refers to the values of UL related to each label.

$$\text{L}_{LAB} = \{\text{much_shorter (msh)}, \text{half_length (hl)}, \text{a_bit_shorter (absh)}, \text{similar_length (sl)}, \text{a_bit_longer (abl)}, \text{double_length (dl)}, \text{much_longer (ml)}\}$$

$$\text{L}_{INT} = \{(0, 0.4], (0.4, 0.6], (0.6, 0.9], (0.9, 1.1], (1.1, 1.9], (1.9, 2.1], (2.1, \infty)\}$$

The last feature to be considered is the Convexity (denoted by \mathbf{C}) at point P , which is obtained from the oriented line built from the previous point to the next point and by ordering the qualitative description of the shape clockwise. For example from Figure 2(b), if point P_j is on the left of the segment defined by P_{j-1} and P_{j+1} , then P_j is *convex*; otherwise P_j is *concave*.

Thus, the complete shape of an object is described as a set of qualitative descriptions of relevant points as ³:

$$[[\text{KEC}_0, \text{A}_0 \mid \text{TC}_0, \text{L}_0, \text{C}_0], \dots, [\text{KEC}_{n-1}, \text{A}_{n-1} \mid \text{TC}_{n-1}, \text{L}_{n-1}, \text{C}_{n-1}]]$$

where \mathbf{n} is the total number of relevant points of the object.

Finally, note that the intervals of values that define the qualitative tags representing the features *angle*, *type of curvature* and *compared length* (A_{INT} , TC_{INT} and L_{INT} , respectively) have been calibrated according to our application and system.

³ $\text{A}_i \mid \text{TC}_i$ denotes that the angle or the type of curvature that occurs at the point P_i .

3.1.2 Qualitative Colour Description

Our approach for Qualitative Colour Description (QCD) translates the Red, Green and Blue (RGB) colour channels into Hue, Saturation and Lightness (HSL) coordinates, which are more suitable for dividing into intervals of values corresponding to colour names.

In contrast to the RGB model, HSL is considered a more natural colour representation model as it is broken down according to physiological criteria: hue refers to the pure spectrum colours and corresponds to dominant colour as perceived by a human; saturation corresponds to the relative purity or the quantity of white light that is mixed with hue; and luminance refers to the amount of light in a colour (Sarifuddin and Missaoui, 2005). Furthermore, as W3C mentions⁴, additional advantages of HSL are that it is symmetrical to lightness and darkness (which is not the case with HSV, for example). This means that: (i) in HSL, the saturation component takes values from fully saturated colour to the equivalent grey, but in HSV, considering the value component at the maximum, it goes from saturated colour to white, which is not intuitive; and (ii) the lightness in HSL always spans the entire range from black through the chosen hue to white, while in HSV, the value component only goes halfway, from black to the chosen hue.

From the HSL colour coordinates obtained, a reference system for qualitative colour description is defined as: $QCRS = \{UH, US, UL, QC_{LAB1..5}, QC_{INT1..5}\}$ where UH is the Unit of Hue; US is the Unit of Saturation; UL is the Unit of Lightness; $QC_{LAB1..5}$ refers to the qualitative labels related to colour; and $QC_{INT1..5}$ refers to the intervals of HSL colour coordinates associated with each colour label.

For our approach, the QC_{LAB} and QC_{INT} are the following:

$$QCRS_{LAB1} = \{black, dark_grey, grey, light_grey, white\}$$

$$QCRS_{INT1} = \{[0 ul, 20 ul[, [20 ul, 30 ul[, [30 ul, 40 ul[, [40 ul, 80 ul[, [80 ul, 100 ul[/ \forall UH \wedge US \in [0, 20]\}$$

$$QCRS_{LAB2} = \{red, yellow, green, turquoise, blue, purple, pink\}$$

$$QCRS_{INT2} = \{[335 uh, 360 uh] \wedge [0 uh, 40 uh], [40 uh, 80 uh], [80 uh, 160 uh], [160 uh, 200 uh], [200 uh, 260 uh], [260 uh, 297 uh] [297 uh, 335 uh] / \forall UL \in [40, 55] \wedge US \in [50, 100]\}$$

$$QCRS_{LAB3} = \{pale_ + QCRS_{LAB2}\}$$

$$QCRS_{INT3} = \{ \forall UH \wedge US \in [20, 50] \wedge UL \in [40, 55] \}$$

$$QCRS_{LAB4} = \{light_ + QCRS_{LAB2}\}$$

$$QCRS_{INT4} = \{ \forall UH \wedge US \in [50, 100] \wedge UL \in [55, 100] \}$$

$$QCRS_{LAB5} = \{dark_ + QCRS_{LAB2}\}$$

⁴See the CSS3 specification from the W3C (<http://www.w3.org/TR/css3-color/#hsl-color>)

$$\text{QCRS}_{INT5} = \{\forall \text{UH} \wedge \text{US} \in]50, 100] \wedge \text{UL} \in]20, 40]\}$$

The saturation coordinate of the HSL colour space (US) determines if the colour corresponds to the grey scale or to the rainbow scale: QCRS_{LAB1} and QCRS_{LAB2} , respectively, in our QCRS. This coordinate also determines the intensity of the colour (pale or strong). The colours in the rainbow scale are considered the strong ones, whereas the pale colours are given an explicit name in QCRS_{LAB3} . The hue coordinate of the HSL colour space (UH) determines the division into colour names inside each scale. Finally, the lightness coordinate (UL) determines the luminosity of the colour: dark and light colours are given an explicit name in QCRS_{LAB4} and QCRS_{LAB5} , respectively.

Note that the intervals of HSL values that define the colour tags ($\text{QC}_{INT1..5}$) have been calibrated according to our application and system.

Colour identification depends on illumination, but HSL colour space deals with lighting conditions through the L coordinate, which separates the lightness of the colour while its corresponding hue or colour spectrum remains the same.

Finally, our approach obtains the qualitative colour of the centroid of each object detected in the image and the qualitative colour of the relevant points of its shape and the most frequent colour is defined as the colour of the object. Note that colour patterns are not handled at all.

3.2 Describing Spatial Features of the Regions in the Image

The spatial features considered for any region or object in the image are its topology relations (Subsection 3.2.1) and its fixed and relative orientation (Subsection 3.2.2). Orientation and topology relations describe the situation of the objects in the two-dimensional space regardless of the proximity of the observer (robot/person) to them. Moreover, topology relations also implicitly describe the relative distance between the objects.

3.2.1 Topological Description

In order to represent the topological relationships of the objects in the image, the intersection model defined by Egenhofer and Franzosa (1991) for region configurations in R^2 is used.

However, as information on depth cannot be obtained from digital images, the topological relations *overlap*, *coveredBy*, *covers* and *equal* defined in Egenhofer and Franzosa (1991) cannot be distinguished by our approach and are all substituted by *touching*.

Therefore, the topology situation in space (invariant under translation, rotation and scaling) of an object A with respect to (wrt) another object B (A wrt B), is described by:

$$\text{Topology} = \{disjoint, touching, completely_inside, container\}$$

Our approach determines if an object is *completely_inside* or if it is the *container* of another object. It also defines the *neighbours* of an object as all the other objects with the same container. The neighbours of an object can be (i) *disjoint* from the object, if they do not have any edge or vertex in common; (ii) or *touching* the object, if they have at least one vertex or edge in common or if the Euclidean distance between them is smaller than a certain threshold set by experimentation.

3.2.2 Fixed and Relative Orientation Description

Our approach describes the orientation of the objects in the image using the model defined by Hernández (1991) and the model defined by Freksa (1992).

A Fixed Orientation Reference System (FORS) is defined by using the model by Hernández (1991), which obtains the orientation of an object A with respect to (wrt) its container or the orientation of an object A wrt an object B, neighbour of A. This reference system divides the space into eight regions (Figure 3(a)) which are labelled as:

$$\text{FORS}_{labels} = \{front (f), back (b), left (l), right (r), left_front (lf), right_front (rf), left_back (lb), right_back (rb), centre (c)\}$$

In order to obtain the fixed orientation of each object wrt another or wrt the image, our approach locates the centre of the FORS on the centroid of the reference object and its *front* area is fixed to the upper edge of the image. The orientation of an object is determined by the union of all the orientation labels obtained for each of the relevant points of the object. If an object is located in all the regions of the reference system, it is considered to be in the *centre*.

A Relative Orientation Reference System (RORS) is defined by using Freksa (1992) double cross orientation model. This model divides the space by means of a Reference System (RS) which is formed by an oriented line determined by two reference points *a* and *b*. The information that can be represented by this model is the qualitative orientation of a point *c* wrt the RS formed by the points *a* and *b*, that is, *c wrt ab* (Figure 3(b)). This model divides the space into 15 regions which are labelled as:

$$\text{RORS}_{labels} = \{left_front (lf), straight_front (sf), right_front (rf), left (l), identical_front (idf), right (r), left_middle (lm), same_middle (sm), right_middle (rm), identical_back_left (ibl), identical_back (ib), identical_back_right (ibr), back_front (bf), same_back (sb), back_right (br)\}$$

In order to obtain the relative orientation of an object, our approach establishes reference systems (RORSs) between all the pairs of disjoint neighbours of that object. The points *a* and *b* of the RORS are the centroids of the objects that make

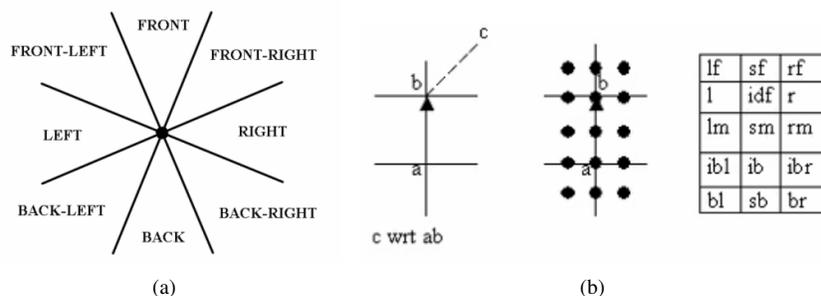


Figure 3: Models of Orientation used by our Approach for Image Description. (a) Hernandez's orientation model; (b) Freksa's orientation model and its iconical representation: 'l' is left, 'r' is right, 'f' is front, 's' is straight, 'm' is middle, 'b' is back and 'i' is identical.

up the RORS. The relevant points of each object are located with respect to the corresponding RORS and the orientation of an object with respect to a RORS is calculated as the union of all the orientation labels obtained for all the relevant points of the object.

Note that, as we have already mentioned, in our approach, orientation relations between the objects in the image are structured in levels of containment. The fixed orientation (Hernández, 1991) of a region is defined with respect to its *container* and *neighbours of level*, while the relative orientation of a region (Freksa, 1992) is defined with respect to its *disjoint neighbours of level*. Therefore, as the spatial features of the regions are relative to the other regions in the image, the number of spatial relationships that can be described depends on the number of regions located at the same level of containment, as shown in Table 1. The advantage of providing a description structured in levels of containment is that the level of detail to be extracted from an image can be selected. For example, the system can extract all the information in the image or only the information about the objects whose container is the image and not another object, which could be considered a more general or abstract description of the image.

The reason for using two models for describing the orientation of the objects or regions in the image is the different kind of information each provides. According to the classification of reference frames by Hernández (1991), we can consider that:

- the reference system or frame in the model developed by Hernández (1991) is intrinsic because the orientation is given by some inherent property of the reference object. This property is defined by our approach by fixing the object front to the upper edge of the image. Therefore, the orientations provided by this model are implicit because they refer to the intrinsic orien-

Table 1: Spatial features described depending on the number of objects at each level

| <i>Spatial Features Described</i> | | <i>Objects within the same container</i> | | |
|-----------------------------------|----------------------|--|----------|---------------|
| | | 1 | 2 | > 2 |
| Wrt its Container | Topology | x | x | x |
| | Fixed Orientation | x | x | x |
| Wrt its Neighbours | Topology | - | x | x |
| | Fixed Orientation | - | x | x |
| | Relative Orientation | - | - | x |

tation of the parent object or the object of reference. Here, implicit and intrinsic orientations coincide as the front of all the objects is fixed to the same location a priori. Therefore, the point of view is influenced by the orientation of the image given by an external observer.

- in the model developed by Freksa (1992), an explicit reference system or frame is necessary to establish the orientation of the point of view with respect to the reference objects. Moreover, this reference system is extrinsic, since an oriented line imposes an orientation and direction on the reference objects. However, the orientation between the objects involved is invariant to the orientation of the image given by an external observer, because even if the image rotates, the orientations obtained by our RORS remain the same.

Therefore, in practice, considering both models, our approach can: (i) describe the implicit orientations of the objects in the image from the point of view of an external observer (robot camera) and regardless of the number of objects within the image, and (ii) describe complex objects contained in the image (which must be composed of at least three objects or regions) in an invariant way, that is, regardless of the orientation of the image given by an external observer (which could be very useful in a vision recognition process in the near future).

3.3 Obtaining a Qualitative Description of Any Digital Image

The approach presented here qualitatively describes any image by describing the visual and spatial features of the main regions of interest within it.

In an image, different colours or textures usually indicate different regions of interest to the human eye. Therefore, image region segmentation approaches are more cognitive than edge-based image segmentation approaches because the extracted edges are defined by the boundaries between regions and all of them are closed (Palmer, 1999).

In our approach, the regions of interest in an image are extracted by a graph-based region segmentation method (Felzenszwalb and Huttenlocher, 2004) based on intensity differences complemented by algorithms developed to extract the boundaries of the segmented regions. Felzenszwalb and Huttenlocher (2004) mention that the problems of image segmentation and grouping remain great challenges for computer vision because, a useful segmentation method has to: (i) capture perceptually important groupings or regions, which often reflect global aspects of the image; and (ii) be highly efficient, running in nearly linear time in the full number of image pixels. The approach developed by Felzenszwalb and Huttenlocher (2004) is suitable for our approach because it meets the above criteria and it also preserves detail in low-variability image regions while ignoring detail in high-variability regions with adjustment of its segmentation parameters: σ , used to smooth the input image before segmenting it; k , the value for the threshold function in segmentation, the larger the value, the larger the components in the result; and min , minimum size of the extracted regions in pixels enforced by post-processing.

Once all the relevant colour regions of an image are extracted, image processing algorithms that obtain the contour of each region are applied. Then, our approach describes the shape, colour, topology and orientation of each region as explained above. As an example, a digital image of the corridors of a building of our university where our offices are located is presented in Figure 4 and its qualitative description is presented in Table 2.

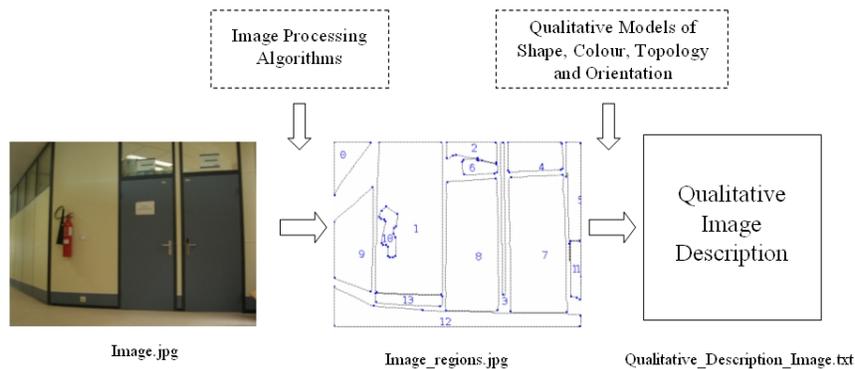


Figure 4: Schema of our approach for qualitative image description.

Figure 4 shows the original image and the image obtained after extracting the regions obtained by Felzenszwalb and Huttenlocher’s (2004) segmentation method (using as segmentation parameters: $\sigma = 0.4$, $k = 500$ and $min = 1000$) and the contour of these regions by our boundary extraction algorithms. Note that

some objects in the original image (such as the door sign, the door handles and the electrical socket) are not extracted in the segmentation process because of their small size (< 1000 pixels).

Table 2: An excerpt of the qualitative description obtained for the image in Figure 4.

| |
|--|
| <p>[SpatialDescription,</p> <p>[1, [Container, Image], [Orientation wrt Image: front, front_left, back_left, back], touching, 2, 8, 9, 13], [disjoint, 0, 3, 4, 5, 6, 7, 11, 12], [completely_inside, 10], [Orientation wrt Neighbours: [0, front_right, right, back_right, back], [2, left, back, back_left], (...) [Relative Orientation wrt Neighbours Disjoint: [[0, 4], rm], (...) [[4, 7], br, rf], (...) [[11, 12], rm, rf]]](...)]</p> <p>[10, [[Container, 1] [Orientation wrt 1: left, back_left, back], [None Neighbours of Level]]](...)]</p> <p>[VisualDescription,</p> <p>[7, dark-grey, Boundary_Shape, [line-line, right, much_shorter, convex] [line-line, right, much_longer, convex], [line-line, right, half_length, convex], [line-line, right, much_longer, convex], [Vertices.Orientation, front, back, back, front]],], (...)]</p> <p>[10, dark-red, Boundary_Shape, [line-line, obtuse, half_length, convex], (...) [line-line, very_obtuse, similar_length, convex]] [Vertices.Orientation, front, front, front_right, right, back_right (...)]],], (...)]</p> |
|--|

Table 2 presents an excerpt of the qualitative description obtained from the digital image presented in Figure 4. Specifically, this table shows the qualitative spatial description of regions 1 and 10 and the qualitative visual description of regions 7 and 10. Note that the object identifiers correspond to those located on the objects in *Image_regions.jpg* in Figure 4.

The spatial description of region 1 can be intuitively read as follows: its *container* is the Image and it is located wrt to the Image at *front*, *front_left*, *back*, *back_left*. Its *touching neighbours* are the regions 2, 8, 9, 13 (Note that some of these are not technically touching but are closer to region 1 than the threshold set

by experimentation for our application). Its *disjoint neighbours* are the regions 0, 3, 4, 5, 6, 7, 11 and 12 and finally, object 10 is *completely_inside* 1. The *fixed orientation* of region 1 wrt region 0 is *front_right*, *right*, *back_right*, *back*, wrt region 2 it is *left*, *back*, *back_left*, wrt region 3 it is *back_right* and in a similar way, the fixed orientation of region 1 is described wrt all its *neighbours of level*. Finally, the relative orientation wrt the disjoint neighbours of region 1 is given: from region 0 to region 4, region 1 is located *right_middle (rm)*; from region 4 to region 7, region 1 is located *back right (br)* and also *right front (rf)*, from region 11 to region 12, region 1 is located *right middle (rm)* and *right front (rf)*.

The spatial description of region 10 is also given in Table 2: its container is region 1 with respect to which it is located at *left*, *back left*, *back*. Region 10 has no *neighbours of level*, as it is the only region contained by region 1.

The visual description of region 7 in Table 2 shows that its colour is *dark_grey* and that the shape of its boundary is qualitatively described as composed of four *line_line* segments whose angles are all *right* and *convex* and whose compared distances are respectively *much_shorter*, *much_longer*, *half*, *much_longer*. Finally, the orientation of its vertices with respect to the centroid of the region is in a clockwise direction: *front*, *back*, *back*, *front*. Note that region 10 is described similarly.

4 Formal Representation of Qualitative Descriptions

Our approach describes any image using qualitative information, which is both visual (e.g. shape, colour) and spatial (e.g. topology, orientation). Here the use of ontologies is proposed in order to give a formal meaning to the qualitative labels associated with each object. Thus, ontologies will provide a logic-based representation of the knowledge within the robot system.

An ontology is a *formal specification of a shared conceptualization* (Borst et al., 1997) providing a non-ambiguous and formal representation of a domain. Ontologies usually have specific purposes and are intended for use by computer applications rather than humans. Therefore, ontologies should provide a common vocabulary and meaning to allow these applications to communicate with each other (Guarino, 1998).

The aim of using a description logics (DL) based ontology is to enhance image interpretation and classification. Furthermore, the use of a common vocabulary and semantics is also intended to facilitate potential communication between agents. The main motives for using DL-based ontologies within our system are:

- **Symbol Grounding.** The association of the right qualitative concept with quantitative data (a.k.a. symbol grounding) and the precise relationships between qualitative concepts is still an open research line (Williams, 2008; Williams et al., 2009). The description logic family was originally called

terminological or concept language due to its concept-centred nature. Thus, DL-based ontologies represent a perfect formalism for providing high-level representations of low-level data (e.g. digital image analysis).

- **Knowledge sharing.** The use of a common conceptualization (vocabulary and semantics) may enhance communication between agents involved in performing similar tasks (e.g. searching for a fire extinguisher in a university environment). Moreover, the adoption of a standard ontology language gives our approach a mechanism for publishing our qualitative representations of images so that they can be reused by other agents.
- **Reasoning.** The adoption of a DL-based representation allows our approach to use DL reasoners that can infer new knowledge from explicit descriptions. This gives some freedom and flexibility when inserting new facts (e.g. new image descriptions), because new knowledge can be automatically classified (e.g. a captured object is a door, a captured image contains a fire extinguisher).

In this section, we present `QImageOntology`⁵, a DL-based ontology to represent qualitative description of images (see Section 4.1), and how we have dealt with the Open World Assumption (OWA) (Hustadt, 1994) in order to infer the expected knowledge (see Section 4.2).

4.1 Three-Layer Representation for `QImageOntology`

`QImageOntology` has adopted DL and OWL as the formalisms for representing the qualitative descriptions extracted from the images. `QImageOntology` was developed using the ontology editor Protégé⁶. Additionally the DL reasoner HermiT⁷ was used for classifying new captured images and for inferring new knowledge.

DL systems make an explicit distinction between the terminological or intensional knowledge (a.k.a. Terminological Box or TBox), which refers to the general knowledge about the domain, and the assertional or extensional knowledge (a.k.a. Assertional Box or ABox), which represents facts about specific individuals. `QImageOntology` also makes a distinction between the general object descriptions and the facts extracted from concrete images. Additionally, our approach includes a knowledge layer within the TBox dealing with contextualized object descriptions (e.g. a UJI office door).

This three-layer architecture is consistent with our purposes and image descriptions are classified with the TBox part of `QImageOntology`. Moreover, the contextualized knowledge can be replaced to suit a particular scenario or environment

⁵Available at: <http://krono.act.uji.es/people/Ernesto/qimage-ontology/>

⁶Protégé: <http://protege.stanford.edu>

⁷HermiT: <http://hermit-reasoner.com/>

(e.g. Jaume I University, Valencia City Council). Thus, the three-layer architecture is composed of:

- a *reference conceptualization*, which is intended to represent knowledge (e.g. the description of a *Triangle* or the assertion of *red* as a *Colour_type*) that is supposed to be valid in any application. This layer is also known as *top level knowledge*⁸ by the community;
- the *contextualized knowledge*, which is application oriented and is mainly focused on the specific representation of the domain (e.g. characterization of doors at Jaume I University) and could be in conflict with other context-based representations; and
- the *image facts*, which represent the assertions or individuals extracted from the image analysis, that is, the set of particular qualitative descriptions.

It is worth noting that the knowledge layers of *QImageOntology* are considered to be three different modules and they are stored in different OWL files. Nevertheless, both the *contextualized knowledge* and the *image facts* layers are dependent on the *reference conceptualization* layer, and thus they perform an explicit import of this reference knowledge.

Currently, the *reference conceptualization* and *contextualized knowledge* layers of *QImageOntology* have a *SHOIQ* DL expressiveness and contain: 51 concepts (organized into 80 subclass axioms, 14 equivalent axioms and 1 disjointness), 46 object properties (characterized with 30 subproperty axioms, 5 property domain axioms, 10 property range axioms, 19 inverse property axioms, and 2 transitive properties), and 51 general individuals (with 51 class assertion axioms and 1 different individual axiom).

An excerpt of the *reference conceptualization* of *QImageOntology* is presented in Table 3: partial characterizations of an *Object_type*, the definition of a *Shape_type* as a set of at least 3 relevant points, the definition of a *Quadrilateral* as a *Shape_type* with exactly 4 points connecting two lines and so on.

Table 4 represents an excerpt from the *contextualized knowledge* of *QImageOntology*, where four objects are characterized: (1) the definition of the wall of our corridor (*UJI_Wall*) as a *pale_yellow*, *dark_yellow*, *pale_red* or *light_grey* object contained by the image; (2) the definition of the floor of the corridor (*UJI_Floor*) as a *pale_red* object located inside the image and located at *back_right*, *back_left* or *back* but not at *front*, *front_right* or *front_left* with respect to the centre of the image; (3) the definition of an office door (*UJI_Office_Door*) as a *grey* or *dark_grey* quadrilateral object located inside the image; (4) the definition of

⁸We have created this knowledge layer from scratch. In the near future it would be interesting to integrate our reference conceptualization with standards such as MPEG-7 for which an OWL ontology is already available (Hunter, 2001, 2006), or top-level ontologies such as DOLCE (Gangemi et al., 2002)

Table 3: Excerpt from the Reference Conceptualization of QImageOntology

| | |
|------------|---|
| α_1 | Image_type \sqsubseteq \exists is_container_of.Object_type |
| α_2 | Object_type \sqsubseteq \exists has_colour.Colour_type |
| α_3 | Object_type \sqsubseteq \exists has_fixed_orientation.Object_type |
| α_4 | Object_type \sqsubseteq \exists is_touching.Object_type |
| α_5 | Object_type \sqsubseteq \exists has_shape.Shape_type |
| α_6 | Shape_type \sqsubseteq ≥ 3 has_point.Point_type |
| α_7 | Quadrilateral \sqsubseteq Shape_type \sqcap = 4 has_point.line_line |
| α_8 | is_left \sqsubseteq has_fixed_orientation |
| α_9 | Colour_type : red |

a fire extinguisher (*Fire_Extinguisher*) as a *red* or *dark_red* object located inside a *UJI_Wall*. Note that the contextualized descriptions are rather preliminary and they should be refined in order to avoid ambiguous categorizations.

Table 4: Excerpt from the Contextualized Descriptions of QImageOntology

| | | |
|-----------|--------------------------------|--|
| β_1 | UJI_Wall \equiv | Object_type \sqcap \exists has_shape.Quadrilateral \sqcap \exists is_completely_inside.Image_type \sqcap $(\exists$ has_colour.{pale_yellow} \sqcup \exists has_colour.{dark_yellow} \sqcup \exists has_colour.{pale_red} \sqcup \exists has_colour.{light_grey}) |
| β_2 | UJI_Floor \equiv | Object_type \sqcap \exists is_completely_inside.Image_type \sqcap $(\exists$ has_colour.{pale_red} \sqcup \exists has_colour.{light_grey}) \sqcap \exists is_back.Image \sqcap $\neg(\exists$ is_front.Image) |
| β_3 | UJI_Office_Door \equiv | Object_type \sqcap \exists has_shape.Quadrilateral \sqcap \exists is_completely_inside.Image_type \sqcap $(\exists$ has_colour.{grey} \sqcup \exists has_colour.{dark_grey}) |
| β_4 | UJI_Fire_Extinguisher \equiv | Object_type \sqcap \exists is_completely_inside.UJI_Wall $\sqcap(\exists$ has_colour.{red} \sqcup \exists has_colour.{dark_red}) |

4.2 Dealing with the Open World Assumption (OWA)

Currently one of the main problems that users face when developing ontologies is the confusion between the Open World Assumption (OWA) and the Closed World Assumption (CWA) (Hustadt, 1994; Rector et al., 2004). Closed world systems such as databases or logic programming (e.g. PROLOG) consider anything that cannot be found to be false (negation as failure). However, Description Logics (and therefore OWL) assume an open world, that is, anything is true or false unless the contrary can be proved (e.g. two concepts overlap unless they are declared as disjoint, or a fact not belonging to the knowledge base cannot be considered to be false). However, some scenarios such as image interpretation, where the set of relevant facts are known, may require closed world semantics.

In our scenario, the OWA problem arose when characterizing concepts such as *Quadrilateral* (see axiom α_7 from Table 3), where individuals belonging to this class should be a *Shape_type* and have exactly four sides (i.e. four connected points). Intuitively, one would expect that *object1* from Table 5 should be classified as *Quadrilateral* according to axioms $\gamma_1 - \gamma_7$ and α_7 from Table 3. However, the reasoner cannot make such an inference. The open world semantics have a direct influence in this example since the reasoner is unable to guarantee that *shape1* is not related to more points.

Table 5: Basic image facts for a shape

| | |
|------------|----------------------------|
| γ_1 | Object_type : object1 |
| γ_2 | Shape_type : shape1 |
| γ_3 | has_shape(object1, shape1) |
| γ_4 | has_point(shape1, point1) |
| γ_5 | has_point(shape1, point2) |
| γ_6 | has_point(shape1, point3) |
| γ_7 | has_point(shape1, point4) |

In the literature there are several approaches that have attempted to overcome the OWA problem when dealing with data-centric applications. These approaches (Grimm and Motik, 2005; Motik et al., 2009; Sirin et al., 2008; Tao et al., 2010a,b) have mainly tried to extend the semantics of OWL with non-monotonic features such as *Integrity Constraints (IC)*. Thus, standard OWL axioms are used to obtain new inferred knowledge with open world semantics whereas ICs validate instances using closed world semantics. These approaches have also tried to translate IC validation into query answering using rules (e.g., SWRL, SPARQL) in order to make use of the existing reasoning machinery. Nevertheless, as already discussed

in the literature, the use of rules may lead to undecidability, so the expressivity of the rules must be restricted (Motik et al., 2005; Krötzsch et al., 2008).

Our approach has a partial and much simpler solution that overcomes the OWA limitations for our particular setting. We have restricted the domain of interpretation for each image with the following OWL 2 constructors:

- *Nominals*. We consider that all the relevant facts for an image are known, thus, for each image, `QImageOntology` concepts are *closed* using an extensional definition with nominals⁹. For example, the class `Point_type` is defined as a set of all points recognized within the image (see axiom γ_8 from Table 6 for an image with only 7 points).
- *Negative property assertion axioms* explicitly define that an individual is not related to other individuals through a given property. In our example, the potential *quadrilateral* individual must have four connected points, but there must also be an explicit declaration that it does not have any more associated points (see axioms $\gamma_9 - \gamma_{11}$ from Table 6).
- *Different axioms for each individual*. OWL individuals must be explicitly defined as different with the corresponding axioms, otherwise they may be considered as the same fact, since OWL does not follow the Unique Name Assumption (UNA). In our example points *point1-point4* should be declared as different (see axiom γ_{12}) in order to be interpreted as four different points for the *quadrilateral* individual.

Table 6: Closing the world for a shape

| | |
|---------------|--|
| γ_8 | <code>Point_type</code> \sqsubseteq {point1, point2, point3, point4, point5, point6, point7} |
| γ_9 | \neg has_point(shape1, point5) |
| γ_{10} | \neg has_point(shape1, point6) |
| γ_{11} | \neg has_point(shape1, point7) |
| γ_{12} | point1 \neq point2 \neq point3 \neq point4 |

It is worth mentioning that `QImageOntology` also defines, in its *reference conceptualization layer*, disjoint, range and domain axioms in order to make explicit that two concepts do not overlap and to restrict the use of the properties within the proper concept (e.g. *has_point* only links *Shape_type* with *Point_type*).

In summary, our approach proposes restricting/closing the world for each particular image using the above constructors within the *image facts layer* of `QImageOntology`. The number of *extra* axioms to add is reasonable for our setting

⁹It is well known that the use of nominals makes reasoning more difficult (Tobies, 2001); however, in this case each image contains a relatively small number of individuals

where processed images contain about 200 concrete individuals with 150 class assertions, 1700 object property assertions, 1000 negative property assertions and 5 different individual axioms.

5 Experimentation and Results

As explained in the previous sections, for any digital image, our approach obtains a qualitative image description and a set of facts according to QImageOntology. In Section 5.1, we present how our approach has been implemented and we also describe the results obtained. In Section 5.2, the tests done in different situations within our robot scenario and the evaluation method used are explained. In Section 5.3, the results obtained are analysed.

5.1 Implementation of Our Approach and Description of the Results Obtained

Figure 5 shows the structure of our approach: it obtains the main regions or objects that characterize any digital image, describes them visually and spatially by using qualitative models of shape, colour, topology and orientation and obtains a qualitative description of the image in a flat format (see Table 2) and also as a set of OWL ontology facts.

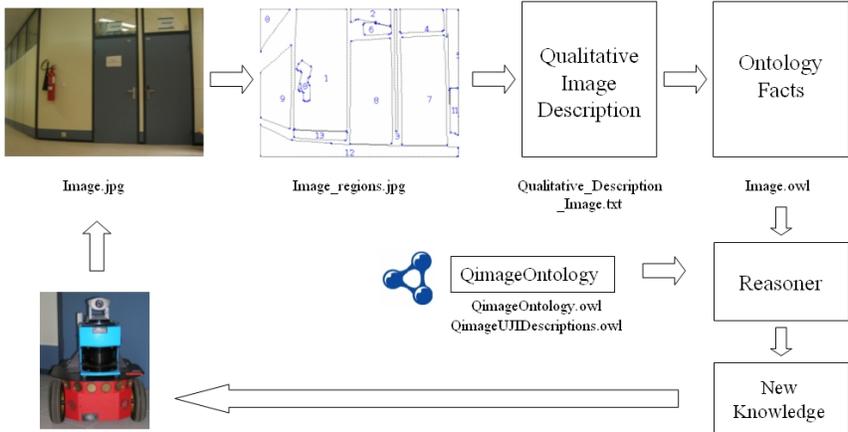


Figure 5: Method overview

The ontology facts obtained (*image facts layer*), together with the *reference conceptualization layer* and the *contextualized knowledge layer* have been automatically classified using the ontology reasoner HermiT, although another rea-

soner (e.g. FaCT++¹⁰ or Pellet¹¹) could have been used. The new inferred knowledge is intended to be reused in the near future by the robot in order to support the decision-making process in localization and navigation tasks.

As an example, note that from the qualitative description of the digital image in Figure 5 (shown in Table 2) and the contextualized descriptions shown in Table 4 the reasoner infers that Object 1 is a *UJI Wall* as it is a *pale yellow* object located *completely inside* the image, that Objects 7 and 8 are *UJI Office doors* as they are *dark grey quadrilaterals* located *completely inside* the image, that Object 10 is a *UJI Fire Extinguisher* as it is a *dark red* object located *completely inside* a *UJI Wall* (Object 1), and finally, that Object 12 is a *UJI Floor* as it is a *pale red* object situated *back left* and *back right* with respect to the centre of the image.

5.2 Testing and Evaluating Our Approach

A collection of digital images extracted from the corridors of our building at Jaume I University (UJI) (our robot scenario) have been processed by our approach and new information has been inferred. Table 7 presents a selection of the tests, where images containing regions classified by our reasoner as *UJI Walls*, *UJI Office Doors*, *UJI Fire Extinguishers* and *UJI Floors* are shown.

Our testing approach and evaluation method is described next. First, our robot explored our scenario with its camera and more than 100 photographs were taken at different locations and points of view with the aim of finding out what kind of objects could be segmented and described by our approach. Walls, floors, office doors, dustbins, fire extinguishers, electrical sockets, glass windows, etc. were properly qualitatively described. The walls, the floor, the office doors and the fire extinguishers were selected as the objects of interest and we adjusted the parameters of the segmentation method to the specific lighting conditions of each test and to define the minimum size of the objects to capture. Second, around 30 photos containing those objects at different locations and points of view were selected and described qualitatively and using description logics. The proper classification of the ontology facts obtained in accordance with *QImageOntology* was checked using Protégé as front-end and a Hermit reasoner. Around 80% of the selected photos (25/30) were correctly classified and some borderline cases appeared because of:

- the adjustment of segmentation parameters: in some cases a *door* region is joined to a *wall* region and extracted as a whole region whose shape is not a *quadrilateral*, and therefore, a *door* cannot be characterized by our approach.

¹⁰FaCT++: <http://owl.man.ac.uk/factplusplus/>

¹¹Pellet: <http://clarkparsia.com/pellet/>

- the colour identification: some extracted regions can be composed of more than a colour, for example the same quantity of *dark-red* and *black* relevant points or pixels can define a *fire-extinguisher* and, therefore, defining the correct colour of the object in those cases is difficult, as our approach does not deal with patterns.

5.3 Analysing Our Results

The results obtained show that our approach can characterize regions of images in our robot scenario as *walls*, *floors*, *office doors* and *fire extinguishers*, under different illumination conditions and from different points of view.

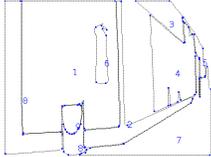
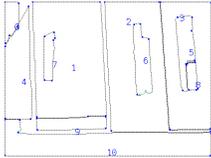
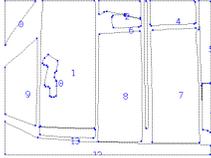
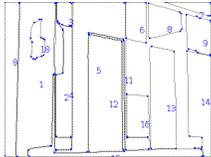
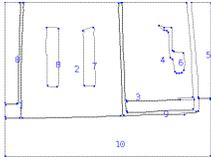
The extraction of the main regions in the image depends on the segmentation parameters used (Felzenszwalb and Huttenlocher, 2004). These parameters are adjusted in order to determine the level of detail extracted from the image. In our tests, regions of small size (such as door signs or handles) are not extracted so as to avoid obtaining much more detail than is needed for our characterization of objects. However, the regions of all tested images that are most easily perceived by the human eye have been obtained and described without problems.

The characterization of qualitative colours using our approach depends on the illumination. This is the main reason that the colour of some objects is defined with different colour names, for example, when identifying doors (*grey* or *dark grey*) or walls (*pale yellow*, *dark yellow* or *light grey*). However, the colour names used in the characterization of an object are very similar from a human point of view and the use of different colour names in an object definition is not a problem for our approach. Therefore, the problems involving different lighting conditions are resolved in this way.

Moreover, it should be noted that our qualitative model for image description provides much more information than is later used in the *contextualized descriptions* of our ontology, which define new kinds of objects based on this information. This is an advantage, as our system could apply this extra information to the characterization of new regions or objects presented in other robot scenarios, where more precise information may be needed in order to differentiate types of regions or objects. For example, our approach has defined *UJI Office Doors* as *dark grey* or *grey quadrilaterals* contained *completely inside* the image. This definition could have been extended by adding that the relevant points of the quadrilateral must be located two at *front* and two at *back* with respect to the centroid of the quadrilateral. Although this information is not needed by the system in order to distinguish *UJI Office Doors* from other objects in our scenario, it could be used in other scenarios in the future.

Finally, as future applications in robotics, we believe that our approach could be usefully applied for *general* and *concrete* robot localization purposes. By extending our approach for characterizing objects to a different scenario (e.g. laborato-

Table 7: Some images of the corridors of our building containing *UJI Fire Extinguishers*, *UJI Walls*, *UJI Office Doors* and *UJI Floor*

| Image | Described Objects | Inferred Information |
|---|---|---|
|  |  | Object 6 is a <i>UJI.Fire.Extinguisher</i> . Objects 1, 4, 5, 7, 8 and 9 are <i>UJI.Walls</i> . |
|  |  | Object 7 is a <i>UJI.Fire.Extinguisher</i> . Object 1 is a <i>UJI.Wall</i> . Object 3 is a <i>UJI.Office.Door</i> . |
|  |  | Object 10 is a <i>UJI.Fire.Extinguisher</i> . Objects 0-6, 9, 11, 12 and 13 are <i>UJI.Walls</i> . Objects 8 and 7 are <i>UJI.Office.Doors</i> . Object 12 is a <i>UJI.Floor</i> . |
|  |  | Object 10 is a <i>UJI.Fire.Extinguisher</i> . Objects 1-4, 6 and 16 are <i>UJI.Walls</i> . Objects 12 and 13 are <i>UJI.Office.Doors</i> . |
|  |  | Objects 6 is a <i>UJI.Fire.Extinguisher</i> . Objects 0-5 and 8 are <i>UJI.Walls</i> . Object 2 is a <i>UJI.Office.Door</i> . Object 9 and 10 are <i>UJI.Floors</i> . |

ries/classrooms/libraries or outdoor areas), it could be used for *general* localization, that is, for determining the kind of scenario the robot is navigating through. Moreover, by defining a matching process for comparing qualitative descriptions of images taken by the robot camera, we could recognize descriptions correspond-

ing to similar or possibly the same visual landmarks and those landmarks could be used to localize the robot specifically in the world.

6 Conclusions

This paper presented a novel approach to represent the qualitative description of images by means of a DL-based ontology. Description logics enable us to balance the need for expressive power with good computational properties for our setting. Our approach obtains a visual and a spatial description of all the characteristic regions/objects contained in an image. In order to obtain this description, qualitative models of shape, colour, topology, and fixed and relative orientation are applied. These qualitative concepts and relations are stored as instances of an ontology and contextualized descriptions that characterize kinds of objects are defined in the ontology schema. Currently, we do not have a contextualized ontology definition for every possible object detected in an image (e.g. printer, office desk or chair). Nevertheless, our approach can automatically process any random image and obtain a set of DL-axioms which describe it visually and spatially.

Our approach has been tested using digital images of the corridors of our building at the university (our robot scenario) and results show that our approach can characterize regions of the image as *walls*, *floor*, *office doors* and *fire extinguishers*, under different illumination conditions and from different observer viewpoints.

As future work on the qualitative description of images, we intend to: (1) extend our model in order to introduce distances captured by the robot laser sensor for obtaining depth information for the images described; and (2) combine our model, which can describe unknown landmarks, with an invariant feature detector, such as SIFT (Lowe, 2004), for detecting known landmarks in the image. Moreover, as further work on our DL-based ontology of images we intend to (1) integrate a reasoner into the robot system, so that the new knowledge obtained can be provided to the robot in real time; (2) reuse non-standard reasoning services such as modularization to improve scalability when dealing with images with a large set of objects; (3) integrate our current ontology with other domain ontologies (e.g., DOLCE (Gangemi et al., 2002)) and standards such as MPEG-7 (Hunter, 2006); (4) extend our current ontology in order to characterize other objects from other robot environments.

Acknowledgments

This work has been partially supported by Generalitat Valenciana (BFPI06/219, BFPI06/372), Spanish MCyT (TIN2008-01825/TIN) and Universitat Jaume I - Fundació Bancaixa (P11A2008-14). We acknowledge Ismael Sanz, Rafael Berlanga and the reviewers of the journal for the valuable feedback and comments they

made us to improve this manuscript. We also thank Pedro F. Felzenszwalb for making his image segmentation source code available in his website.

References

- Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Bhatt, M. and Dylla, F. (2009). A qualitative model of dynamic scene analysis and interpretation in ambient intelligence systems. *Robotics for Ambient Intelligence, International Journal of Robotics and Automation*, 4(3).
- Bohlken, W. and Neumann, B. (2009). Generation of rules from ontologies for high-level scene interpretation. In *International Symposium on Rule Interchange and Applications, RuleML*, volume 5858 of *Lecture Notes in Computer Science*, pages 93–107. Springer.
- Borst, P., Akkermans, H., and Top, J. L. (1997). Engineering ontologies. *Int. J. Hum.-Comput. Stud.*, 46(2):365–406.
- Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., and Sattler, U. (2008). OWL 2: The next step for OWL. *J. Web Semantics*, 6(4):309–322.
- Dasiopoulou, S. and Kompatsiaris, I. (2010). Trends and issues in description logics frameworks for image interpretation. In *Artificial Intelligence: Theories, Models and Applications, 6th Hellenic Conference on AI, SETN*, volume 6040 of *Lecture Notes in Computer Science*, pages 61–70. Springer.
- Egenhofer, M. J. and Al-Taha, K. K. (1992). Reasoning about gradual changes of topological relationships. In Frank, A. U., Campari, I., and Formentini, U., editors, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Intl. Conf. GIS—From Space to Territory*, volume 639 of *Lecture Notes in Computer Science*, pages 196–219, Berlin. Springer.
- Egenhofer, M. J. and Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174.
- Falomir, Z., Almazán, J., Museros, L., and Escrig, M. T. (2008). Describing 2D objects by using qualitative models of color and shape at a fine level of granularity. In *Proc. of the Spatial and Temporal Reasoning Workshop at the 23rd AAAI Conference on Artificial Intelligence, ISBN: 978-1-57735-379-9*.
- Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.

- Freksa, C. (1991). Qualitative spatial reasoning. In Mark, D. M. and Frank, A. U., editors, *Cognitive and Linguistic Aspects of Geographic Space*, NATO Advanced Studies Institute, pages 361–372. Kluwer, Dordrecht.
- Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In Frank, A. U., Campari, I., and Formentini, U., editors, *Theories and methods of spatio-temporal reasoning in geographic space*, volume 639 of *LNCS*, pages 162–178. Springer, Berlin.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening ontologies with dolce. In *13th International Conference on Knowledge Engineering and Knowledge Management, EKAW*, volume 2473 of *Lecture Notes in Computer Science*, pages 166–181. Springer.
- Grimm, S. and Motik, B. (2005). Closed world reasoning in the Semantic Web through epistemic operators. In *Proceedings of the OWLED Workshop on OWL: Experiences and Directions*, volume 188 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Guarino, N. (1998). Formal ontology in information systems. In *International Conference on Formal Ontology in Information Systems (FOIS98)*, Amsterdam, The Netherlands, The Netherlands. IOS Press.
- Hernández, D. (1991). Relative representation of spatial knowledge: The 2-D case. In Mark, D. M. and Frank, A. U., editors, *Cognitive and Linguistic Aspects of Geographic Space*, NATO Advanced Studies Institute, pages 373–385. Kluwer, Dordrecht.
- Horrocks, I., Kutz, O., and Sattler, U. (2006). The even more irresistible *SR_{OL}Q*. In *KR 2006*, pages 57–67.
- Horrocks, I., Patel-Schneider, P. F., and van Harmelen, F. (2003). From *SHIQ* and RDF to OWL: the making of a web ontology language. *J. Web Sem.*, 1(1):7–26.
- Hunter, J. (2001). Adding multimedia to the Semantic Web: Building an MPEG-7 ontology. In *Proceedings of the first Semantic Web Working Symposium, SWWS*, pages 261–283.
- Hunter, J. (2006). Adding multimedia to the Semantic Web: Building and applying an MPEG-7 ontology. In Stamou, G. and Kollias, S., editors, *Chapter 3 of Multimedia Content and the Semantic Web*. Wiley.
- Hustadt, U. (1994). Do we need the closed world assumption in knowledge representation? In Baader, F., Buchheit, M., Jeusfeld, M. A., and Nutt, W., editors, *Reasoning about Structured Objects: Knowledge Representation Meets*

Databases, Proceedings of 1st Workshop KRDB'94, Saarbrücken, Germany, September 20-22, 1994, volume 1 of *CEUR Workshop Proceedings*. CEUR-WS.org.

- Johnston, B., Yang, F., Mendoza, R., Chen, X., and Williams, M.-A. (2008). Ontology based object categorization for robots. In *PAKM '08: Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management*, volume 5345 of *LNCS*, pages 219–231, Berlin, Heidelberg. Springer-Verlag.
- Katz, Y. and Cuenca Grau, B. (2005). Representing qualitative spatial information in OWL-DL. In *Proceedings of the Workshop on OWL: Experiences and Directions, OWLED*, volume 188 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Kosslyn, S. M. (1994). *Image and brain: the resolution of the imagery debate*. MIT Press, Cambridge, MA, USA.
- Kosslyn, S. M., Thompson, W. L., and Ganis, G. (2006). *The Case for Mental Imagery*. Oxford University Press, New York, USA.
- Krötzsch, M., Rudolph, S., and Hitzler, P. (2008). ELP: Tractable rules for OWL 2. In *The Semantic Web, 7th International Semantic Web Conference, ISWC*, volume 5318 of *Lecture Notes in Computer Science*, pages 649–664. Springer.
- Kuhn, W., Raubal, M., and Gärdenfors, P. (2007). Editorial: Cognitive semantics and spatio-temporal ontologies. *Spatial Cognition & Computation: An Interdisciplinary Journal*, 7(1):3–12.
- Lovett, A., Dehghani, M., and Forbus, K. (2006). Efficient learning of qualitative descriptions for sketch recognition. In *20th International Workshop on Qualitative Reasoning*.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.
- Maillot, N. and Thonnat, M. (2008). Ontology based complex object recognition. *Image Vision Comput.*, 26(1):102–113.
- Motik, B., Horrocks, I., and Sattler, U. (2009). Bridging the gap between owl and relational databases. *J. Web Sem.*, 7(2):74–89.
- Motik, B., Sattler, U., and Studer, R. (2005). Query answering for owl-dl with rules. *J. Web Sem.*, 3(1):41–60.

- Neumann, B. and Möller, R. (2008). On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82–101.
- Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42(3):145–175.
- Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. MIT Press.
- Qayyum, Z. U. and Cohn, A. G. (2007). Image retrieval through qualitative representations over semantic features. In *Proceedings of the 18th British Machine Vision Conference (BMVC)*, pages 610–619.
- Quattoni, A. and Torralba, A. (2009). Recognizing indoor scenes. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:413–420.
- Rector, A. L., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., and Wroe, C. (2004). OWL pizzas: Practical experience of teaching OWL-DL: Common errors and common patterns. In Motta, E., Shadbolt, N., Stutt, A., and Gibbins, N., editors, *EKAW*, volume 3257 of *Lecture Notes in Computer Science*, pages 63–81. Springer.
- Sarifuddin, M. and Missaoui, R. (2005). A new perceptually uniform color space with associated color similarity measure for contentbased image and video retrieval. In *Multimedia Information Retrieval Workshop, 28th annual ACM SIGIR conference*, pages 3–7.
- Schill, K., Zetzsche, C., and Hois, J. (2009). A belief-based architecture for scene analysis: From sensorimotor features to knowledge and ontology. *Fuzzy Sets and Systems*, 160(10):1507–1516.
- Sirin, E., Smith, M., and Wallace, E. (2008). Opening, closing worlds - on integrity constraints. In *Proceedings of the Fifth OWLED Workshop on OWL: Experiences and Directions*. CEUR Workshop Proceedings.
- Socher, G., Geleit, Z., et al. (1997). Qualitative scene descriptions from images for integrated speech and image understanding. Technical report: <http://www.techfak.uni-bielefeld.de/techfak/persons/gudrun/pub/d.ps.gz>.
- Tao, J., Sirin, E., Bao, J., and McGuinness, D. (2010a). Extending OWL with integrity constraints. In *Proceedings of the International Workshop on Description Logics (DL)*, volume 573 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Tao, J., Sirin, E., Bao, J., and McGuinness, D. L. (2010b). Integrity constraints in OWL. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. AAAI Press.

- Tobies, S. (2001). Complexity results and practical algorithms for logics in knowledge representation. Technical report, RWTH Aachen, Germany. PhD thesis.
- Williams, M. (2008). Representation = grounded information. In *PRICAI '08: Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence*, volume 5351 of *LNCS*, pages 473–484, Berlin, Heidelberg. Springer-Verlag.
- Williams, M., McCarthy, J., Gärdenfors, P., Stanton, C., and Karol, A. (2009). A grounding framework. *Autonomous Agents and Multi-Agent Systems*, 19(3):272–296.