Computer Vision and Image Understanding 116 (2012) 698–714

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A model for the qualitative description of images based on visual and spatial features ☆

Zoe Falomir [a,b,*], Lledó Museros [a], Luis Gonzalez-Abril [c], M. Teresa Escrig [b], Juan A. Ortega [d]

[a] Department of Engineering and Computer Science, Universitat Jaume I, ES Tecnologia i Ciéncies Experimentals, Av. Vicent Sos Baynat s/n, E-12071 Castelló, Spain
[b] Cognitive Robots SL, Universitat Jaume I, Espaitec I, Av. Vicent Sos Baynat s/n, E-12071 Castelló, Spain
[c] Department of Applied Economics I, Universidad de Sevilla, Facultad de Ciencias Económicas, Avda. Ramón y Cajal, 1, E-41018 Sevilla, Spain
[d] Department of Languages and Computer Systems, Universidad de Sevilla, ETS Ingeniería Informática, Avda. Reina Mercedes s/n, E-41012 Sevilla, Spain

ARTICLE INFO

Article history:
Received 27 October 2010
Accepted 31 January 2012
Available online 16 February 2012

Keywords:
Computer vision
Qualitative shape
Qualitative colours
Qualitative orientation
Spatial description

ABSTRACT

An approach that provides a qualitative description of any image is presented in this paper. The main visual features (shape and colour) and the main spatial features (fixed orientation, relative orientation and topology) of each object within the image are described. This approach has been tested in two real scenarios that involve agents and human interaction: (i) images captured by the webcam of a mobile robot while it navigates, and (ii) images of tile compositions captured by an industrial camera used to select tile pieces to be used in assembling tile mosaics. In both scenarios, promising results have been obtained.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Digital images are fully integrated within modern daily life. Digital cameras are used to take photographs of trips and holidays, mobile phone cameras allow users to capture any scene in everyday life, and webcams in laptops are used to communicate the images of users' surroundings instantly across the network. The digital images generated can be easily copied, deleted, edited, sent by email or multimedia messages, included in web pages, etc. and computer systems and programs have been developed to provide all these possibilities. However, there is still no system capable of describing a digital image cognitively, that is, in a similar way to how people do it.

Psychological studies carried out on how people describe images [1–4] explain that people find the most relevant content in the images and use words (qualitative tags) to describe it. Usually different colours/textures in an image indicate different objects/regions of interest to people [5]. Moreover, cognitive studies [6] explain that, although 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, the knowledge about this image that can be retrieved from memory is qualitative because absolute locations, wavelengths and intensities cannot be retrieved from human memory. Thus, qualitative representations are similar in many ways to the *mental images* that people report when they describe what they have seen from memory or when they attempt to answer questions on the basis of visual memories [7,8].

Therefore, a cognitive description of any digital image must be a qualitative description that could be understood and interpreted by human users, which would allow the user-machine communication in many applications to be enhanced. For example, the qualitative description obtained may be used as the key search in image retrieval from data bases, and it may also be easily post-processed to produce a written narrative description of any image that may be included in a user-interface or read aloud by a speech synthesiser application for blind users to know what the image shows. Moreover, a cognitive description of a digital image must be able to describe any kind of object/region in the image by its features, regardless of whether it has been seen or unseen previously, as people can describe a scene or objects that they have never seen before or the names of which they cannot recall. Finally, it must be possible to extract and compute a cognitive description of any digital image automatically. It must also be independent of the image segmentation methods used to obtain the relevant regions in

☆ This paper has been recommended for acceptance by R. Bergevin.

* Corresponding author at: Department of Engineering and Computer Science, Universitat Jaume I, ES Tecnologia i Ciéncies Experimentals, Av. Vicent Sos Baynat s/n, E-12071 Castelló, Spain.

E-mail addresses: zfalomir@uji.es, zfalomir@gmail.com (Z. Falomir), museros@uji.es (L. Museros), luisgon@us.es (L. Gonzalez-Abril), escrigm@uji.es (M.T. Escrig), jortega@us.es (J.A. Ortega).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2012.01.007

the image, so that the system could apply a new more efficient method when it appears in the literature.

However, using computers to extract visual information from space and interpreting it in a meaningful way as human beings can do remains a challenge. Because digital images represent visual data numerically, most image processing has been carried out by applying mathematical techniques to obtain and describe image content. In the recent literature there are methods that describe and compare digital images numerically in order to obtain a degree of similarity between them [9–12]. All these approaches succeeded in the task they were designed for. However, they produce huge numerical file descriptions that cannot be interpreted or given a meaning without a correspondence of descriptions produced by other images of visualised scenes or objects previously stored in memory. Moreover, most of them need training or learning techniques. The main disadvantage of these methods is that a repository of all the possible images of scenes or objects existing in the world is still not possible. Therefore, they only succeed in specific delimited contexts. And they are not cognitive because they are not able to describe any feature of an object or scene that they have not seen before, that is, that have not been previously stored in memory.

Furthermore, extracting semantic information from images is still an on-going area of research in computer vision. The association of meaning with the representations of objects obtained by robotic systems, also known as the *symbol-grounding problem*, is a prominent issue within the field of Artificial Intelligence [13]. A qualitative description of images can contribute in this topic because it would be understandable not only by people but also by intelligent agents. The advantage of a description based on qualitative tags is that a semantic meaning can be assigned to them by means of ontologies. Therefore, the knowledge of any agent able to describe images qualitatively would be increased, i.e. a software agent could 'know' the content of the images on the Web or a physical agent (i.e. a mobile robot) could 'know' the features of all the objects in the images captured by its webcam even if these objects have not been seen before.

In this paper, the contribution presented is the automatic extraction of a qualitative description of any digital image based on the description of the visual and spatial information of the relevant regions within it, which are extracted independently of the segmentation method applied to the image. Specifically, the Qualitative Image Description approach (QID approach) uses qualitative models of shape and colour to describe the visual features of each relevant region in the image, and the qualitative models of topology [14] and orientation [15,16] to describe their spatial features.

The QID approach is intended to contribute to the solution of the open research issues previously presented. In this paper, the QID approach is applied to two real-world scenarios that involve agents and human interaction: (i) images captured by the webcam of a mobile robot while it navigates, and (ii) images of tile compositions captured by an industrial camera used to select tile pieces to be used in assembling tile mosaics. In addition, here the QID approach is generalised to show its flexibility in all kind of images and also its adaptability to other scenarios (i.e. medical image description, geographical image description, etc.).

This flexibility opens the way to future applications that involve extending the QID approach for: (1) extracting meaning from images and improving the understanding of those images by web agents by translating the QID to a description logics-based ontology; (2) obtaining a semantic similarity measure between the meaning of two ontological descriptions, where this similarity would calculate the resemblance between the different instances generated; (3) measuring the similarity of two qualitative image descriptions from the point of view of human thinking; (4) using that similarity measure for visual image/scene recognition and retrieval from a cognitive point of view (applicable to design automation processes, psychological research, and other fields involving imitation and study of human perception); (5) generating a human-language description for each processed image using a context free grammar which may produce a written paragraph describing the scene and which may be read aloud for a speech-synthesiser application for blind users to understand; etc.

The remainder of this paper is organised as follows. Section 2 presents the related work. Section 3 presents the QID approach and its implementation is explained in Section 4. Section 5 details the two scenarios where the experimentation was carried out and the results obtained. Finally, conclusions are drawn in Section 6.

2. Related work

Similar approaches that extract qualitative or semantic information from images representing scenes have appeared in the literature [17–20,10]. Socher et al. [17] provided a robotic manipulator system with a verbal description of an image so that it can identify and pick up an object that has been previously modelled geometrically and then categorised 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. [18] proposed a qualitative description for sketch image recognition, which described lines, arcs and ellipses as basic elements and also the relative position, length and orientation of their edges. Qayyum and Cohn [19] divided landscape images using a grid for their description so that semantic categories (grass, water, etc.) could be identified and qualitative relations of relative size, time and topology could be used for image description and retrieval in data bases. Oliva and Torralba [20] obtained 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) that enable the images to be classified in the following semantic categories: coast, countryside, forest, mountain, highway, street, close-up and tall building. Quattoni and Torralba [10] proposed an approach for classifying images of indoor scenes in semantic categories such as book store, clothing store, kitchen, bathroom, restaurant, office, classroom, etc. This approach combined global spatial properties and local discriminative information (i.e. information about objects contained in places) and used learning distance functions for visual recognition.

All the studies described above provide evidence for the effectiveness of using qualitative information to describe images. However, in the approach developed by Socher et al. [17], a previous object recognition process is needed before it becomes possible to give a qualitative description of the image of the scene the robot manipulator has to manage, whereas the QID approach is able to describe the image of the scene in front of the robot without this prior process. The approach of Lovett et al. [18] is applied to sketches, while the QID approach is applied to digital images captured from the real robot environment. Qayyum and Cohn [19] used a grid to divide the image and to describe what is inside each grid square (grass, water, etc.), which is suitable for their application. However, the objects are divided into an artificial number of parts that depend on the size of the cell, while the QID approach extracts complete objects. Oliva and Torralba's [20] approach 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. Quattoni and Torralba's [10] approach performs well for recognising indoor scenes, although it uses a learning distance function and, therefore, it must be

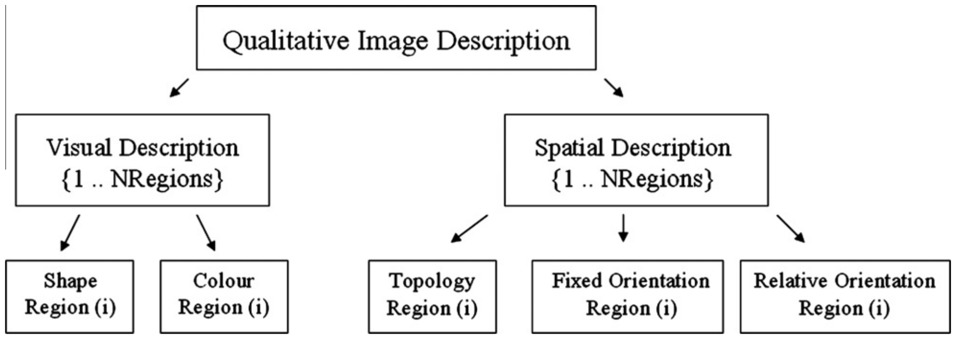


Fig. 1. Structure of the qualitative image description obtained by the QID approach.

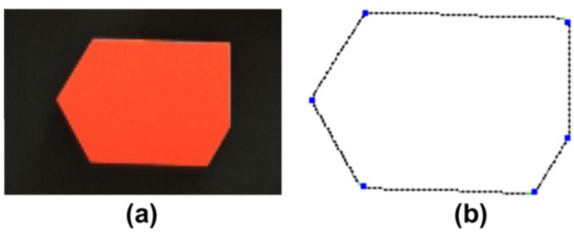


Fig. 2. Example of (a) an object within a digital image and (b) the extraction of the relevant points of the boundary of the object.

trained on a dataset, while the QID approach does not require training.

3. The Qualitative Image Description (QID) approach

Some studies on how people describe images can be found in the literature [1–4] whose main objective is analysing people's image descriptions for image retrieval in databases.

Jörgensen's research [1] studied image attributes typically noted by participants in a series of describing tasks involving activities such as viewing images, describing them for a retrieval system and describing them from memory. Their results show that the mentioned attributes may be distributed in the following classes: objects, people, colour, visual elements (e.g. shape, texture), location, description (e.g. number of objects), people-related attributes (e.g. emotion, social status), art-historical information (e.g. a picture, a photo, a print), abstract concepts (e.g. theme, atmosphere), content/story (e.g. activity, event), external relation (e.g. comparison, similarity) and viewer response (e.g. conjecture, uncertainty). These classes were also found in studies by Laine-Hernandez and Westman [2] analysing of how people describe journalistic photographs.

Psychological studies by Greisdorf and O'Connor [3] showed that seven basic attributes generally ascribed to images when computer users look at them are: objects, colour, shape, texture, location, actions and affects. On the other hand, research by Wang et al. [4] on how humans describe relative positions of image objects show that the relations of direction (*right*, *left*, *above*, *below*), topology (*overlap*, *separate*, *touch*, *in*, *out*, etc.) and distance (*far*, *near*, etc.) are the most used.

These studies provide us an idea of all the concepts people pay attention to when an image is looked at. In order to simplify the amount of information to extract from an image, the QID approach tackles the problem of qualitative image description by describing objects within an image visually and spatially as it is shown in Fig. 1. As visual features, the QID approach describes the shape (Section 3.1) and colour (Section 3.2) of each region, which are absolute properties that only depend on the region itself. As spatial features, the QID approach describes the topological relations (Section 3.3) and the qualitative orientation relations (Section 3.4) of the regions within the image, which are properties defined with respect to other regions.

3.1. Qualitative Shape Description (QSD)

Cognitively shape is defined as an aspect of a stimulus that remains invariant despite changes in size, position and orientation [21]. By knowing the shape of an object, people can predict more facts about that object than by knowing any other property (e.g. its size, what kind of object it is, what it is used for and so on) [5].

Usually, when people describe the shape of an object, they distinguish between straight sides and curved ones, describe angles and their convexity, and compare the lengths of the sides of the object. Hence, these features are used in QSD [22] which is briefly presented next.

Given a digital image containing an object (see the example in Fig. 2a), the boundary of this object are firstly extracted by the QSD and by analysing the slope defined by groups of points contained on this boundary, a set of relevant points, denoted by $\{P_0, P_1, \ldots, P_N\}$, are obtained, as a second step (see the example in Fig. 2b). Then, each relevant point P is described by a set of four features $\langle \mathrm{KEC}_P, A_P$ or $\mathrm{TC}_P, \mathrm{L}_P, \mathrm{C}_P\rangle$, which are defined below.

The first feature is the Kind of Edges Connected (denoted by **KEC**) and it indicates the connection 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 straight line and a curve;
- *curve-line*, if *P* connects a curve and a straight 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 discretised by using the Angle Reference System or ARS = $\{°, A_{LAB}, A_{INT}\}$ where, degrees (°) indicates the unit of measurement of the smaller angle at each relevant point; A_{LAB} refers to the set of labels for the angles; and A_{INT} refers to the values in degrees ° related to each label: $A_{LAB} = \{A_1, A_2, \ldots, A_{KA}\}$, and $A_{INT} = \{[0, a_1], (a_1, a_2], \ldots, (a_{KA-1}, 180]\}$ where *KA* is the number of labels[1].

On the other hand, if the **KEC** is a curvature-point or a *p-curve*, 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 = $\{°, TC_{LAB}, TC_{INT}\}$. As it is shown in Fig. 3a, the Type of

[1] The number of labels in this feature and in the rest of features must be defined in each situation as it can be seen in Section 5.

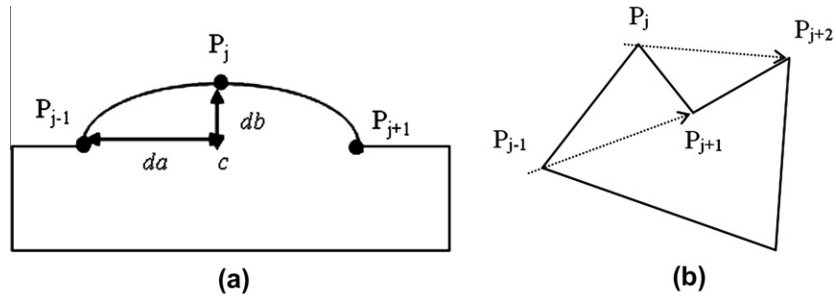
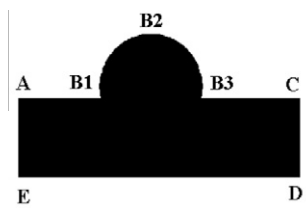


Fig. 3. Characterization of: (a) point P_j as a point of curvature, and (b) points P_j as *convex* and P_{j+1} as *concave*.

Table 1
QSD of an object containing straight segments and curves before parametrising of the features of shape: Angle (*A*), Type of Curvature (*TC*) and Length (*L*).

Object 1		QSD
		*QSD(Object 1) = [*
	A:	[line-line, *A*(*A*), *L*(*A*), convex],
	B1:	[line-curve, *A*(*B*1), *L*(*B*1), concave],
	B2:	[curvature-point, *TC*(*B*2), *L*(*B*2), convex],
	B3:	[curve-line, *A*(*B*3), *L*(*B*3), concave],
	C:	[line-line, *A*(*C*), *L*(*C*), convex],
	D:	[line-line, *A*(*D*), *L*(*D*), convex],
	E:	[line-line, *A*(*E*), *L*(*E*), convex]].

Curvature at P_j is determined by calculating first the point *c*, which is the half-point of the line between P_{j-1} (initial point of the curve) and P_{j+1} (final point of the curve). Next, the distance between P_{j-1} and *c*, named *da*, and the distance between P_j and *c*, named *db*, are calculated, and finally, the angle that determines the TC is obtained by $Angle(P) = 2 \cdot atan2(da/db) \cdot 180/\pi$ in degrees (°). In TCRS, TC_{LAB} refers to the set of labels for curvature; and TC_{INT} refers to the values of degrees (°) related to each label: $TC_{LAB} = \{TC_1, TC_2, \ldots, TC_{KTC}\}$, and $TC_{INT} = \{[0,tc_1], (tc_1,tc_2], \ldots, (tc_{KTC-1},180]\}$ where *KTC* is the number of labels.

The third feature considered is the compared length (denoted by **L**) which is defined by the Length Reference System or LRS = {UL, L_{LAB}, 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* = (*length of* 1*st edge*)/(*length of* 2*nd 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: $L_{LAB} = \{L_1, L_2, \ldots, L_{KL}\}$, and $L_{INT} = \{[0,l_1], (l_1,l_2], \ldots, (l_{KL-1},l_{KL})\}$ where *KL* is the number of labels and l_{KL} is the maximum number of times that an edge of a shape can be larger than another edge connected by a relevant point, which will be determined by the kind of shapes processed in the application.

The last feature to be considered is the Convexity (denoted by **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, as Fig. 3b shows: P_j is characterised as *convex*, whereas P_{j+1} is characterised as *concave*. Note that mathematically P_j cannot be within the oriented line from P_{j-1} to P_{j+1}, otherwise it will not be a relevant point of the shape. Moreover, the convexity of the angle at the relevant point indicates if the smaller angle is measured from the inside or the outside of the object.

Thus, the shape is described as a set of qualitative descriptions of relevant points as:

$$[[KEC_0, A_0|TC_0, L_0, C_0], \ldots, [KEC_{n-1}, A_{n-1}|TC_{n-1}, L_{n-1}, C_{n-1}]]$$

where *n* is the total number of relevant points of the object, KEC_i describes the Kind of Edges Connected by the first relevant point of the shape of the object, $A_i|TC_i$ describes the Angle or the Type of Curvature defined by the first relevant point of the shape of the object, L_i describes the compared length of the edges connected by the first relevant point of the shape of the object and finally, C_i describes the convexity of the first relevant point of the shape of the object. By convention, the first relevant point to be described (denoted by P_0) is always the one closest to the upper-left corner of the image and the rest of the relevant points are described cyclically in a clockwise direction. An example of the QSD of an object is shown in Table 1.

3.2. Qualitative Colour Description (QCD)

Although millions of colours can be defined in computer systems, the basic colours that can be named by users are limited to about 10–20 [23]. Moreover, a real fact in human cognition is that people go beyond the purely perceptual experience to classify things as members of categories and attach linguistic labels to them, and colour is not an exception: fresh blood and ripe tomatoes are all classified as *red*, even though they produce their own particular colour coordinates [5].

Colour naming models are designed to relate a numerical colour space with semantic colour names used in natural language. In the literature, different colour naming models were defined based on different colour spaces: CIE Lab colour space [24–26]; CIE Lab and HSL colour spaces [27]; Musell colour space (L*C*H) [28]; HSV colour space [29]; and so on. The colour naming model used in the QID approach is based on HSL colour space which translates the standard Red, Green and Blue (sRGB) colour channels of the centroid of each segmented object in an image into coordinates of Hue, Saturation and Lightness (HSL) colour space (see Fig. 4).

The colour of the centroid is the representative of a region because the image processing algorithms used extract colour homogeneous regions and do not deal with the problem of patterns. Moreover, according to Sarifuddin and Missaoui [30], the HSL colour space is intuitive to use for giving a name to the perceptual colour of the object and for adding semantic labels to this name in order to refer to the richness (saturation) or the brightness of the colour (lightness).

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

HSL colour space distributes colours in the following way (see Fig. 4). The *rainbow colours* are located in the horizontal central circle. The colour lightness changes in the vertical direction, therefore *light rainbow colours* are located above, while *dark rainbow colours* are located below. The colour saturation changes from the boundary of the two cone bases to the axis of the cone bases, therefore, *pale rainbow colours* are located inside the horizontal central circle. As a consequence of the changing colour saturation and lightness, the vertical axis locates the qualitative colours corresponding to the *grey scale*. According to this, the QID approach considers $M = 5$ colour sets: (1) grey colours, (2) rainbow colours, (3) pale rainbow colours, (4) light rainbow colours and (5) dark rainbow colours where the QC_{LAB} and QC_{INT} are:

$$QC_{LAB_1} = \{G_1, G_2, G_3, \ldots, G_{KG}\}$$

$$QC_{INT_1} = \{[0, g_{ul_1}], (g_{ul_1}, g_{ul_2}], (g_{ul_2}, g_{ul_3}], \ldots, (g_{ul_{KG-1}}, 100] \in UL/UH \in [0, 360] \text{ and } US \in [0, g_{us_{MAX}}]\}$$

The saturation coordinate of the HSL colour space (US) determines if the colour corresponds to the grey scale or to the rainbow scale. The colour name set for the grey scale is defined by QC_{LAB_1} whose corresponding intervals of values in HSL are determined by QC_{INT_1}. All the colours in this set can take any value of hue, values of saturation between 0 and $g_{us_{MAX}}$ and values of lightness which determine the different colour names defined.

$$QC_{LAB_2} = \{R_1, R_2, R_3, \ldots, R_{KR}\}$$

$$QC_{INT_2} = \{[0, r_{uh_1}], (r_{uh_1}, r_{uh_2}], (r_{uh_2}, r_{uh_3}], \ldots, (r_{uh_{KR-1}}, 360] \in UH; \text{ where } UL \in (r_{ul_{MIN}}, r_{ul_{MAX}}] \text{ and } US \in (r_{us_{MIN}}, 100]\}$$

The colours in the rainbow scale are defined by the names in QC_{LAB_2} and they are considered the more saturated ones or the strong ones. In QC_{INT_2}, their saturation can take values between $r_{us_{MIN}}$ and 100, their lightness between $r_{ul_{MIN}}$ and $r_{ul_{MAX}}$ and the different values of hue are those which determine the colour names defined for this set.

$$QC_{LAB_3} = \{pale\_ + QC_{LAB_2}\}$$

$$QC_{INT_3} = \{[0, r_{uh_1}], (r_{uh_1}, r_{uh_2}], (r_{uh_2}, r_{uh_3}], \ldots, (r_{uh_{KR-1}}, 360] \in UH; \text{ where } UL \in (r_{ul_{MIN}}, r_{ul_{MAX}}] \text{ and } US \in (g_{us_{MAX}}, r_{us_{MIN}}]\}$$

The pale colour name set (QC_{LAB_3}) is defined by adding the prefix *pale_*to the colours defined for the rainbow scale (QC_{LAB_2}). These colour names are defined in QC_{INT_3} by the same hue and lightness intervals, they differ from rainbow colours by their saturation which can take values between $g_{us_{MAX}}$ and $r_{us_{MIN}}$.

$$QC_{LAB_4} = \{light\_ + QC_{LAB_2}\}$$

$$QC_{INT_4} = \{[0, r_{uh_1}], (r_{uh_1}, r_{uh_2}], (r_{uh_2}, r_{uh_3}], \ldots, (r_{uh_{KR-1}}, 360] \in UH; \text{ where } UL \in (r_{ul_{MAX}}, 100] \text{ and } US \in (r_{us_{MIN}}, 100]\}$$

$$QC_{LAB_5} = \{dark\_ + QC_{LAB_2}\}$$

$$QC_{INT_5} = \{[0, r_{uh_1}], (r_{uh_1}, r_{uh_2}], (r_{uh_2}, r_{uh_3}], \ldots, (r_{uh_{KR-1}}, 360] \in UH; \text{ where } UL \in (rd_{ul}, r_{ul_{MIN}}] \text{ and } US \in (r_{us_{MIN}}, 100]\}$$

The lightness coordinate (UL) determines the luminosity of the colour: dark and light colours are distinguished and given an explicit name in QC_{LAB_4} and QC_{LAB_5}, respectively, by adding the prefixes *dark_* and *light_* to the colour names in the rainbow scale (QC_{LAB_2}). The intervals of values for dark and light colour sets (QC_{INT_4} and QC_{INT_5}, respectively) can take the same values of hue and saturation as those taken by the rainbow colours in QC_{INT_2}. However, they take different values for lightness: light colours between $r_{ul_{MAX}}$ and 100 and dark colours between rd_{ul} and $r_{ul_{MIN}}$.

Note that, although colour identification depends on illumination, 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. In the same way, the QCRS deals with the lighting conditions maintaining the name of the colour depending on the hue and managing the lightness and saturation changes using the semantic prefixes *light*, *dark* and *pale*.

As an example, according to the previous definitions, if the colour of the object in Table 1 would have as HSL colour coordinates [0,0,0], the colour name assigned to it would be one defined in the grey scale (QC_{LAB_1}).

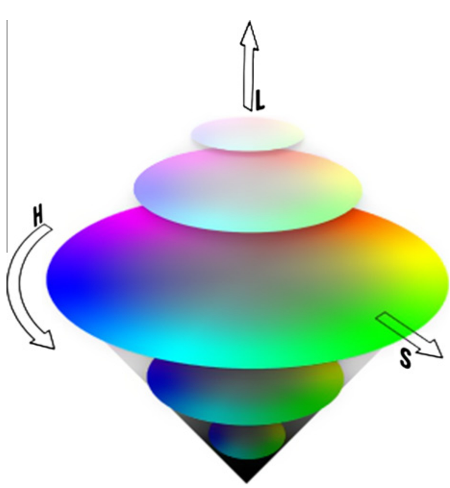


Fig. 4. HSL colour space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

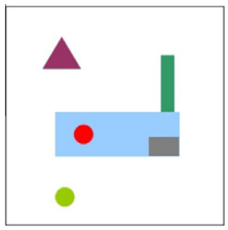
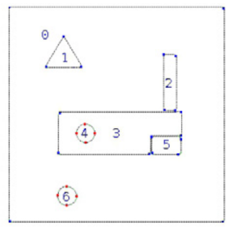
3.3. Topological description

Topological relations are spatial relations that are invariant under topological transformations, such as translation, rotation and scaling and also describe implicitly the relative distance between the objects [31].

In order to represent the topological relationships of the objects in the image, the intersection model defined by Egenhofer and Franzosa [14] 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 by that model cannot be distinguished by the QID approach and are all substituted by *touching*. For example, as Table 2 shows, in an image in two dimensions, it cannot be known: (i) if the green rectangle is overlap by the blue object or if they are touching; (ii) if the blue object is a blue rectangle covered by a grey rectangle or if they are a blue hexagon and a grey rectangle touching; and (iii) if any object covers a smaller or equal object and makes it invisible.

Therefore, the topological situation in space of an object A with respect to (wrt) another object B (A wrt B), is described by:

Table 2
Drawing for exemplifying the topological relations between object 3 and the other objects within the image described by the QID approach.

Image		Topology description
		... [3, [Container, 0], [touching, 2, 5], [disjoint, 1, 6], [completely_inside, 4]], ...

Topology = {*disjoint*, *touching*, *completely_inside*, *container*}

The QID approach determines if an object is *completely_inside* or if it is the *container_of* another object, and it also defines the *neighbours of level* of an object as all the other objects within the same container. The *neighbours of level* 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 (*DistanceThreshold*) set by experimentation.

Finally, as an example, the topological situation of the blue object in the drawing in Table 2 is described as having: (i) one *container* (the image, Object 0); (ii) an object located *completely inside* (the red circle, Object 4); (iii) two neighbours *touching* (the green and grey rectangles, Objects 2 and 5), and (iv) two neighbours *disjoint* (the purple triangle and the yellow circle, Objects 1 and 6).

3.4. Orientation description

Metric orientation information locates a point at any position on a line segment from the origin of a Cartesian reference system with a given angle. However, orientation information expressed in this way is not cognitive, because for human beings this kind of information is impossible to obtain for two main reasons: (i) our perceptual measurements (without any suitable tool) are quite imprecise, and we usually think of orientation as *left* or *right* but rarely as '*15 degrees to the north*'; and (ii) we hardly ever think that our orientation or position is somewhere with respect to an external Cartesian reference system unless we are using a compass. Therefore, qualitative models of orientation are used cognitively in many applications because they enable users to express their orientation in terms of non-metric information and also enable them to differentiate between given orientations and to reason about them.

The QID approach applies two kinds of qualitative orientation models: (i) the model by Hernández [15] in order to provide the orientation relations of the objects within the image fixed by the point of view of an external observer; and (ii) the Freksa's double-cross orientation model [16] in order to provide the orientation of the objects relative to other objects within the image and regardless of the orientation of the image given by an external observer.

Note that the QID approach discretizes the regions in the image by describing some relevant points of their boundary. Therefore, both region-based [15] and point-based [16] orientation models are suitable of application, as the orientation of a region is determined as the union of all the orientations obtained by the relevant points of its boundary.

This section is organised as follows. The fixed orientation model applied is explained in SubSection 3.4.1 and its relative orientation model used is described in SubSection 3.4.2. Moreover, in Section 3.4.3, a comparison of the reference frames used by each model of orientation is given. Finally, as the orientation relations of each region in an image depend on the total number of regions contained within this image, Section 3.4.4 explains which orientation relations can be computed according to the number of objects within the same container.

3.4.1. Qualitative model of fixed orientation

A Fixed Orientation Reference System (FORS) is defined which obtains the orientation of an object A 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 (see Fig. 5) which are labelled as:

$$FO = \{front(f), back(b), left(l), right(r), left\_front(lf), right\_front(rf), left\_back(lb), right\_back(rb), centre(c)\}$$

In order to obtain the fixed orientation of each object wrt another object or wrt the image, the QID 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*.

Moreover, the fixed orientation of the relevant points of all the objects in the image is also obtained wrt its centroid. This information is included in the visual description of the region because the obtained fixed orientations are related to each of the relevant points of the shape of the object and arranged wrt them.

Note that the FO information would change if there is a significant rotation of the image or if there is a significant translation or rotation of any of the objects in the image.

As an example, the fixed orientation (FO) of the purple triangle (Object 1) in the drawing in Table 3 is described as located: *front-left* wrt its container (the image or Object 0); *left* wrt the Object 2; *front-left* wrt the Object 3 and wrt Object 5; and *front* wrt the Object 6. Note that the FO wrt the red circle (Object 4) is not given because it is not a *neighbour of level* of the purple triangle (Object 1) since the red circle is *completely_inside* the blue rectangle (Object 3). Finally, the orientation of the vertices of the triangle wrt its centroid are provided: *front*, *back-right* and *back-left*.

3.4.2. Qualitative model of relative orientation

A Relative Orientation Reference System (RORS) is defined using Freksa's double-cross orientation model [16]. 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*

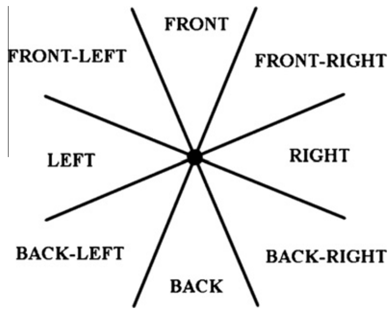


Fig. 5. The reference system used in the qualitative orientation model by Hernández [15].

and b, that is, *c wrt ab* (Fig. 6). This model divides the space into 15 regions labelled as:

$$\begin{aligned}\mathrm{RO} = \{&\mathit{left\_front}(\mathit{lf}), \mathit{straight\_front}(\mathit{sf}), \mathit{right\_front}(\mathit{rf}), \mathit{left}(\mathit{l}),\\ &\mathit{dentical\_front}(\mathit{idf}), \mathit{right}(\mathit{r}), \mathit{left\_middle}(\mathit{lm}), \mathit{same\_middle}(\mathit{sm}),\\ &\mathit{right\_middle}(\mathit{rm}), \mathit{identical\_back\_left}(\mathit{ibl}), \mathit{identical\_back}(\mathit{ib}),\\ &\mathit{identical\_back\_right}(\mathit{ibr}), \mathit{back\_front}(\mathit{bf}), \mathit{same\_back}(\mathit{sb}),\\ &\mathit{back\_right}(\mathit{br})\}\end{aligned}$$

In order to obtain the relative orientation of an object, the QID 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 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 the RO information would change if there is a significant translation of any of the objects in the image, whereas it would remain invariant to image and object rotations.

As an example, some relative orientations (RO) that can be extracted from the objects in the drawing in Table 4 are described next. The purple triangle (Object 1) is *right-middle, right-front* (*rm, rf*) wrt the reference system built from the green rectangle to the blue hexagon (RS(2,3)) but *left-middle, left-front* (*lm, lf*) in the opposite direction, that is, wrt RS(3,2). Note that, as the relations of orientation obtained directly from opposite RSs are exactly the opposite orientations, they are not given in the QID approach. In the same way: (i) the green rectangle (Object 2) is *left-middle* wrt the reference system built from the purple triangle to the green rectangle (RS(1,5)); (ii) the blue hexagon (Object 3) is *left-middle, right-middle* wrt the RS built from the triangle to the yellow circle (RS(1,6)); (iii) the grey rectangle (Object 5) is *right-middle, right-front* wrt the RS built form the purple triangle to the green rectangle (RS(1,2)) and, wrt the same RS, the yellow circle (Object 6) is *right-middle*. The rest of the orientations described in Table 4 are described similarly. Note that the relative orientation of the red circle (Object 4) cannot be provided as this object has not any neighbours of level (objects contained by the same container).

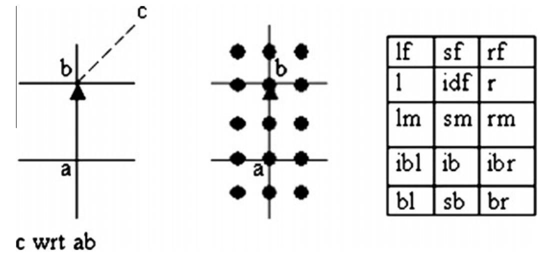


Fig. 6. The qualitative orientation model by Freksa [32] and its iconical representation: *l is left, r is right, f is front, s is straight, m is middle, b is back and i is identical.*

3.4.3. Reference frames of fixed and relative orientation

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 made by Hernández [15], it is considered that:

- the reference system or frame in the FORS 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 orientation 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 RORS, 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:

(a) describe the implicit orientations of the objects in the image from the point of view of an external observer (i.e. robot camera) and regardless of the number of objects within the image, and

Table 3
Drawing for exemplifying the fixed orientation relations of object 1 described by the QID approach.

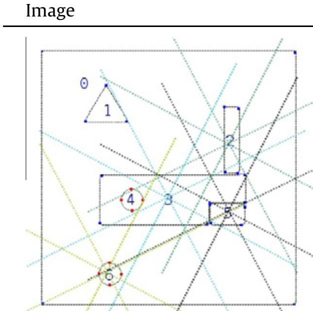
Image	Fixed orientation description
	. . . [1, [Container, 0], [Orientation wrt 0: front_left], [Orientation wrt Neighbours: [2, left], [3, front_left], [5, front_left], [6, front]], . . . [Vertices_Orientation, front, back_right, back_left],],], . . .

Table 4
Drawing for exemplifying the relative orientation relations of object 1 described by the QID approach.

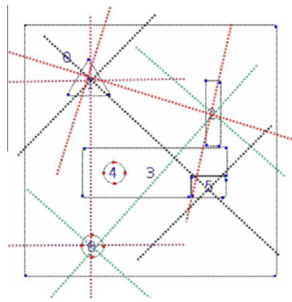
Image	Relative orientation description
	[1, [RO: [[2,3], rm, rf], [[2,5], br], [[2,6], rm], [[3,5], bl], [[3,6], br], [[5,6], rm]], [2, [RO: [[1,5], lm], [[1,6], lm], [[5,6], br]]], [3, [RO: [[1,6], lm, rm]] [4, [RO: -], [5, [RO: [[1,2], rm, rf], [[1,6], lm], [[2,6], lm]], [6, [RO: [[1,2], rm], [[1,3], rf], [[1,5], rm], [[2,3], lf], [[2,5], lf], [[3, 5], rm, br]],

Table 5
Spatial features described depending on the number of objects at each level. The 'x' symbol indicates that the description can be provided with this amount of objects, while '–' means the opposite.

Spatial features described	Objects within the same container		
	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

(b) 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).

3.4.4. Organization of orientation relations

In our approach, orientation relations between the objects in the image are structured in levels of containment. The fixed orientation [15] of a region is defined with respect to its *container* and *neighbours of level*, while the relative orientation of a region [32] 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 5.

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.

3.5. Structure of the Qualitative Image Description (QID)

Summarizing, the general structure of the Qualitative Image Description (QID) provided by our approach is defined as a set of qualitative tags such that:

$$\textbf{QID}(\text{IdImage}) = [\text{SpatialDescription}(\text{NRegions}), \text{VisualDescription}(\text{NRegions})]$$

For each object/region detected in the image, the spatial information described consists of the identifier of the object/region, its topological relations wrt its container and the other objects in the image, its fixed orientation wrt its container and wrt its neighbours, and its relative orientation wrt all the reference systems defined by its neighbours:

$$\textbf{SpatialDescription}(1..\text{NRegions}) = [\text{IdRegion}, \text{Topology}(\text{Container}), \text{FixedOrientation}(\text{Container}), \text{Topology}(\text{Region}), \text{FixedOrientation}(\text{Neighbours}), \text{RelativeOrientation}(\text{RSs})]$$

Topology(Container) = [*Container*, IdContainer]
FixedOrientation(Container) = [*Orientation wrt* IdContainer: FixedOrientationTags]
Topology(Region) = [*touching*(IdRegions), *disjoint*(IdRegions), *completely_inside*(IdRegions)]
FixedOrientation (1 .. NNeighbours) = [*Orientation wrt Neighbours*: [IdNeighbour, FOs]]
RelativeOrientation (1 .. NRSs) = [*Relative Orientation wrt Neighbours Disjoint*: [RSs, ROs]]
RS = [IdNeighbour_A, IdNeighbour_B]
FO ∈ {*front, back, left, right, left_front, right_front, left_back, right_back, centre*}
RO ∈ {*lf, sf, rf, l, idf, r, lm, sm, rm, ibl, ib, ibr, bf, sb, br*}

For each object/region detected in the image, the visual information described consists of its identifier, its colour and the description of the shape of each vertex:

$$\textbf{VisualDescription}(1..NRegions) = [IdRegion, QCD, QSD(RPs), Orientation(RPs)]$$

$\text{QCD} \in \text{QC}_{LAB1\ldots5}$ where,

$$\text{QC}_{LAB_1} = \{\text{G}_1, \text{G}_2, \text{G}_3, \ldots, \text{G}_{KG}\}$$

$$\text{QC}_{LAB_2} = \{\text{R}_1, \text{R}_2, \text{R}_3, \ldots, \text{R}_{KR}\}$$

$$\text{QC}_{LAB_3} = \{pale\_ + \text{QC}_{LAB_2}\}$$

$$\text{QC}_{LAB_4} = \{light\_ + \text{QC}_{LAB_2}\}$$

$$\text{QC}_{LAB_5} = \{dark\_ + \text{QC}_{LAB_2}\}$$

QSD(1..nRP) = [KEC, A or TC, L, C] where,

$$\mathrm{KEC} \in \{line-line, line-curve, curve-line, curve-curve, curvature-point\};$$

$$\mathrm{A} \in \mathrm{A}_{LAB} = \{\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_{KA}\}$$

$$\mathrm{TC} \in \mathrm{TC}_{LAB} = \{\mathrm{TC}_1, \mathrm{TC}_2, \ldots, \mathrm{TC}_{KTC}\}$$

$$\mathrm{C} \in \{convex, concave\}$$

$$\mathrm{L} \in \mathrm{L}_{LAB} = \{\mathrm{L}_1, \mathrm{L}_2, \ldots, \mathrm{L}_{KL}\}$$

$$\mathrm{Orientation}(1..\mathrm{nRP}) = [\mathrm{FOs}]$$

4. A computational model for the QID approach

The computational model for the QID approach first obtains the relevant regions of an image by applying image processing algorithms and then describes the visual and spatial features of these regions by applying qualitative models presented in Section 3 (see Fig. 7).

The approaches used to extract the main regions of a digital image are discussed in Section 4.1 whereas Section 4.2 outlines the proposed algorithm. The structure of the qualitative description obtained has been explained in Section 3.5.

4.1. Obtaining the relevant regions of any digital image

Region segmentation is defined by Palmer [5] as the process of dividing an image into mutually exclusive areas based on the uniformity of an image-based property, such as luminance, chromatic colour, texture, motion or binocular disparity. Two ways of approaching this task are also distinguished:

- *Boundary-based methods*, in which the visual system detects differences (or gradients) in local visual properties that divide one region from another. The methods that first detect the edges or boundaries in a digital image and then, from the obtained boundaries, extract the regions within it are included in this group. An example is the well-known Canny's segmentation method [33].
- *Region-based methods*, which consider that in an image, different colours or textures usually indicate different regions of interest to the human eye. The methods included in this group are those that extract the different regions of colour/texture/etc., from an image and then define the boundaries of these regions as the edges. An example of these methods is the one by Felzenszwalb and Huttenlocher [34].

As Felzenszwalb and Huttenlocher [34] mention, the problems of image segmentation and grouping remain great challenges for computer vision because, for obtaining a useful segmentation method, it has to: (i) capture perceptually important groupings or regions, which often reflect global aspects of the image; and (ii) be highly efficient, running in time nearly linear in the number of image pixels.

Generally, image region-based segmentation methods are considered more cognitive than boundary-based segmentation methods because the extracted edges are defined by the boundaries between colour regions and all of these regions are closed. In Fig. 8, the results of both methods applied to the segmentation of the same image can be compared. While Canny's segmentation method [33] obtains open edges, the boundaries extracted from Felzenszwalb and Huttenlocher's method [34] are all closed.

It is worth noting that the QID approach is not dependent on the region segmentation method used, and therefore, the most convenient method that obtains closed regions can be selected from the literature depending on the application.

4.2. The proposed algorithm

The computational procedure for the QID approach is outlined in Algorithm 1, which is described in the following subsection.

Algorithm 1. Obtaining the qualitative description of a digital image.

```
ImageRegions ← Image_Region_Segmentation(Image,Method)
for all Region R in ImageRegions do
  R.Points ← Find_Relevant_Points(R)
  R.Container ← Find_Container(R,ImageRegions)
  R.Centroid ← Find_Centroid(R)
  R.QC ← Qualitative_Colour(R)
  for all P in R.Points do
    R.QSD ← Qualitative_Shape(P,R)
  end for
  for all P in R.Points do
    Fixed_Orientation(P,R.Centroid)
    Fixed_Orientation(P,R.Container.Centroid)
  end for
  for all r in{ImageRegions|r.Container = R.Container} do
    if Touching(r,R,DistanceThreshold) then
      R.Neighbours_Touching ← r
    else
    R.Neighbours_Disjoint ← r
    end if
  end for
  for all r in{R.Neighbours_Touching or R.Neighbours_Disjoint}
  do
    for all P in R.Points do
      Fixed_Orientation_wrt_Neighbours_of_Level(P,r.Centroid)
    end for
  end for
  if R.Neighbours_Disjoint ⩾ 2 then
    RS ← Build_Reference_Systems(R.Neighbours_Disjoint)
    for all rs in RS do
      for all P in R.Points do
        Relative_Orientation(rs,P)
      end for
    end for
  end if
end for
```

First, the captured digital image (*Image*) is segmented into regions of interest (*ImageRegions*) by the selected method, which, as mentioned earlier, can be a boundary-based or a region-based segmentation method, depending on the application. Then, for each region (R) of interest:

- its boundary is processed and the relevant points (*R.Points*) that characterise its shape are extracted;
- its container (*R.Container*) is obtained and, as an inverse relationship, the current region is located *completely inside* its container;
- its centroid (*R.Centroid*) is calculated;
- its qualitative colour (*R.QC*) is obtained from the centroid of the region;
- its qualitative shape description (*R.QSD*) is obtained by describing the features of each of the relevant points on its boundary (*R.Points*);

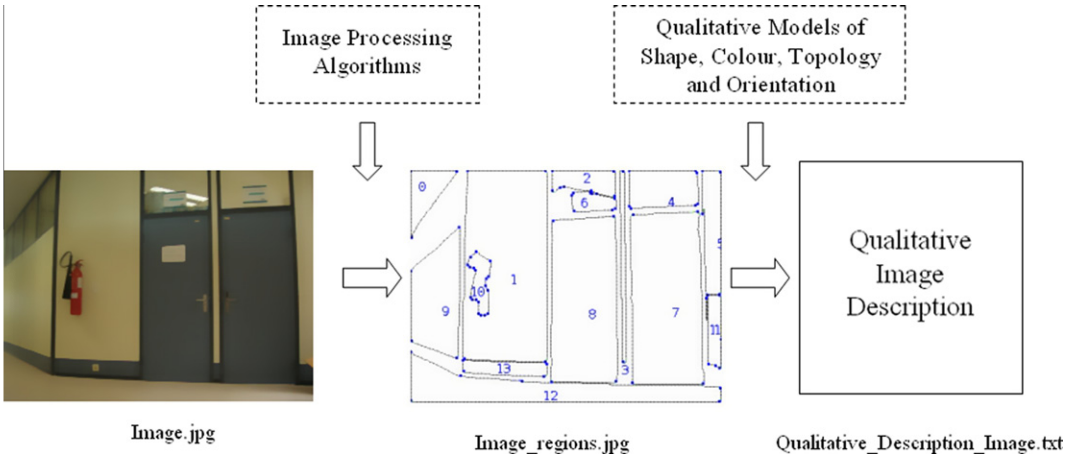


Fig. 7. Schema of the QID approach for qualitative image description.

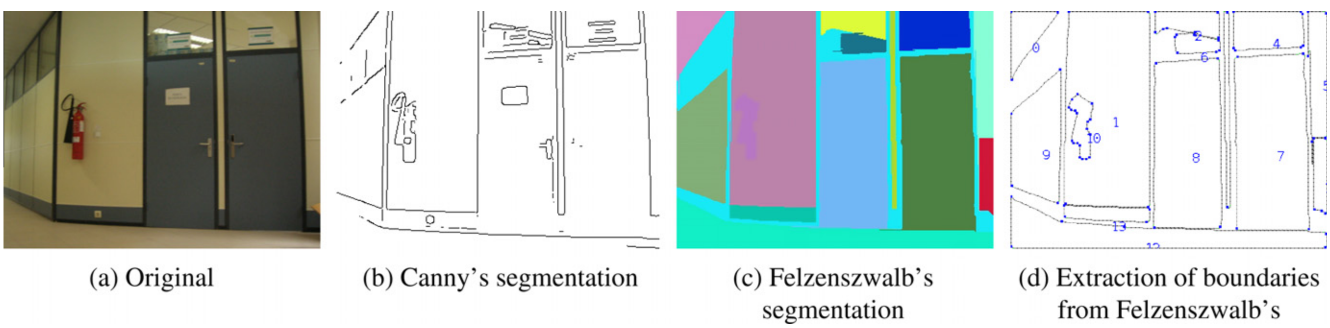


Fig. 8. Comparison of segmentation methods: (a) original image from the robot environment; (b) segmentation obtained by Canny's boundary based method [33]; (c) segmentation obtained by Felzenszwalb and Huttenlocher's region based method [34]; and (d) boundaries of the coloured regions in (c) extracted by our algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- the fixed orientation of each of the relevant points in the region (*R.Points*) is described with respect to its centroid (*R.Centroid*);
- its fixed orientation with respect to the centroid of its container (*R.Container.Centroid*) is obtained;
- its neighbours that are touching (or within a distance threshold) are obtained from all the regions with the same container (*R.Neighbours_Touching*). As an opposite relationship, all the neighbours that are disjoint are also obtained (*R.Neighbours_Disjoint*);
- its fixed orientation with respect to its neighbours on the same level (touching or disjoint) is obtained;
- a reference system (*RS*) is built and the relative orientations of the current region wrt all these reference systems are obtained for each pair of disjoint neighbours (*R.Neighbours_Disjoint*) of the current region.

The computational cost of the algorithm is $O(PR^3)$, where P is the largest number of relevant points that define a region in the image and R is the total number of regions in the image. Clearly, the computational cost of our QID-Algorithm peaks when a lot of regions are extracted in the image and those regions have irregular boundaries that are described by a large number of relevant points.

5. Experimentation and results

In this paper, the QID approach presented is applied to two real robot-working scenarios (the ones available at our laboratory at Universitat Jaume I), where digital images are managed for interaction between human users and physical agents (i.e. a mobile robot) or software agents (i.e. computer applications). Specifically, these scenarios are: (i) Scenario I: the description of images of the world captured by a webcam located on a mobile robot (Section 5.2), and (ii) Scenario II: the description of images of tile compositions captured by an industrial camera located on a platform that is used by a robot arm to assemble mosaics automatically (Section 5.3). The parameters selected for both scenarios are shown in Section 5.1 and an analysis of the results is shown in Section 5.4.

It is important to note that, although both scenarios presented were suitable for the test experiments, the parameters and the image segmentation method used by the QID approach can be adjusted for any kind of scenario that manages digital images and needs to improve human user interaction or the knowledge of the agents involved.

5.1. Parameter selection

In order to determine the interval values of the reference systems defined for describing the angles and lengths of the shape (ARS, TCRS and LRs) and the colour (QCRS) of the image regions, several experts in the implementation area (robotics and industrial tile mosaics) were asked for their opinions and know-how. First, they were asked about the most suitable granularity for describing each feature, that is, the number of qualitative labels to use to describe it. Then they were shown some angles, lengths and colours and they were asked to match a qualitative label to them. The AMEVA algorithm [35] was used to obtain the classes of the intervals, which were rounded off afterwards in order to obtain the same sets of intervals used in both applications.[2]

For the Angle Reference System or ARS = {°, A_{LAB}, A_{INT}}, the chosen granularity was 5 and the set of labels for the qualitative angles and the values of degrees (°) related to each label were:

$$A_{LAB} = \{very\_acute, acute, right, obtuse, very\_obtuse\}$$
$$A_{INT} = \{[0,40],(40,85],(85,95],(95,140],(140,180]\}$$

For the Type of Curvature Reference System or TCRS = {°, TC_{LAB}, TC_{INT}}, the chosen granularity was also 5 and the set of labels for the type of curvature and the values of degrees (°) related to each label were:

$$TC_{LAB} = \{very\_acute, acute, semicircular, plane, very\_plane\}$$
$$TC_{INT} = \{[0,40],(40,85],(85,95],(95,140),[140,180]\}$$

For the Length Reference System or LRS = {UL, L_{LAB}, L_{INT}}, the chosen granularity was 7, the set of labels that were related in order to compare length and the values related to each label were:

$$L_{LAB} = \{much\_shorter(msh), half\_length(hl), a\_bit\_shorter(absh), similar\_length(sl), a\_bit\_longer(abl), double\_length(dl), much\_longer(ml)\}$$

$$L_{INT} = \{(0,0.4),[0.4,0.6],(0.6,0.9),[0.9,1.1],(1.1,1.9),[1.9,2.1],(2.1,10)\}$$

For the Qualitative Colour Reference System or $QCRS = \{UH, US, UL, QC_{LAB_{1\ldots5}}, QC_{INT_{1\ldots5}}\}$, the qualitative labels related to each colour scale and the interval values which refers to the intervals of HSL colour coordinates associated with each colour label were the following (See Tables 8–12):

$$QC_{LAB_1} = \{black, dark\_grey, grey, light\_grey, white\}$$
$$QC_{INT_1} = \{[0,20),[20,30),[30,40),[40,80),[80,100) \in UL; \text{ where } UH \in [0,360] \text{ and } US \in [0,20]\}$$

$$QC_{LAB_2} = \{red, yellow, green, turquoise, blue, purple, pink\}$$

$$QC_{INT_2} = \{(335,360] \text{ and } [0,40],(40,80],(80,160],(160,200],(200,260],(260,297],(297,335] \in UH; \text{ where } US \in (50,100] \text{ and } UL \in (40,55]\}$$

$$QC_{LAB_3} = \{pale\_ + QC_{LAB_2}\}$$

$$QC_{INT_3} = \{(335,360] \text{ and } [0,40],(40,80],(80,160],(160,200],(200,260],(260,297],(297,335] \in UH; \text{ where } US \in (20,50] \text{ and } UL \in (40,55]\}$$

$$QC_{LAB_4} = \{light\_ + QC_{LAB_2}\}$$

$$QC_{INT_4} = \{(335,360] \text{ and } [0,40],(40,80],(80,160],(160,200],(200,260],(260,297],(297,335] \in UH; \text{ where } US \in (50,100] \text{ and } UL \in (55,100]\}$$

$$QC_{LAB_5} = \{dark\_ + QC_{LAB_2}\}$$

$$QC_{INT_5} = \{(335,360] \text{ and } [0,40],(40,80],(80,160],(160,200],(200,260],(260,297],(297,335] \in UH; \text{ where } US \in (50,100] \text{ and } UL \in (20,40]\}$$

For QC_{LAB_1}, the chosen granularity was 5, while for $QC_{LAB_{2\ldots5}}$, the chosen granularity was 7. Therefore, in the final QCRS, 10 basic colours are defined (*black, grey, white, red, yellow, green, turquoise, blue, purple, pink*) and by adding the semantic descriptors *pale_, light_* and *dark_*, a total of 5 + 7 · 4 = 33 colour names are obtained. Conway's research [23] showed that, although strictly speaking it may be accurate, people tend not to describe a colour as *dark pale blue* and may even consider this a contradiction. They also recommended that, in order to produce more cognitive colour name descriptions, no more than one adjective should be applied to a basic colour name. Furthermore, if a lightness and saturation modifier appears equally applicable to a particular colour, the saturation modifier should be chosen. This aspect is reflected in the QCRS.

5.2. Scenario I

Scenario I consists of the description of images captured by a webcam located on the top of a Pioneer 2 mobile robot[3] (see Fig. 9). For this scenario, Felzenszwalb's segmentation method [34] is used since it captures the perceptually important regions in an image, it is highly efficient, running nearly linear in time as regards the number of image pixels, and it also preserves detail in low-variability image regions while ignoring detail in high-variability regions. This is achieved by adjusting 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*, the minimum size of the extracted regions in pixels enforced by post-processing.

An excerpt of the qualitative description of the digital image shown in Fig. 9 is given in Table 6. Specifically, the qualitative spatial description of regions 1 and 10 and the qualitative visual description of regions 7 and 10 is shown in this table.

The spatial description of region 1 can be 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 and 13 (Note that some of these are not technically touching but are closer to region 1 than the threshold determined for this application). Its *disjoint neighbours* are the regions 0, 3, 4, 5, 6, 7, 11 and 12 and finally, the 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 is *left, back, back_left*, wrt region 3 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 6: 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 6 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 *much_shorter, much_longer, half,* and *much_longer*, respectively. 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.

With respect to computation time, it should be noted that, for the image shown in Fig. 9, the time of execution for the extraction

[2] For the sake of simplicity, the same sets of intervals were chosen, since there are few differences between them the two scenarios.

[3] http://www.mobilerobots.com

(a) Webcam located on a mobile robot (b) Image obtained (c) Felzenszwalb's segmentation: σ = 0.4, *k* = 500 and *min* = 1000

Fig. 9. The QID approach applied to Scenario I: describing an image of University Jaume I corridor taken by a webcam located on a Pioneer 2 mobile robot.

Table 6
An excerpt of the qualitative description obtained for the image captured by the Pioneer 2 webcam in Fig. 9.

[**SpatialDescription**,
(...)
[**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]]
]
(...)
[**10**, [[Container, 1] [Orientation wrt 1: left, back_left, back],
[None Neighbours of Level]]
],
(...)
]
[**VisualDescription**,
(...)
[**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]],
],
(...)
[**10**, dark_red,
[Boundary_Shape,
[line-line, obtuse, half_length, convex],
[line-line, obtuse, similar_length, convex],
(...)
[line-line, very_obtuse, similar_length, convex]]
[Vertices_Orientation, front, front, front_right, right, back_right (...)]],
],
(...)
]

of the main regions in the image is around 0.83 seconds and the time for generating the qualitative description of the image is around 2.19 seconds. The total time of execution is thus around 3.02 seconds using a computer with an Intel Core i5 processor at 2.27 GHz and 4 GB of RAM, running under an Ubuntu 10.04 (lucid) with a Linux kernel 2.6.32-21-generic.

The images of indoor scenes captured by a webcam located on a Pioneer 2 robot shown in Fig. 10 were used to test our approach. The segmentation result of each image and the QID obtained are available on-line.[4] The execution times obtained in the tests are shown in Fig. 11. The vertical axis represents the time in seconds, whereas the horizontal axis represents the computational cost of each QID description ($O(P\ R^3)$). This depends on the number of regions in the image (R) and the maximum or the largest number of relevant points that define a region in the image (P), because the temporal cost depends on both values. For example, the variable *SC17:10;8* in the horizontal axis indicates that the image *SC17* was described by the QID approach as being composed by 10 regions of with a maximum number of relevant points for each region of 8. The time needed for obtaining the QID of image *SC17* is 0.5 seconds.

5.3. Scenario II

Scenario II consists of the description of images of tile compositions captured by an industrial camera AVT-Guppy F033C located on a platform from which a robot arm picks up and places tile pieces for assembling tile mosaics (Fig. 12). In this scenario, the Canny method of segmentation [33] is used. This method obtains fast, good results for this application, since the boundaries of the objects are clearly defined and distinguished, as they are usually made up of straight edges or simple curves.

[4] http://dl.dropbox.com/u/17361913/CVIUTests.rar

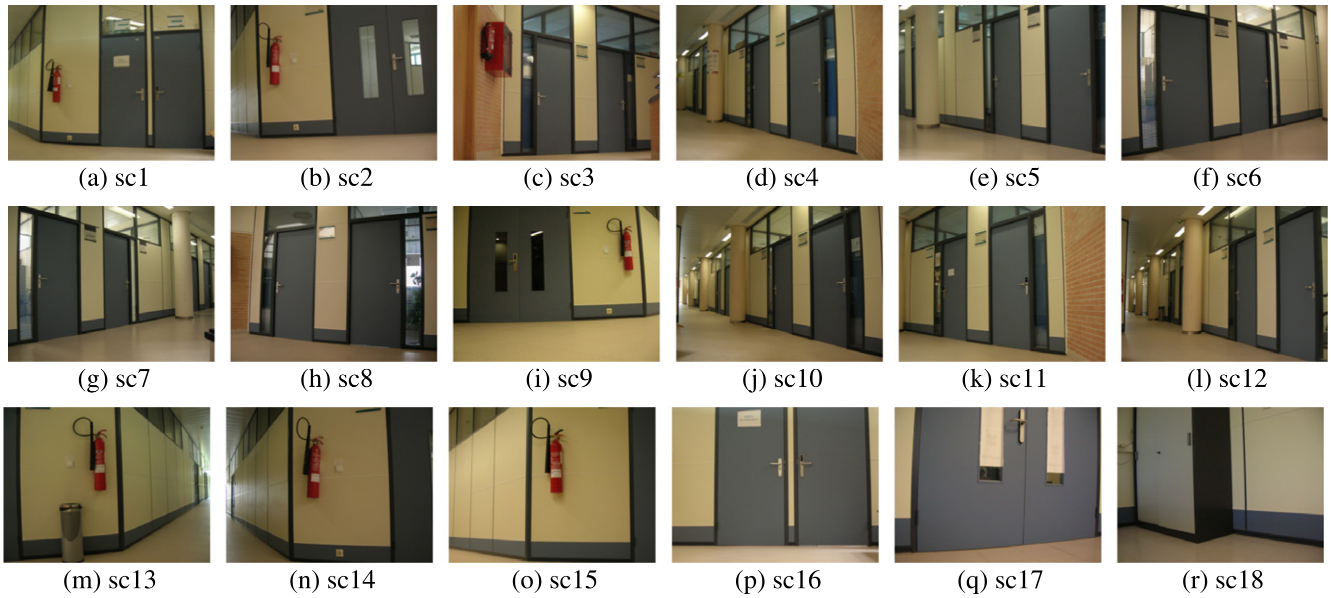


Fig. 10. Images of indoor scenes (SC) used for testing our approach.

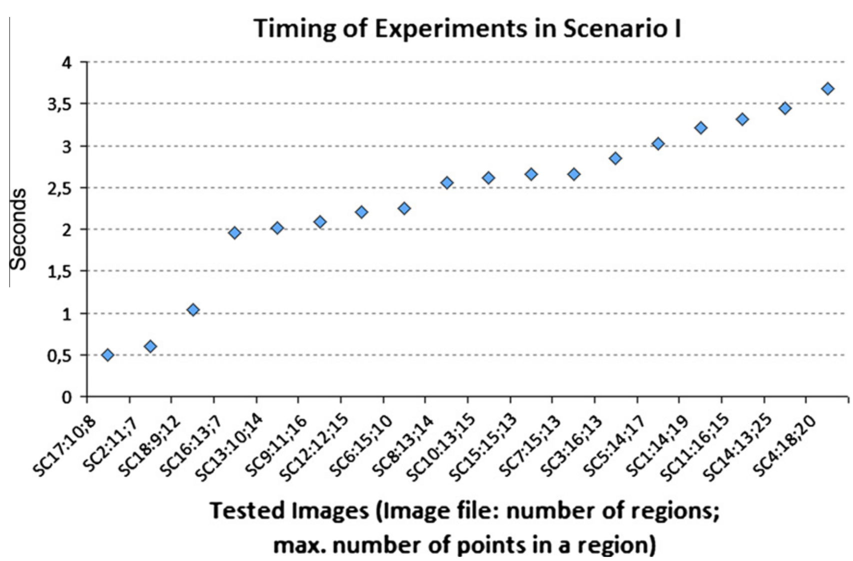


Fig. 11. Execution times of QID obtained for images in Fig. 10 related to the computational cost of QID-Algorithm for each situation.

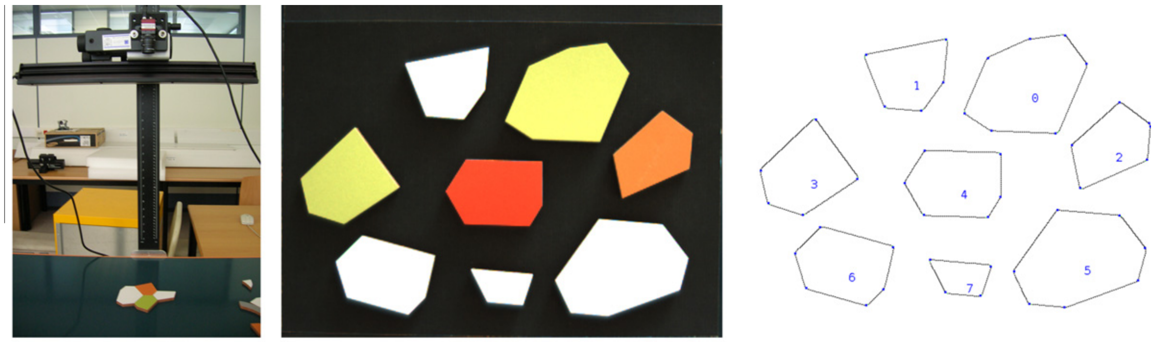
An excerpt of the qualitative description of the digital image shown in Fig. 12 is given in Table 7. Specifically, the qualitative spatial description of regions 3 and 5 and the qualitative visual description of region 3.

The spatial description of region 3 can be read as follows: its *container* is the Image and it is located wrt to the Image at *left*. Its *disjoint neighbours* are the rest of the pieces of tile composition since none of them is touching region 3. The *fixed orientation* of region 3 wrt region 0 is *left, back_left*, wrt region 1 is *left, back, back_left* and in a similar way, the fixed orientation of region 3 is described wrt all its *neighbours of level*. Finally, the relative orientation wrt the disjoint neighbours of region 3 is given: from region 0 to region 1, region 3 is located *left_front (lf)*; from region 0 to region 2, region 3 is located *back_right(br)*, from region 0 to region 6, region 3 is located *right_front (rf), right_middle (rm)* and so on. Note that the spatial description of region 5, also given in Table 7, is explained similarly.

The visual description of region 3 in Table 7 shows that its colour is *pale_yellow* and that the shape of its boundary is qualitatively described as composed of five *line-line* segments whose angles are: one *acute* and *convex*, one *right* and *convex* and three *obtuse* and *convex* and whose compared distances are *similar_lenght, similar_lenght, a_bit_longer, similar_length,* and *half_length*, respectively. The orientation of its vertices with respect to the centroid of the region is in a clockwise direction: *front, right, back, back_left, left*.

With respect to the computation time, it should be noted that, for the image shown in Fig. 12, the time of execution for the extraction of the main regions in the image is around 0.77 seconds and the time for generating the qualitative description of the image is around 0.89 seconds. The total time of execution is thus around 1.66 seconds using the same computer as in Scenario I.

The images of tile compositions captured by an industrial camera located above the table where a robot arm assembles tile mosa-



(a) Industrial camera located on a platform used by a robot arm

(b) Image obtained

(c) Canny's segmentation

Fig. 12. The QID approach applied to Scenario II: describing an image of a tile composition taken by an industrial camera located in a platform which is used by a robot arm to assemble tile mosaics.

Table 7
An excerpt of the qualitative description obtained for the mosaic image captured by the industrial camera in Fig. 12.

```
[SpatialDescription,
    (...)
  [3, [Container, Image], [Orientation wrt Image: left],
      [disjoint, 0, 1, 2, 4, 5, 6, 7],
      [Orientation wrt Neighbours: [0, left, back_left], [1, left, back, back_left] (... ) [7, front_left, left]],
      [Relative Orientation wrt Neighbours Disjoint: [[0, 1], lf], [[0, 2], br] (...) [[0, 6], rf, rm] (...)
      [[1, 2], br], [[1, 4], rm] (... )[[4, 6], rm, rf] (...) [[6, 7], bl, lm]]
    ]
  (...)
  [5, [Container, Image], [Orientation wrt Image: right, back_right, back],
      [disjoint, 0, 1, 2, 3, 4, 6, 7],
      [Orientation wrt Neighbours: [0, back, back_right], [1, back_right] (...) [6, right], [7, front_right, right]],
      [Relative Orientation wrt Neighbours Disjoint: [[0, 1], bl], [[0, 2], rm, rf], [[0, 3], lm, bl] (...) [[1, 2], rm, rf]
   (...)
      [[4, 6], bl, lm], [[4, 7], lm, lf], [[6, 7], lf, rf]]
    ]
    (...)
]
[VisualDescription,
  [3, pale_yellow,
      [Boundary_Shape,
          [line-line, acute, similar_lenght, convex],
          [line-line, right, similar_lenght, convex],
          [line-line, obtuse, a_bit_longer, convex],
          [line-line, obtuse, similar_lenght, convex],
          [line-line, obtuse, half_length, convex]],
      [Vertices_Orientation,front, right, back, back_left, left],],
    ],
      (...)
]
```

ics shown in Fig. 13 were used to test our approach. The results of segmentation of each image by Canny's method and the QID obtained are available on-line.[5] The execution times obtained in the tests are shown in Fig. 14, which has the same structure as previously described Fig. 11.

5.4. Analysis of the results

As the QID approach is a general approach aimed at extracting qualitative knowledge from any digital image, it can be applied in a lot of scenarios by adjusting its parameters and selecting the suitable image segmentation method.

The two testing scenarios shown in this paper prove the functionality and adaptability of our approach. There is not any predefined task that our approach has to solve in these scenarios. However, examples of possible applications can be given for clarification and for outlining possible future work. In Scenario I, the QID approach can help in localisation and object-recognition tasks after matching the qualitative descriptions obtained. And it can also help in the communication between a human user and a robot by obtaining the qualitative description produced in a narrative language form and reading it aloud by a speech synthesiser program. In Scenario II, on the other hand, the QID approach can help to find the similarity between two mosaics by solving the problem of matching of two qualitative descriptions of tile compositions, and it can also enhance human-machine communication by naming the visual features of the tile pieces currently located on the table.

From the charts shown in Figs. 11 and 14, note that the computation cost of our QID-Algorithm peaks when a lot of regions are

[5] http://dl.dropbox.com/u/17361913/CVIUTests.rar

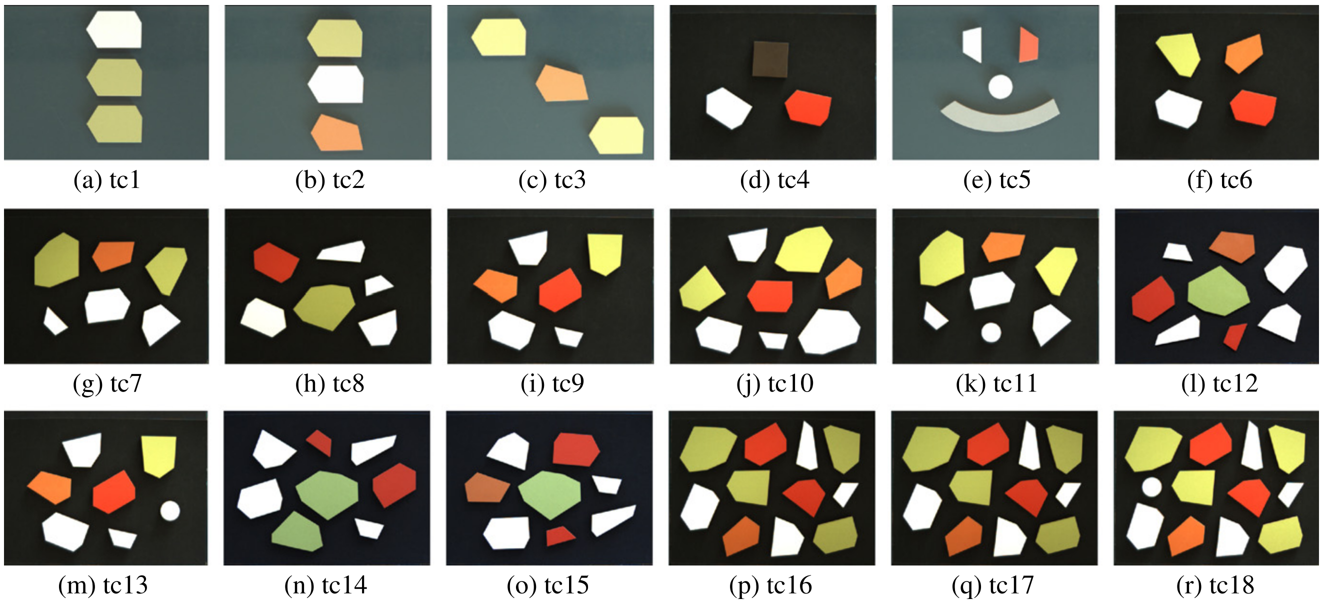


Fig. 13. Images of tile compositions (TC) used for testing our approach.

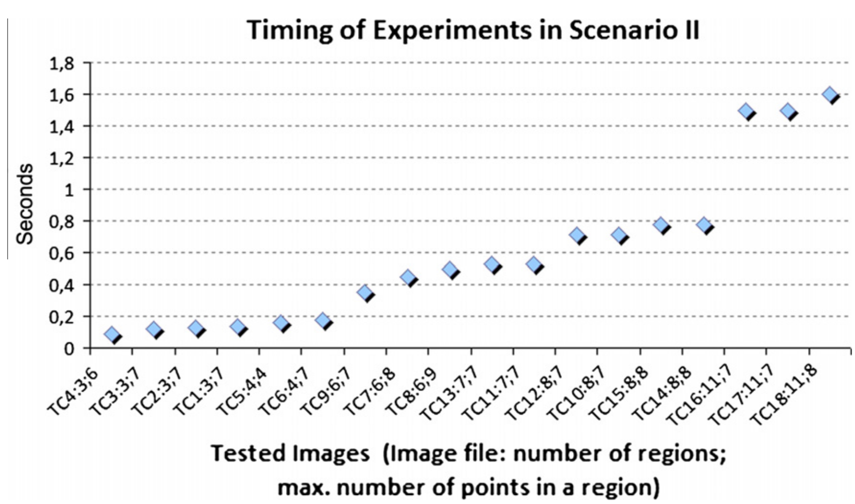


Fig. 14. Execution times of QID obtained for images in Fig. 13 related to the computational cost of QID-Algorithm for each situation.

extracted in the image and those regions have irregular boundaries described by a large number of relevant points. Note also that having more than 10 regions in an image significantly increases the time needed to describe that image. Furthermore, if an image has few regions but a very irregular shape with a high number of relevant points, the time obtained is more than that needed to describe an image with more regions and more regular shapes. Fig. 11 shows an example, where image *SC18* (with 9 regions and a maximum of 12 points defining a region) needs more time for the QID than image *SC2* with 11 regions and a maximum of 7 points. Moreover, the worst time obtained is less than 4 seconds for an image of more than 15 regions, but in images with few regions, like those in Scenario II, the time decreases a lot and the experiments are performed in less than 1.6 seconds.

6. Conclusions

A computational approach to the qualitative description of any digital image based on the visual and spatial features of all the characteristic regions/objects within the image is presented in this paper. This description was obtained from qualitative models of shape, colour, topology, and fixed and relative orientation.

The QID approach is not dependent on the region segmentation method used, and therefore, the most convenient segmentation method that obtains closed regions can be selected from the literature depending on the application. Moreover, the parameters of the QID approach can be adjusted by experts according to the selected application.

The approach presented here was applied to extract qualitative information from two different scenarios: (i) images of indoor environments, mainly visual landmarks, captured from a mobile robot camera in which a region-based image segmentation method is used; and (ii) images of tile compositions captured by an industrial camera located on a platform which is used by a robot arm to assemble mosaics automatically, in which a boundary-based image segmentation method is applied. In both cases, successful results were obtained with a low computational cost, thus showing the utility and flexibility of the QID approach.

As future work, we intend to apply the QID approach to solve tasks in both scenarios presented here by, for example: (i) matching the QID obtained to help in robot localisation and object recognition; (ii) enhancing human-machine communication by translating the produced QID into a natural language form; and (iii) measuring the similarity between two mosaics by solving the approximate matching of two QIDs of tile compositions in order to detect design plagiarisms. Moreover, other intended future applications are the following: (i) extracting meaning from images and improving the understanding of those images by web agents by translating the QID to a description logics-based ontology; (ii) obtaining a semantic similarity measure between the meaning of two ontological descriptions, where this similarity would calculate the resemblance between the different instances generated; (iii) measuring the similarity of two qualitative image descriptions from the point of view of human thinking; (iv) using that similarity measure for visual image/scene recognition and retrieval from a cognitive point of view (applicable to design automation processes, psychological research, and other fields involving imitation and study of human perception); and (v) applying the QID approach to other scenarios, such as, medical image description, geographical image description, and so on.

Acknowledgements

This work has been partially supported by Generalitat Valenciana (Grants BFPI06/219 and BFPI06/372), Universitat Jaume I – Fundació Bancaixa (Grant P11A2008-14) and Spanish MCyT (Grant TIN2009-14378-C02-01). We thank Pedro F. Felzenszwalb for making his source code available in his website and we thank also Pablo Ródenas and Jon Almazán for their support.

Appendix A

Tables for parameterizing qualitative colours in the HSL colour space.

Table 8
HSL intervals (QC_{INT_1}) for colour names in QC_{LAB_1}.

QC_{LAB_1}	UH	US	UL
black	[0, 360]	[0, 20]	[0, 20)
dark_grey	[0, 360]	[0, 20]	[20, 30)
grey	[0, 360]	[0, 20]	[30, 40)
light_grey	[0, 360]	[0, 20]	[40, 80)
white	[0, 360]	[0, 20]	[80, 100)

Table 9
HSL intervals (QC_{INT_2}) for colour names in QC_{LAB_2}.

QC_{LAB_2}	UH	US	UL
red	(335, 360]∧[0, 40]	(50, 100]	(55, 100]
yellow	(40, 80]	(50, 100]	(55, 100]
green	(80, 160]	(50, 100]	(55, 100]
turquoise	(160, 200]	(50, 100]	(55, 100]
blue	(200, 260]	(50, 100]	(55, 100]
purple	(260, 297]	(50, 100]	(55, 100]
pink	(297, 335]	(50, 100]	(55, 100]

Table 10
HSL intervals (QC_{INT_3}) for colour names in QC_{LAB_3}.

QC_{LAB_3}	UH	US	UL
pale_red	(335, 360]∧[0, 40]	(20, 50]	(40, 55]
pale_yellow	(40, 80]	(20, 50]	(40, 55]
pale_green	(80, 160]	(20, 50]	(40, 55]
pale_turquoise	(160, 200]	(20, 50]	(40, 55]
pale_blue	(200, 260]	(20, 50]	(40, 55]
pale_purple	(260, 297]	(20, 50]	(40, 55]
pale_pink	(297, 335]	(20, 50]	(40, 55]

Table 11
HSL intervals (QC_{INT_4}) for colour names in QC_{LAB_4}.

QC_{LAB_4}	UH	US	UL
light_red	(335, 360]∧[0, 40]	(50, 100]	(55, 100]
light_yellow	(40, 80]	(50, 100]	(55, 100]
light_green	(80, 160]	(50, 100]	(55, 100]
light_turquoise	(160, 200]	(50, 100]	(55, 100]
light_blue	(200, 260]	(50, 100]	(55, 100]
light_purple	(260, 297]	(50, 100]	(55, 100]
light_pink	(297, 335]	(50, 100]	(55, 100]

Table 12
HSL intervals (QC_{INT_5}) for colour names in QC_{LAB_5}.

QC_{LAB_5}	UH	US	UL
dark_red	(335, 360]∧[0, 40])	(50, 100]	(20, 40])
dark_yellow	(40, 80]	(50, 100]	(20, 40]
dark_green	(80, 160]	(50, 100]	(20, 40]
dark_turquoise	(160, 200]	(50, 100]	(20, 40]
dark_blue	(200, 260]	(50, 100]	(20, 40]
dark_purple	(260, 297]	(50, 100]	(20, 40]
dark_pink	(297, 335]	(50, 100]	(20, 40]

References

[1] C. Jörgensen, Attributes of images in describing tasks, Information Processing Management: An International Journal 34 (2–3) (1998) 161–174.
[2] M. Laine-Hernandez, S. Westman, Image semantics in the description and categorization of journalistic photographs. in: A. Grove, J. Stefl-Mabry (Eds.), Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology, vol. 43, 2006, pp. 1–25.
[3] H. Greisdorf, B. O'Connor, Modelling what users see when they look at images: a cognitive viewpoint, Journal of Documentation 58 (1) (2002) 6–29.
[4] X. Wang, P. Matsakis, L. Trick, B. Nonnecke, M. Veltman, A study on how humans describe relative positions of image objects, in: Lecture Notes in Geoinformation and Cartography, Headway in Spatial Data Handling, Springer, Berlin, Heidelberg, 2008, ISBN 978-3-540-68565-4, pp. 1–18. ISSN:1863-2246.
[5] S. Palmer, Vision Science: Photons to Phenomenology, MIT Press, 1999.
[6] C. Freksa, Qualitative spatial reasoning, in: D.M. Mark, A.U. Frank (Eds.), Cognitive and Linguistic Aspects of Geographic Space, NATO Advanced Studies Institute, Kluwer, Dordrecht, 1991, pp. 361–372.
[7] S.M. Kosslyn, Image and Brain: The Resolution of The Imagery Debate, MIT Press, Cambridge, MA, USA, 1994.
[8] S.M. Kosslyn, W.L. Thompson, G. Ganis, The Case for Mental Imagery, Oxford University Press, New York, USA, 2006.
[9] A. Ramisa, Localization and object recognition for mobile robots, Ph.D. thesis, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, SPAIN, supervised by Prof. Ramon López de Mántaras and Dr. Ricardo Toledo, 2009.
[10] A. Quattoni, A. Torralba, Recognizing indoor scenes, IEEE Computer Society Conference on Computer Vision and Pattern Recognition 0 (2009) 413–420.
[11] J. Wu, H.I. Christensen, J.M. Rehg, Visual place categorization: problem, dataset, and algorithm, in: Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'09, IEEE Press, Piscataway, NJ, USA, 2009, pp. 4763–4770.
[12] C. Liu, J. Yuen, A. Torralba, SIFT flow: dense correspondence across scenes and its applications, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (5) (2011) 978–994.
[13] M.-A. Williams, Representation = grounded information, in: PRICAI '08: Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence, LNCS, vol. 5351, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 473–484.
[14] M.J. Egenhofer, R. Franzosa, Point-set topological spatial relations, International Journal of Geographical Information Systems 5 (2) (1991) 161–174.
[15] D. Hernández, Relative representation of spatial knowledge: the 2-D case, in: D.M. Mark, A.U. Frank (Eds.), Cognitive and Linguistic Aspects of Geographic Space, NATO Advanced Studies Institute, Kluwer, Dordrecht, 1991, pp. 373–385.
[16] C. Freksa, Using orientation information for qualitative spatial reasoning, in: A.U. Frank, I. Campari, U. Formentini (Eds.), Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, LNCS, vol. 639, Springer, Berlin, 1992, pp. 162–178.
[17] G. Socher, G. Sagerer, P. Perona, Bayesian reasoning on qualitative descriptions from images and speech, Image and Vision Computing 18 (2) (2000) 155–172.
[18] A. Lovett, M. Dehghani, K. Forbus, Efficient learning of qualitative descriptions for sketch recognition, in: 20th International Workshop on Qualitative Reasoning, 2006.
[19] Z.U. Qayyum, A.G. Cohn, Image retrieval through qualitative representations over semantic features, in: BMVC, 2007.

[20] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, International Journal of Computer Vision 42 (3) (2001) 145–175.
[21] R.A. Wilson, F.C. Keil (Eds.), The MIT Encyclopedia of the Cognitive Sciences, The MIT Press, Cambridge, Massachusetts, 1999.
[22] Z. Falomir, L. Gonzalez-Abril, L. Museros, J. Ortega, Measures of similarity between objects from a qualitative shape description, Spatial Cognition and Computation, in press.
[23] D. Conway, An experimental comparison of three natural language colour naming models, in: Proceedings of East-West International Conference on Human–Computer Interactions, 1992, pp. 328–339.
[24] G. Menegaz, A.L. Troter, J. Sequeira, J.M. Boi, A discrete model for color naming, EURASIP Journal on Applied Signal Processing 1 (2007) 1–10. Special Issue on Image Perception 2007.
[25] J.V. de Weijer, C. Schmid, Applying color names to image description, in: ICIP (3), IEEE, 2007, pp. 493–496.
[26] R. Benavente, M. Vanrell, R. Baldrich, Parametric fuzzy sets for automatic color naming, Journal of The Optical Society of America A – Optics Image Science and Vision 25 (10) (2008) 2582–2593.
[27] A. Mojsilovic, A computational model for color naming and describing color composition of images, IEEE Transactions on Image Processing 14 (5) (2005) 690–699.
[28] M. Seaborn, L. Hepplewhite, T.J. Stonham, Fuzzy colour category map for the measurement of colour similarity and dissimilarity, Pattern Recognition 38 (2) (2005) 165–177.
[29] Y. Liu, D. Zhang, G. Lu, W. Ying Ma, Region-based image retrieval with perceptual colors, in: Proceedings of Pacific-Rim Multimedia Conference (PCM2004), 2004, pp. 931–938.
[30] M. Sarifuddin, R. Missaoui, A new perceptually uniform color space with associated color similarity measure for contentn++based image and video retrieval, in: Multimedia Information Retrieval Workshop, 28th Annual ACM SIGIR Conference, 2005, pp. 3–7.
[31] M. Egenhofer, A formal definition of binary topological relationships, in: 3rd International Conference on Foundations of Data Organization and Algorithms, FODO 1989, Lecture Notes in Computer Science, vol. 367, Springer-Verlag, 1989, pp. 457–472.
[32] C. Freksa, Using orientation information for qualitative spatial reasoning, in: A.U. Frank, I. Campari, U. Formentini (Eds.), Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. International Conference on GIS—From Space to Territory, Lecture Notes in Computer Science, vol. 639, Springer, Berlin, 1992, pp. 162–178.
[33] J.F. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 8 (1986) 679–697.
[34] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, International Journal of Computer Vision 59 (2) (2004) 167–181.
[35] L. Gonzalez-Abril, F. Cuberos, F. Velasco, J. Ortega, Ameva: an autonomous discretization algorithm, Expert Systems with Applications 36 (3, Part 1) (2009) 5327–5332.