# Qualitative Distances and Qualitative Description of Images for Indoor Scene Description and Recognition in Robotics

Zoe Falomir Llansola

A thesis submitted in partial fulfillment of the requirements of University Jaume I for the Dr. degree and of the University of Bremen for the Dr.-Ing. degree.

Directors:

Dr. M. Teresa Escrig Monferrer, *Cognition for Robotics Research* (C4R2), Engineering and Computer Science Department, Universitat Jaume I, and Cognitive Robots (C-Robots) SL, Castellón, Spain.

Prof. Dr. Christian Freksa, Cognitive Systems (CoSy), University of Bremen, Germany.

Figueroles (Spain), May 13, 2011

*Als meus pares Maribel i Rafael,*
*per criar-me i educar-me amb tanta estima,*
*al meu germà Juan,*
*per qué hem jugat, rist i aprés junts,*
*i a tots aquells que m'estimen*
*per compartir la seua vida amb mi.*

# Abstract

The automatic extraction of knowledge from the world by a robotic system as human beings interpret their environment through their senses is still an unsolved task in Artificial Intelligence. A robotic agent is in contact with the world through its sensors and other electronic components which obtain and process mainly numerical information. Sonar, infrared and laser sensors obtain distance information. Webcams obtain digital images that are represented internally as matrices of red, blue and green (RGB) colour coordinate values. All this numerical values obtained from the environment need a later interpretation in order to provide the knowledge required by the robotic agent in order to carry out a task.

Similarly, light wavelengths with specific amplitude are captured by cone cells of human eyes obtaining also stimulus without meaning. However, the information that human beings can describe and remember from what they see is expressed using words, that is qualitatively. The exact process carried out after our eyes perceive light wavelengths and our brain interpret them is quite unknown. However, 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.

As the information provided by all the electronic components incorporated in a robotic agent is numerical, the approaches that first appeared in the literature giving an interpretation of this information followed a mathematical trend. In this thesis, this problem is addressed from the other side, its main aim is to process these numerical data in order to obtain qualitative information as human beings can do.

The research work done in this thesis tries to narrow the gap between the acquisition of low level information by robot sensors and the need of obtaining high level or qualitative information for enhancing human-machine communication and for applying logical reasoning processes based on concepts. Moreover, qualitative concepts can be added a meaning by relating them to others. And they can be used for reasoning applying qualitative models that have been developed in the last twenty years for describing and interpreting metrical and mathematical concepts such as orientation, distance, velocity, acceleration, and so on. And they can be understood by human-users both written and read aloud.

The first contributions presented are the definition of a method for obtaining fuzzy distance patterns (which include qualitative distances such as *near*, *far*, *very far* and so on) from the data obtained by any kind of distance sensors incorporated in a mobile robot and the definition of a factor to measure the dissimilarity between those fuzzy patterns. Both have been applied to the

integration of the distances obtained by the sonar and laser distance sensors incorporated in a Pioneer 2 dx mobile robot and, as a result, special obstacles have been detected as *glass window*, *mirror*, and so on. Moreover, the fuzzy distance patterns provided have been also defuzzified in order to obtain a smooth robot speed and used to classify orientation reference systems into *open* (it defines an open space to be explored) or *closed*.

The second contribution presented is the definition of a model for qualitative image description (QID) by applying the new defined models for qualitative shape and colour description, the topology model by Egenhofer and Al-Taha [1992] and the orientation models by Hernández [1991] and Freksa [1992]. This model can qualitatively describe any kind of digital image and is independent of the image segmentation method used. The QID model have been tested in two scenarios in robotics: (i) the description of digital images captured by the camera of a Pioneer 2 dx mobile robot and (ii) the description of digital images of tile mosaics taken by an industrial camera located on a platform used by a robot arm to assemble tile mosaics.

In order to provide a formal and explicit meaning to the qualitative description of images generated, a Description Logic (DL) based ontology has been designed and presented as the third contribution. Our approach can automatically process any random image and obtain a set of DL-axioms that describe it visually and spatially. And objects included in the images are classified according to the ontology schema and using a DL reasoner. Tests have been carried out using digital images captured by a webcam incorporated in a Pioneer 2 dx mobile robot. The images taken correspond to the corridors of a building at University Jaume I and objects within them have been classified into *walls*, *floor*, *office doors* and *fire extinguishers*, under different illumination conditions and from different observer viewpoints.

The final contribution is the definition of a similarity measure between qualitative descriptions of shape, colour, topology and orientation. And the integration of those measures into the definition of a general similarity measure between two qualitative descriptions of images. These similarity measures have been applied to: (i) extract objects with similar shapes from the MPEG7 CE Shape-1 library; (ii) assemble tile mosaics by qualitative shape and colour similarity matching; (iii) compare images of tile compositions; and (iv) compare images of natural landmarks in a mobile robot world for their recognition.

The contributions made in this thesis are only a small step forward in the direction of enhancing robot knowledge acquisition from the world. And it is also written with the aim of inspiring others in their research, so that bigger contributions can be achieved in the future which can improve the life quality of our society.

**Keywords:** *Sensor data integration; Fuzzy set theory; Qualitative representation and modelling; Image segmentation; Qualitative shape; Qualitative colour; Qualitative orientation; Qualitative topology; Ontology; Description Logics; Similarity; Conceptual Neighbourhood Diagrams; Interval Distances; Robotics.*

# Resümee

Die Erzielung einer automatischen Wissensextraktion der Roboterumgebung auf eine ähnliche Weise, wie wir Menschen die Umwelt mit unseren Sinnen erfassen, ist eine noch unerledigte Aufgabe der Künstlichen Intelligenz. Der Kontakt der Roboter mit ihrer Umgebung vollzieht sich anhand ihrer Sensoren und weiterer elektronischer Elemente, die vor allem nummerische Daten erfassen und verarbeiten. Die Sonar-, Laser- und Infrarotsensoren ermitteln Entfernungsdaten. Die Digitalkameras erzielen Aufnahmen, die vom System als Koordinatennetz in Rot, Grün und Blau (RGB) dargestellt werden. Alle diese nummerischen Daten, die der Umgebung entnommen werden, müssen anschlie$\beta$end interpretiert werden, um den Roboter mit dem zur Ausübung einer bestimmten Aufgabe erforderlichen Wissen zu versorgen.

Ganz ähnlich werden Lichtwellen mit einer bestimmten Amplitude von den Zapfenzellen im menschlichen Auge erfasst und sorgen für neutrale Reize. Allerdings drücken Personen die Informationen, mit denen sie beschreiben oder sich ins Gedächtnis rufen, was sie sehen oder gesehen haben, mit Worten - also qualitativ - aus. Der Prozess, der in Gang gesetzt wird, sobald Lichtwellen von unseren Augen aufgefangen und von unserem Gehirn verarbeitet werden, ist nicht genau bekannt. Es gilt aber als Tatsache der menschlichen Kognition, dass wir über die blo$\beta$e Erfahrung der Wahrnehmung hinaus Dinge in Kategorien ordnen und ihnen ein sprachliches Etikett zuteilen.

Da es sich bei den Informationen, welche die Elektronikelemente eines Roboters übermitteln, um nummerische Daten handelt, folgten die ersten Methoden in der Fachliteratur mathematischen Approximationen. In der vorliegenden Dissertation wird das Problem aus einer anderen Perspektive behandelt, die vor allem darauf abzielt, diese nummerischen Werte zu verarbeiten, um qualitative Informationen zu erhalten, also genau so, wie wenn wir die mit unseren Sinnen erfassten Informationen verarbeiten.

Die Forschungsarbeit im Rahmen dieser Dissertation versucht, die Lücke zwischen der Erfassung minderwertiger Informationen durch Robotersensoren und dem Bedarf an hochwertigen bzw. qualitativen Informationen zu verringern, um die Kommunikation zwischen Mensch und Maschine zu verbessern und konzeptbasierte logische Schlussfolgerungen zu ermöglichen. Au$\beta$erdem kann man den erzielten Qualitätskonzepten eine Bedeutung zuordnen und diese durch Verknüpfung mit anderen Konzepten optimieren. Die genannten Konzepte können auch für Schlussfolgerungen anhand von Qualitätsmodellen verwendet werden, die in den letzten zwanzig Jahren entwickelt wurden, um metrische und mathematische Konzepte wie Orientierung, Entfernung, Geschwindigkeit, Beschleunigung, usw. zu beschreiben und zu interpretieren. Qualitative Informationen werden ferner von den Menschen sowohl in schriftlicher als auch in

gesprochener Form verstanden.

Die ersten der in dieser Dissertation präsentierten Beiträge sind die Bestimmung einer Methode zur Erzielung von Mustern für Fuzzy-Abstände (darunter qualitative Abstandsinformationen wie nah, weit entfernt, sehr weit entfernt, usw.) auf der Grundlage von Daten, die wiederum mit einem beliebigen Abstandssensor eines Roboters erfasst wurden, sowie die Bestimmung eines Faktors zur Messung der Unähnlichkeit dieser Muster. Beide Beiträge wurden bei der Integrierung der Abstände angewendet, welche von den Sonar- und Lasersensoren eines Roboters Pioneer 2 dx ermittelt wurden. Infolgedessen konnten besondere Hindernisse wie Glasscheiben, Spiegel, usw. entdeckt werden. Die ermittelten Muster fr Fuzzy-Abstände wurden in eine geringe Geschwindigkeit bei der Roboter-Annäherung an Hindernisse übertragen und zur Unterscheidung von Orientierungssystemen in offen (die näher erkundet werden können) und geschlossen (in denen der Roboter keine weiteren Informationen zur Lokalisierung und Orientierung erfassen kann) genutzt.

Der zweite Beitrag ist die Definition eines Modells zur qualitativen Bildbeschreibung anhand der Anwendung neuer Qualitätsmodelle zur Form- und Farbbeschreibung sowie der topologischen Modelle von Egenhofer und Al-Taha [1992] und der Modelle zur qualitativen Orientierung von Hernández [1991] und Freksa [1992]. Dieses Modell kann ein beliebiges digitales Bild qualitativ beschreiben und ist unabhängig von der verwendeten Segmentierungsmethode. Au$\beta$erdem wurde das Modell in zwei echten Arbeitsszenarien mit Robotern getestet: (i) zur Beschreibung digitaler Aufnahmen, die von der Kamera eines mobilen Roboters Pioneer 2 dx stammen; und (ii) zur Beschreibung digitaler Aufnahmen von Fliesenmustern, die eine Industriekamera auf einer Plattform erstellt, von der ein Roboterarm Keramikteile zur Zusammenstellung von Mosaiken nimmt.

Als dritter Beitrag wurde eine auf Beschreibungslogik basierende Ontologie aufgebaut, um der qualitativen Bildbeschreibung eine förmliche, ausdrückliche Bedeutung zu geben. Die erzielte Methode kann automatisch jedes beliebige Bild verarbeiten und einen Satz beschreibungslogischer Axiome erzielen, die das Bild optisch und räumlich beschreiben. Die im Bild enthaltenen Objekte werden dem ontologischen Schema entsprechend mithilfe eines Beschreibungslogik-Reasoners klassifiziert. Zur Bewertung der erzielten Methode wurden Tests mit Aufnahmen eines mobilen Roboters Pioneer 2 dx durchgeführt, der sich durch die Flure eines Gebäudes der Universität Jaume I bewegte. Dabei konnten die auf diesen Aufnahmen erfassten Objekte bei unterschiedlichen Beleuchtungsbedingungen und unter verschiedenen Gesichtspunkten des Betrachters als Wände, Boden, Bürotüren und Feuerlöscher klassifiziert werden.

Der letzte Beitrag der Dissertation besteht aus der Definition eines Ma$\beta$es für die Ähnlichkeit zwischen qualitativen Form-, Farb-, Topologie- und Orientierungsbeschreibungen und der Eingliederung dieser Ma$\beta$e in die Definition eines allgemeinen Ähnlichkeitsma$\beta$es zwischen zwei qualitativen Bildbeschreibungen. Diese Ähnlichkeitsma$\beta$e wurden angewandt auf: (i) die Extraktion von Objekte aus dem MPEG7-CE-Shape-1-Datensatz; (ii) die Zusammensetzung von Mosaiken anhand von Fliesenübereinstimmungen, die auf Ähnlichkeitsma$\beta$en für qualitative Form- und Farbbeschreibungen basieren; (iii) den Vergleich und die Klassifizierung von Aufnahmen von Keramiksätzen; und (iv) den Vergleich von Aufnahmen von Fluren und Ecken zur Identifizierung von Bezugspunkten zur Lokalisierung und Orientierung unseres Roboters.

Die Beiträge im Rahmen dieser Dissertation sind nur ein kleiner Fortschritt

hin zu einem besseren Wissenserwerb von Robotern in Bezug auf ihre Umgebung. Sie sollen andere Wissenschaftler bei ihren Forschungen inspirieren und künftig grö$\beta$ere Beiträge ermöglichen, um die Lebensqualität unserer Gesellschaft zu verbessern.

**Stichwörter:** Integrierung von Sensorendaten; Fuzzy-Set-Theorie; Qualitative Darstellung und Modellierung; Bildsegmentierung; Qualitative Form; Qualitative Farbe; Qualitative Orientierung; Qualitative Topologie; Ontologien; Beschreibungslogik; Ähnlichkeit; Diagramme konzeptueller Nachbarschaften; Intervallabstände; Robotik.

# Resumen

La extracción automática de conocimiento del entorno de un robot realizada de forma similar a cómo los humanos interpretan el mundo a través de sus sentidos es una tarea que aún no está resuelta en Inteligencia Artificial. Los robots están en contacto con su entorno a través de sus sensores y otros dispositivos electrónicos que obtienen y procesan principalmente información numérica. Los sensores sónar, láser o infrarrojos obtienen valores de distancias. Las cámaras digitales obtienen imágenes que son representadas por el sistema como matrices de coordenadas de color rojo, azul y verde (RGB en inglés). Toda esta información numérica extraída del entorno necesita una interpretación posterior para proporcionar al robot conocimiento necesario para que éste pueda realizar una tarea encomendada.

De forma similar, ondas de luz de una determinada amplitud son capturadas por las células cono del ojo humano, que obtiene estímulos sin significado. No obstante, la información que las personas pueden describir y recordar acerca de lo que ven o han visto la expresan con palabras, es decir, de forma cualitativa. El procedimiento que se produce después de que nuestros ojos perciban las ondas de luz y nuestro cerebro las interprete no se conoce con exactitud. Sin embargo, un hecho real en la cognición humana es que las personas van más allá de la experiencia de percibir para clasificar cosas como miembros de categorías y asignarles una etiqueta lingüística.

Como la información proporcionada por los componentes electrónicos incorporados en un robot es numérica, los métodos que primero aparecieron en la literatura para dar interpretación a dicha información siguieron aproximaciones matemáticas. En esta tesis, este problema se aborda desde otra perspectiva, el motivo principal de la cual es procesar esos valores numéricos para obtener información cualitativa, al igual que las personas hacemos cuando procesamos información capturada por nuestros sentidos.

El trabajo de investigación realizado en esta tesis intenta cerrar la brecha existente entre la adquisición de información de bajo nivel realizada por los sensores de un robot y la necesidad de información de alto nivel o cualitativa para mejorar la comunicación hombre-máquina y para aplicar el razonamiento lógico basado en conceptos. Además, se puede asignar un significado a los conceptos cualitativos obtenidos y se puede incrementar éste relacionándolos con otros. Dichos conceptos también pueden ser usados para razonar aplicando modelos cualitativos que se han desarrollado durante los últimos veinte años para describir e interpretar conceptos métricos y matemáticos como la orientación, la distancia, la velocidad, la aceleración, etc. Además, la información cualitativa es entendida por las personas tanto si es escrita como reproducida oralmente.

Las primeras contribuciones presentadas en esta tesis son la definición de un

método para obtener patrones de distancias difusas (que extraen información cualitativa de distancias como *cerca*, *lejos*, *muy lejos*, etc.) a partir de datos obtenidos por cualquier tipo de sensor de distancia incorporado en un robot móvil y la definición de un factor para medir la disimilitud entre dichos patrones. Ambos han sido aplicados a la integración de las distancias proporcionadas por los sensores sónar y láser de un robot Pioneer 2 dx, y como resultado, se han podido detectar obstáculos especiales como *cristal*, *espejo*, etc. Los patrones de distancias difusas proporcionados se han traducido en velocidades suaves de aproximación del robot a obstáculos y utilizado para clasificar sistemas de referencia de orientación en *abiertos* (que se pueden explorar más exhaustivamente) o *cerrados* (donde el robot ya no puede encontrar más información para localizarse y orientarse).

La segunda contribución que se presenta es la definición de un modelo para la descripción cualitativa de cualquier imagen mediante la aplicación de nuevos modelos cualitativos de descripción de formas y color y los modelos de topología de Egenhofer and Al-Taha [1992] y los modelos de orientación cualitativa de Hernández [1991] y Freksa [1992]. El modelo presentado puede describir cualitativamente cualquier imagen digital y es independiente del método de segmentación utilizado. Además, dicho modelo ha sido probado en dos escenarios de trabajo reales en donde se incluyen robots: (i) la descripción de imágenes digitales capturadas por la cámara de un robot móvil Pioneer 2 dx, y (ii) la descripción de imágenes digitales de composiciones de azulejos tomadas por una cámara industrial colocada sobre una plataforma desde donde un brazo robótico coge piezas cerámicas para montar mosaicos.

Como tercera contribución, se ha construido una ontología basada en lógica descriptiva para dar un significado formal y explícito a la descripción cualitativa de cualquier imagen. El método obtenido puede procesar automáticamente cualquier imagen y obtener un conjunto de axiomas basados en lógica descriptiva que describen visualmente y espacialmente dicha imagen. Los objetos incluidos en las imágenes son clasificados de acuerdo con el esquema de la ontología utilizando para ello un razonador de lógica descriptiva. Para evaluar el método obtenido se han realizado pruebas con las imágenes obtenidas por una cámara situada en un robot móvil Pioneer 2 dx mientras éste navegaba por los pasillos de un edificio de la Universitat Jaume I y se ha logrado clasificar objetos detectados en dichas imágenes como *paredes*, *suelo*, *puertas de despachos* y *extintores*, bajo diferentes condiciones de iluminación y diferentes puntos de vista del observador.

Finalmente, la última contribución de esta tesis es la definición de una medida de similitud entre descripciones cualitativas de forma, color, topología y orientación, y la integración de dichas medidas en una medida de similitud general entre dos descripciones cualitativas de imágenes. Dichas medidas de similitud han sido aplicadas en: (i) la comparación de objetos de la librería de reconocimiento de formas MPEG7 CE Shape-1; (ii) el ensamblado de mosaicos utilizando correspondencias de azulejos basados en medidas de similitud entre descripciones cualitativas de forma y color; (iii) la comparación y clasificación de imágenes de composiciones de piezas cerámicas; y en (iv) la comparación de imágenes de pasillos y esquinas para la identificación de puntos de referencia para la localización y orientación de nuestro robot.

Las contribuciones realizadas en esta tesis son sólo un pequeño paso para mejorar la adquisición de conocimiento del entorno de un robot y han sido

escritas con la intención de inspirar a otros en sus investigaciones, para que mayores contribuciones se puedan alcanzar en el futuro que mejoren la calidad de vida de la sociedad.

**Palabras clave**: *Integración de datos de sensores; Teoría de conjuntos difusos; Representación y modelado cualitativo; Segmentación de imágenes; Forma cualitativa; Color cualitativo; Orientación cualitativa; Topología cualitativa; Ontologías; Lógica descriptiva; Similitud; Diagramas de vecindad conceptual; Distancias intervalares; Robótica.*

# Acknowledgements

After five years of research this dissertation has been finished successfully. Many people have been working with me during this time giving their professional and also their personal support. Many thanks to all of you.

Després de cinc anys d'investigació aquesta tesi ha vist la llum. Molta gent ha estat al meu costat durant aquest temps, recolzant-me amb el seu suport tant professional com personal. A tots, us estic molt agraïda.

Después de cinco años de investigación esta tesis es una realidad. Mucha gente ha compartido experiencias conmigo durante este tiempo, apoyándome tanto profesionalmente como personalmente. A todos, mi más sincero agradecimiento.

- A Tere Escrig per iniciar-me en el món de la investigació, i també per les seues valioses idees i la seua orientació al llarg d'aquesta tesi.
- Many thanks to Christian Freksa for giving me the opportunity of working with the Cognitive Systems (CoSy) research group at Bremen University. I am very grateful for his time and comments on my work during my stay in Bremen, and also for his direction and revisions to this thesis. I would also like to thank all the members of CoSy for their suggestions, which greatly improved this thesis. I really appreciate their warm welcome and all their help and support during my stay in Germany.
- A Lledó Museros per estar sempre ahí per ajudar-me i acompanyar-me en aquest camí, entre altres al meu primer congrés i a la meua estada en Sevilla.
- Als meus companys del *Cognition for Robotics Research Group (C4R2)* i de *Cognitive Robots (C-Robots) SL*: Vicent Castelló, Jorge Grande, Pablo Ródenas, Juan Carlos Peris. Per la vostra genialitat, bona disposició i ajuda en tants treballs que hem fet junts i que no haurien segut possible sense vosaltres, i per totes les rises i anècdotes que hem viscut dins i fora de l'UJI.
- A Ernesto Jiménez y al resto del grupo *Temporal Knowledge Base Group* por su valiosa colaboración en nuestro trabajo sobre ontologías.
- A Luís González por todas sus cuidadosas revisiones y aportaciones a esta tesis y por enseñarme a difundir todo nuestro trabajo en revistas. A Juan Antonio Ortega y a Paco Velasco por acogerme en su grupo de

investigación y hacer tan enriquecedora mi estancia en Sevilla. También a mis compañeros del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla por su hospitalidad y simpatía.

- Many thanks to Stefan Wölfl from Universität Freiburg, Germany, and Eliseo Clementini from University l'Aquila, Italy, for taking the time to carefully review this thesis and help me to improve it.

- A los compañeros habituales de las *Jornadas de la Asociación de Razonamiento Cualitativo y Apliaciones (JARCA)* y del *Congrés Català d'Intel.ligència Artificial (CCIA)* que con sus sugerencias han inspirado muchas mejoras a esta tesis y con los que he compartido muy buenas experiencias.

- A tots els meus companys i companyes de la Universitat Jaume I, especialment als membres del Departament d'Enginyeria i Ciència dels Computadors, per recolzar aquesta tesi en tot moment. També a les meues companyes de camí, Maria Arregui i Anna Puig, per tantes coses que hem viscut durant aquest temps. Als meus alumnes de projectes i d'estades en pràctiques, companys i amics, especialment a Jon Almazán, William Viana, Higinio Martí i Isabel Martí per posar il·lusió i ganes en els seus treballs que al final han acabat formant part d'aquesta tesi.

- A tota la meua família per estar sempre ahí, creure en mi, recolzar-me i animar-me sempre en tot, i als que ja no estan, per tot el que ens van ensenyar, perqué nosaltres som perqué elles i ells van ser abans.

- I per últim, però no menys importants, a totes les meues amigues, als meus amics i al meu sol, gràcies per acompanyar-me en els mal de caps i en les rises durant tot aquest temps. Però sobretot gràcies pels bons moments que hem passat junts, i que en vinguen de molt millors!!

Finally, this work has been funded by: (1) Generalitat Valenciana under research grant numbers BFPI06/219, BEFPI/2008/027, BEFPI/2009/024; (2) the Ministerio de Educación y Ciencia, under project TIN 2006-14939 titled *Application of cognitive processes to robotics*; (3) the European Commission under FEDER funds for *Using cognitive techniques to obtain full scans of any plane surface*; (4) Fundació Bancaixa - Universitat Jaume I, under project P11A2008-14 titled *Developing and Applying cognitive theories for object recognition*; and (5) the Departament d'Enginyeria i Ciència dels Computadors at Universitat Jaume I (Fons del Pla Estratègic de 2009/2010 i 2010/2011). We also acknowledge collaboration with the interdisciplinary Transregional Collaborative Research Center Spatial Cognition SFB/TR 8 Project R3-[Q-Shape] at Universität Bremen.

# Contents

# List of Figures

# List of Tables

# Glossary

| | |
|---|---|
| $A$ | Angle |
| $C$ | Convexity |
| $CC$ | Computational Cost |
| $CND$ | Conceptual Neighbourhood Diagram |
| $CWA$ | Closed World Assumption |
| $DF$ | Dissimilarity Factor |
| $DL$ | Description Logics |
| $FDP$ | Fuzzy Distance Pattern |
| $FdSet$ | Fuzzy distance Set |
| $FE$ | Features Extration |
| $FO$ | Fixed Orientation |
| $HSL$ | Hue Saturation Lightness |
| $IC$ | Integrity Constraints |
| $KEC$ | Kind of Edges Connected |
| $L$ | Length |
| $OWL$ | Ontology Web Language |
| $OWA$ | Open World Assumption |
| $QCD$ | Qualitative Colour Description |
| $QID$ | Qualitative Image Description |
| $QSD$ | Qualitative Shape Description |
| $RO$ | Relative Orientation |
| $RP$ | Relevant Point |
| $RS$ | Reference System |
| $SimFO$ | Similarity between Fixed Orientations |
| $SimQCD$ | Similarity between Qualitative Colour Descriptions |
| $SimQID$ | Similarity between Qualitative Image Descriptions |
| $SimQSD$ | Similarity between Qualitative Shape Descriptions |
| $SimTop$ | Similarity between Topology relations |
| $TC$ | Type of Curvature |

# Chapter 1

# Introduction

Human beings use references to orientate themselves when they walk through a city. For example, when we think that *the supermarket is opposite the church*, the church is the reference for locating the supermarket, so when we recognize the church we are able to find the supermarket. In the same way, mobile robots use references of place or *landmarks* to locate themselves in their world and to orientate themselves and navigate towards a target.

In mobile robotics there are two main approaches to robot localization and navigation. The first one is to adapt the world for robots to navigate through it, that is, using *artificial landmarks* designed for easy recognition by the robot sensors (such as bar codes or radio frequency tags) placed in the world where the robot will need them. The second is to provide the robot with methods for detecting and recognizing *natural landmarks* in the world, that is, references similar to the ones used by human beings (such as signs, buildings, etc.). This approach is the more cognitive of the two; however, there are still open research problems endeavouring to achieve a robust and meaningful detection and interpretation of *natural landmarks*. The work presented in this thesis is a small step in this direction and it focuses on describing, detecting and recognizing *natural landmarks* in the robot world that will be used in the future for indoor navigation by applying qualitative orientation methods.

The following sections present the context and motivation of this thesis, its specific objectives achieved and its main contributions. Finally an outline of the chapters of this thesis is presented.

## 1.1 Context and Motivation of this Thesis

This work has been conducted in the group *Cognition for Robotics Research (C4R2)* and the spin-off *Cognitive Robots (C-Robots) SL* both located at University Jaume I and in the *Cognitive Systems (CoSy)* research group at Bremen University. Moreover, important contributions were made during my research stay with the *IDINFOR (Investigación, Desarrollo e Innovación en Informática)* research group at Seville University.

The main aim of both C4R2 and C-Robots is to apply qualitative reasoning models to real robot navigation and manipulation in order to provide robotics with human-like or cognitive functionalities. The main objective of this reseach

originated in the work of my supervisor Dr. M. Teresa Escrig [Escrig and Toledo, 2000] in applying the Double Cross Calculus (DCC) orientation model of my supervisor Prof. Dr. Christian Freksa [Freksa, 1992] combined with a constraint-based reasoning system to robot navigation in a simulated environment.

Other approaches in the literature also applied orientation reasoning techniques to simulated robot navigation. In the work by Schlieder [1995], the Dipole Calculus (DC) was applied to robot navigation in a QUALNAV simulated environment. And more recent work by Wolter et al. [2007] and Dylla et al. [2007] used the Oriented Point Relation Algebra ($\mathcal{OPRA}_m$) to regulate the traffic between artificial agents in a simulated sea navigation environment.

All these works succeeded in simulated environments where the locations and appearances of all the landmarks were known exactly, therefore, efforts are now being made to apply them to real environments.

In the literature there is also some research work that has applied qualitative orientation models to real robot navigation [Peris and Escrig, 2005; Frommberger and Wolter, 2008]. Peris and Escrig [2005] apply the Double Cross Calculus (DCC) orientation approach by Freksa [1992] to build a topological map and use it in robot navigation in real environments. And in the work by Frommberger and Wolter [2008], robot navigation is based on the Relative Line Position Representation (RLPR), which was inspired by the direction relation matrix by Goyal and Egenhofer [2001] that combines the qualitative model based on projections by Mukerjee and Joe [1990a] and the cardinal orientation approach by Frank [1991]. In these works the interpretation of the real environment consists in: (i) detecting the corners of the room where the robot is located using a laser scanner [Peris and Escrig, 2005]; and (ii) detecting the walls of the corridors of the robot environment and the intersections of these corridors using a laser scanner [Frommberger and Wolter, 2008]. However, we think that more information can be extracted from a real environment using all the robot sensors (camera, laser sensors, sonar sensors, etc.), just as human beings use all their senses to find their way around.

In the literature there are some approaches dealing with the problem of describing and recognizing *natural landmarks* in the robot world [Tapus and Siegwart, 2006; Ramisa, 2009; Quattoni and Torralba, 2009; Wu et al., 2009; Liu et al., 2011].

Tapus and Siegwart [2006] built fingerprint sequences of visual scenes using a list of characters, where each character represents instances of (i) colour patches -by a capital letter describing the hue- (ii) vertical edges -by a 'v' character- (both extracted from images) and (iii) corners extracted from the laser scanner -by a 'c' character-. These fingerprints are matched for robot localization and navigation using a topological map.

Ramisa [2009] transformed panoramic images of indoor scenes into signatures consisting of a constellation of descriptors computed over the different types of local affine covariant region descriptors (Harris-Affine, Hessian-Affine, MSER) and SIFT and GLOH descriptors (see the work by Mikolajczyk et al. [2005] for an overview of all these methods) and use these signatures for robot localization and navigation using a topological map.

Quattoni and Torralba [2009] extended their original GIST descriptor that identifies outdoor scenes [Oliva and Torralba, 2001] (by transforming original images into energy spectra and generating a multidimensional space in which scenes sharing membership in semantic categories are projected closed together),

by including learning techniques such as support vector machines (SVMs) and methods of segmentation of regions of interest (ROI) for indoor scene recognition.

Wu et al. [2009] transformed digital images into census histograms and apply the principal component analysis operation to compare them. This approach is evaluates faster, but it also needs training using SVMs.

Liu et al. [2011] transformed digital images into SIFT [Lowe, 2004] flow images and compare them by determining the dense correspondence between them. As analysing optical flow or dense sampling of the time domain enables tracking, dense sampling in the space of world images is enables scene alignment.

All these approaches succeded in the task they were designed for. However, the approaches by Ramisa [2009], Quattoni and Torralba [2009], Wu et al. [2009] and Liu et al. [2011] abstract image features of natural landmarks by producing huge numerical file descriptions that cannot be interpreted or given a meaning without a correspondence of descriptions produced by other images of seen scenes or objects previously stored in memory. Moreover, most of them need training or learning techniques.

The idea of the fingerprints built by Tapus and Siegwart [2006] is similar to our aim, however we intend to produce a generalizable approach that extracts descriptions of visual scenes more understandable by human beings.

Therefore, the main aim of this research work is to develope a general approach that obtains a qualitative description of any previously unknown visual scene without using training or learning techniques. The more cognitive or human understandable description obtained, the better, because the applicability of our approach will be wider (i.e. scene retrieval in image data bases, scene description using natural language, visual scene understanding thought ontological representation, etc.).

Our main aims are: (i) to define a model that describes qualitatively any previously seen or unseen visual scene without using training or learning techniques; (ii) to provide a meaning to these qualitative information extracted by means of an ontology interpretable by any intelligent agent; and (iii) to define a similarity measure for comparing the qualitative descriptions obtained in order to recognize natural scene landmarks.

Finally, our future purposes are: (i) to use the qualitative information extracted from the scenes as the input to qualitative orientation models and navigation algorithms and (ii) to test our general approach in all the applications that require generating or recognizing approximatimate human descriptions of visual scenes.

## 1.2 Specific Objectives

In order for a mobile robot to be able to navigate efficiently in its world, it must be able to detect *landmarks* (or reference objects/places), build a map with them by relating these landmarks and then use this map to localize itself in the world and plan its path to the target. For a robot to be able to localize itself in the map, it must first find out which landmarks on the map correspond to those landmarks currently detected.

Previous work developed in the research group *Cognition for Robotics Research (C4R2)* and the spin-off *Cognitive Robots (C-Robots) S. L.* [Peris and

Escrig, 2005] uses the laser sensor of a mobile robot in order to detect corners of a room, determine they are the *landmarks* of the robot world and build a map to localize itself and to navigate efficiently inside a room. However, the only feature of a corner that a one-dimensional laser scan sensor can obtain is its geometrical shape, that is, if it is a concave or convex corner. As a consequence, all the concave corners are equal and all the convex corners are equal for the robot. In order to differentiate between these corners, visual information that could describe the objects/colours/etc. of the corner is needed. In this PhD thesis an approach to extract this visual information is provided.

Moreover, in order for a robot to navigate efficiently, avoiding collisions, it must also be able to detect obstacles in a robust way. Sonar distance sensors are based on sound waves that sometimes obtain inaccurate distances, such as when these sound waves rebound in corners. In contrast, laser sensors are more accurate, but as they are based on light rays, so some of them have problems obtaining the real distance to a glass window or a mirror because of the reflection of light. However, advantages of both kinds of distance sensors can be combined to obtain robust obstacle detection and this goal has been achieved in this PhD thesis.

The main objective of this PhD thesis is the qualitative interpretation of the data captured by distance sensors (such as laser and sonar sensors) and vision systems (such as a webcam) incorporated in a mobile robot. This qualitative interpretation of the sensory information allows us to:

- obtain a qualitative description of the landmarks of an indoor robot environment (i.e. corners and doors),
- extract knowledge of the robot environment by translating the qualitative information obtained into an ontology based on description logics,
- use the qualitative description obtained in order to recognize with a degree of similarity the qualitative descriptions of the same landmark in the robot world.

The structure of this PhD thesis is outlined in Figure 1.1. In one line of work, *numerical distance data* obtained from *laser and sonar distance sensors* are transformed into *fuzzy distance patterns* using *fuzzy logic*. Then *integration algorithms* are applied to obtain *reliable distances to obstacles*, including distances to special obstacles such as corners, glass windows and mirrors. In another line of work, given a *digital image* captured by a *camera* located on a mobile robot, *segmentation techniques* are applied in order to *extract the relevant regions of the image*. Then *qualitative models of shape, colour, topology and orientation* are used to *describe the image visually and spatially*. The qualitative description of any digital image is translated to an *ontology based on description logics* and using an *ontology schema* and a *reasoner*, more knowledge is inferred from the description such as a *classification of the images of the robot environment*. Moreover, if corners are considered as the landmarks of an indoor robot environment (detected from the laser distances in the approach defined by Peris and Escrig [2005]), then the qualitative distance and visual information are further information about those landmarks that can be used in their posterior recognition. Finally, *similarity measures* defined for *shape, colour, topology and orientation* can be used to compare the visual description of a landmark (such

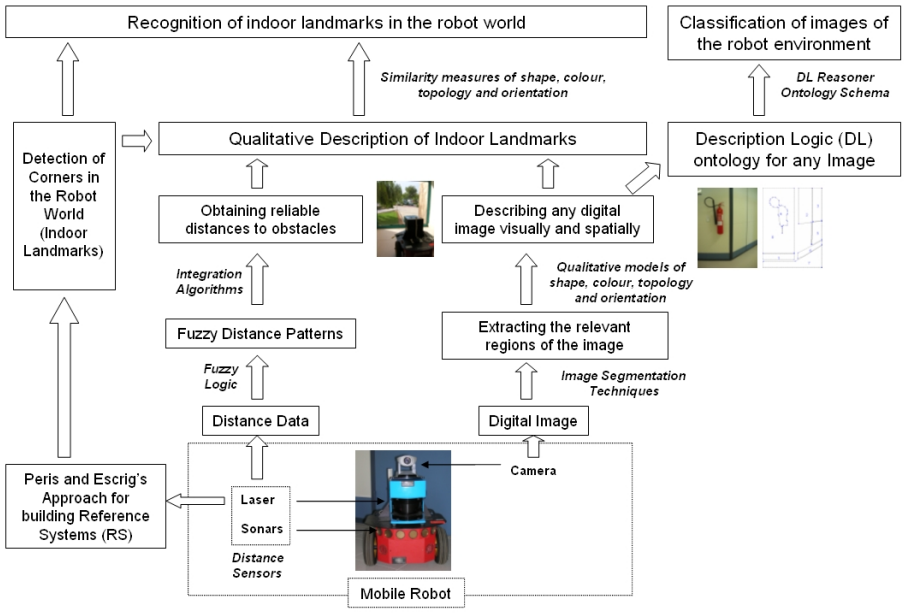



Figure 1.1: Diagram of the main objectives achieved in this PhD thesis.

as a corner) faced by the robot at the moment with the previous description of a known landmark and therefore *recognition of indoor landmarks* are achieved.

## 1.3 Contributions

The main scientific contributions of this thesis are:

1. A Distance Sensor Data Integration and Interpretation approach that obtains Fuzzy Distance Patterns ($FDP$s) from the data obtained by any kind of distance sensor and compares them using a Dissimilarity Factor ($DF$).

2. The application of the previous approach to the integration of the distances obtained by the sonar and laser distance sensors of a Pioneer 2 dx mobile robot and, as a result, the detection of special obstacles such as glass windows.

3. A Qualitative Image Description ($QID$) approach that describes any kind of digital image, is independent of the image segmentation method used and incorporates:

   i. A new Qualitative Shape Description ($QSD$) model based on the previous model by Museros and Escrig [2004].

   ii. A new Qualitative Colour Description ($QCD$) model for qualitative colour description based on Hue, Saturation and Lightness colour coordinates.

   iii. A Topology description based on the model by Egenhofer and Al-Taha [1992].

   iv. A Fixed Orientation ($FO$) description based on the model by Hernández [1991] and a Relative Orientation ($RO$) description based on the model by Freksa [1992].

4. The application of the $QID$ approach to two scenarios in robotics:

   i. the description of digital images captured by the camera of a Pioneer 2 dx mobile robot and

   ii. the description of digital images of tile mosaics taken by an industrial camera located on a platform used by a robot arm to assemble tile mosaics.

5. An approach that obtains a Description Logic (DL) based Ontology from each Qualitative Image Description ($QID$) in order to give a formal and explicit meaning to the qualitative visual and spatial features obtained.

6. The classification of the objects described in the images according to the definitions of an ontology schema and using a DL reasoner.

7. A Similarity measure between Qualitative Image Descriptions ($SimQID$) that incorporates:

   i. A Similarity measure between Qualitative Shape Descriptions ($SimQSD$).

   ii. A Similarity measure between Qualitative Colour Descriptions ($SimQCD$).

   iii. A Similarity measure between Topology descriptions based on the model by Egenhofer and Al-Taha [1992] ($SimTop$).

   iv. A Fixed Orientation ($FO$) description based on the model by Hernández [1991] ($SimFO$).

8. The application of the $SimQSD$ measure to extract objects with similar shapes from the MPEG7 CE Shape-1 library;

9. The application of the $SimQSD$ and $SimQCD$ measures to assemble tile mosaics by qualitative shape and colour similarity matching;

10. The application of the $SimQID$ measure to:

    i. compare images of tile compositions;

    ii. compare images of landmarks in a mobile robot world and identify them with a degree of similarity.

## 1.4 Outline of the Thesis

This thesis is structured in six chapters. Below is a brief outline of the thesis, starting with chapter two.

**Chapter 2: Distance Sensor Data Integration and Interpretation**
In this chapter we introduce a method for obtaining fuzzy distance patterns from the data captured by any kind of distance sensors, the dissimilarity factor defined for comparing fuzzy distance patterns and our algorithms to integrate

distances from sonar and laser distance sensors. Then we apply these methods to the specific case of a Pioneer 2 dx mobile robot and finally we test our development in two real scenarios using the Player Stage as the robot interface.

**Chapter 3: A Computational Model for Qualitative Description of Images**

This chapter presents a model for qualitatively describing any kind of digital image using qualitative models of shape, colour, topology and orientation and the application of this model to: (i) the description of digital images obtained by the camera of a Pioneer 2 dx mobile robot and (ii) the description of digital images of tile compositions taken by an industrial camera located on a platform used by a robot arm to assemble tile mosaics.

**Chapter 4: An Ontology for Qualitative Description of Images**

This chapter details the Description Logic (DL) based ontology designed to provide a formal and explicit meaning to the qualitative description of images generated by our model and the classification of the objects within the images achieved by our approach according to the definitions of an ontology schema and using a DL reasoner.

**Chapter 5: Similarity Measures between Qualitative Image Descriptions**

This chapter presents the definition of a similarity factor between the qualitative descriptions of shape, colour, topology and orientation of objects, which is also extended to obtain a similarity factor between two qualitative descriptions of images. Then these factors are applied to: (i) extract objects with similar shapes from the MPEG7 CE Shape-1 library; (ii) assemble tile mosaics by qualitative shape and colour similarity matching; (iii) compare images of tile compositions; (iv) compare images of landmarks in a mobile robot world and identify them with a degree of similarity.

**Chapter 6: Conclusions**

Finally in this chapter the conclusions of the thesis and future research directions are presented.

# Chapter 2

# Distance Sensor Data Integration and Interpretation

Human beings use many kinds of sensory information (sight, hearing, smell, taste, touch, etc.) in order to obtain a complete and reliable representation of their surroundings. Results of neuroscience studies by Hawkins and Blakeslee [2004] explain that the information captured by human senses is perceived by the cerebral cortex as spatiotemporal patterns that are stored as memories. In the same way, robots can incorporate different kinds of sensors -each one sensitive to a different property of the environment- whose data could be integrated to make the perception of the robot more robust and to obtain new information, otherwise unavailable. Information captured through the robot sensors can be expressed as patterns of concepts, by transforming numerical data obtained into qualitative terms, and then, these patterns can be stored in a knowledge base.

Some robot sensors, such as laser and sonar sensors, capture the same physical magnitude of the environment that is the distance to objects. Although these kinds of sensors are sensitive to different properties of the environment (light and sound properties, respectively), they present different drawbacks which can be overcome by integrating or fusing data obtained by them. If no method to combine data provided by all the robot sensors is effectively used, important information about the robot environment may be lost. This is the reason why sensor data integration is important for robot navigation, as it provides the robot with knowledge about its surroundings that can help it to carry out a task successfully and more efficiently.

Finally, observations in robot-oriented industrial applications by Liu [2008] emphasize that robots should have the capabilities of integrating reasoning, perception and action with conventional industrial tasks. In order to achieve such capabilities, the qualitative representation in a robotic system is required to have a natural connection to its quantitative representation and to provide the atomic representation that could be used to build higher level cognitive functions for robots to enable them to reason, act, and perceive in dynamic, partially unknown, and unpredictable environments. Our research work is a little step to contribute in this direction.

## 2.1 Related Work on Distance Integration

Sensor data fusion represents the process of combining data or information from multiple sensors for detecting obstacles, object recognition, tracking of objects, etc. Generally, in the literature there are many approaches that deal with the problem of distance sensor data integration by using different probabilistic methods and incorporating different kinds of sensors. These methods obtain a high precision in their application but at a high computational cost, and they usually obtain a description of the world that has a higher degree of precision than that required by the task to be performed by the robot.

Sonar and laser data integration for mobile robot navigation has been characterized by the use of probabilistic techniques: covariance intersections [Martin et al., 2006], grid maps and Bayes' theorem [Lai et al., 2005], grid maps and the Dezert-Smarandache theory [Li et al., 2005], Kalman filters [Diosi et al., 2005; Diosi and Kleeman, 2004], Gauss approximation methods and vector maps [Vamossy et al., 2004], fuzzy segment maps [Herrero et al., 2002], etc.

However, the extraction of knowledge from the world by numerical methods is very limited. A later interpretation of the coordinates where the robot is located is needed so that the robot can extract knowledge from the numerical values obtained. If the aim of sensor data integration is not localizing the robot accurately in the world, but extracting knowledge from it, qualitative representations are usually used. These representations define qualitative concepts for each important characteristic to distinguish in the world that can be used later on in decision processes.

In the literature, qualitative representations of sensor data have been applied to very few sensor data fusion approaches. We have found only Reece and Durrant-Whyte's works [Reece and Durrant-Whyte, 1995; Reece, 1997a,b, 2000] which mainly obtain qualitative descriptions of sensor cues. In their earlier work [Reece and Durrant-Whyte, 1995], they extracted regions of constant depth (RCD) from sonar sensor cues and interpret qualitatively the evolution of these RCD as the robot moves through the environment in order to identify planes, edges and corners. This work was extended [Reece, 1997a,b] in order to classify the surface curvature of the robot environment as convex, concave, plane, close-concave, far-concave, etc. and by combining a qualitative model based on intervals and qualitative differential equations (QDE) with a Kalman filter in order to interrelate the values of sensor observations with the system parameters and to estimate the parameters of the system for noisy processes or when the models obtained are incomplete or imprecise. Finally, another Reece's work [Reece, 2000] used qualitative descriptions of sensor cues for image understanding. A sensor cue contains a qualitative descriptor of the tool that processed the image, a qualitative representation of the spectral bands of the observed image -green, red, near-infrared or none of these- and a label denoting the interpretation of the representation. A qualitative reasoning system was built in order to distinguish soil from water in thermal daytime and nighttime images.

The motivation of our approach for distance sensor data integration is not localizing the robot accurately in the world, not interpreting sensor cues. Our main aim is to extract information about distances in the robot world by means of qualitative concepts that improve the knowledge of the robot of its surroundings and which could be used in later decision processes. Moreover, this quali-

tative information about distances is also used to detect obstacles in a robust way and also to characterize the obstacles or landmarks found in the world. Moreover, the semantic meaning of these qualitative names could be improved and related to others in the future by means of an ontology.

A more recent research trend in literature is integrating results from distance sensor fusion with images taken from a camera in order to extract knowledge from the environment by means of an XML dataset [Zivkovic et al., 2008], an ontology [Zender et al., 2008], symbolic/qualitative information [Oliveira et al., 2005], etc. This is the direction of our approach. However, first our aim is to extract knowledge from distance data fusion in an earlier stage and, later on, integrating it with the knowledge extracted from images by our approach based on description logics (see Chapters 3 and 4).

Our approach for distance sensor data integration (1) obtains patterns of fuzzy distances for each kind of distance sensor incorporated in a robot; (2) compares the patterns obtained in order to detect sensor errors; (3) integrates the patterns coming from different kinds of sensors to overcome the drawbacks presented individually and to obtain a more reliable perception of the environment; (4) provides knowledge to the robot by means of qualitative terms that categorize the distance of the robot to the obstacles and also types of obstacles; (5) can be extended and generalized for any kind of distance sensor and any kind of robot that includes distance sensors.

To the best of our knowledge, there is no approach for distance sensor data integration that presents these characteristics so it is not possible to carry out any comparative study with our approach.

## 2.2 Integration of Distance Sensor Data and Interpretation

Our approach for *Distance Sensor Data Integration and Interpretation* consists of the following steps, connected as Figure 2.1 shows.

i. **Obtaining Robust Fuzzy Distance Patterns (*FDP*s)** from each kind of distance sensor (described in Section 2.3), which involves:

   (a) Transforming distance sensor readings into patterns of fuzzy distance zones.

   (b) Comparing patterns of fuzzy distance zones obtained in order to detect those sensors that are not working properly.

   (c) Obtaining a robust fuzzy distance pattern for each kind of distance sensor.

ii. Integrating patterns provided by each kind of sensor in order to overcome each sensor's disadvantages and to obtain a final distance pattern that can **detect any sort of obstacles** (explained in Section 2.4).

iii. Calculating the discontinuities of distances in the final distance pattern and relating them to the corners detected by the approach by Peris and Escrig [2005] in order to **classify Reference Systems (RSs) in the robot world as *open* or *closed*** (as explained in Section 2.5).

iv. Defuzzifying the final distance pattern in order to provide the robot with a **smooth speed** depending on its frontal distance to the obstacle (described in Section 2.6).

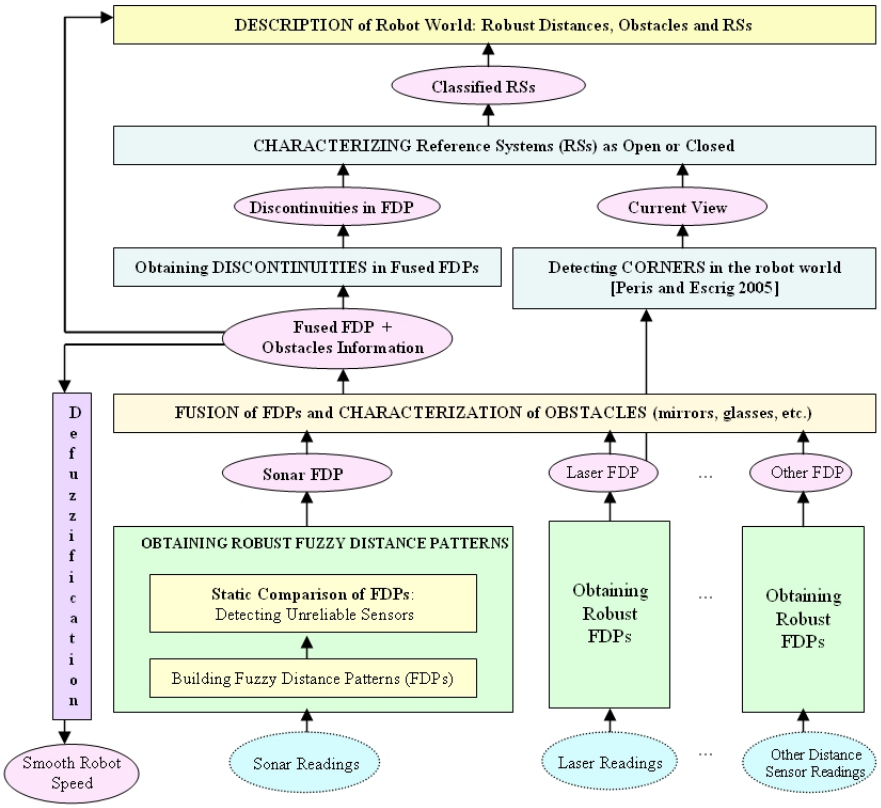



Figure 2.1: Scheme of our approach for Distance Sensor Data Integration and Interpretation.

Our approach can be extended and generalized for any kind of robot incorporating sonar and laser distance sensors and other kinds of distance sensors such as infrared an so on. In Section 2.7, the results of our tests on a real robot platform are detailed and, finally, in Section 2.8, our conclusions are explained.

## 2.3 Obtaining Robust Fuzzy Distance Patterns ($FDP$s)

One of the main objectives of our approach is obtaining a reliable *Fuzzy Distance Pattern* ($FDP$) for each kind of distance sensor on robot. In order to achieve this, first we obtain patterns of fuzzy distance zones from sensor numerical readings, as described in Section 2.3.1. Then we compare these patterns to detect sensors with technical problems, as described in Section 2.3.3. In order to compare fuzzy distance sets ($FdSets$), we have developed a *Dissimilarity Factor*($DF$), as explained in Section 2.3.2.

### 2.3.1 Building $FDP$s

In order to transform numerical distances obtained from the sensor readings into fuzzy distances, a fuzzy distance set is used. The concept of the fuzzy set was introduced by Zadeh [1965] as *a 'class' with a continuum of grades of membership*. Since then they have become the foundation of a methodology for translating the numerical data obtained from the world into linguistic categories or classes that can be given a meaning and used for reasoning.

In our approach, fuzzy distance values are used because (1) they provide linguistic values of distance that can be given a meaning and that can be useful to the robot later on for decision processes, and because (2) they can be easily defuzzified into the original numerical values.

Let us define a fuzzy set as a pair $(FdSet, \mu_{FdSet})$ where $FdSet$ is a set and $\mu_{FdSet} : FdSet \rightarrow [0,1] \in \Re$. For a finite set $FdSet = \{x_1, ..., x_n\}$, for each $x \in FdSet$, a $\mu_{FdSet}(x)$ is obtained and called the grade of membership of $x \in (FdSet, \mu_{FdSet})$.

Let $x \in FdSet$. Then $x$ is called not included in the fuzzy set $(FdSet, \mu_{FdSet})$ if $\mu_{FdSet}(x) = 0.0$, $x$ is called fully included if $\mu_{FdSet}(x) = 1.0$, and $x$ is called fuzzy member if $0.0 < \mu_{FdSet}(x) < 1.0$.

For each numerical distance obtained from the sensor readings, its grade of membership to each defined *Fuzzy distance Set* ($FdSet$) is calculated and those fuzzy distances obtained ($\{x_1, ..., x_n\}$) whose grade of membership is other than zero ($\mu_{FdSet}(x_i) \neq 0$) are selected. Those numerical sensor readings that are negative or exceed the sensor range are characterized as $out\_of\_range$ distances with the maximum grade of membership (1.0).

After transforming all sensor readings into fuzzy distances, those qualitative distances with the same name are grouped into zones and fuzzy distance patterns are obtained.

Let us consider a *Fuzzy Distance Pattern* ($FDP_t$) as a collection of fuzzy distance zones related to the same sensor scan at a time $t$. Each *zone* includes its starting and ending angular position, the event corresponding to the zone and a list of fuzzy distances related to it. The grade of membership of these fuzzy distances is obtained as the mean of all the grades of membership originally included in the zone. And the rest of parameters are defined as:

$FDP_{time}(SensorType)$ =([$\text{Zone}_0$, ... ,$\text{Zone}_K$]).
$SensorType$ = {*sonar*, *laser*, *infrared*, etc.}
$Zone_i = [[Start, End], FdSet, Event]$

$FdSet = [(x_1, \mu_{FdSet}(x_1)), ..., (x_i, \mu_{FdSet}(x_i)), ..., (x_n, \mu_{FdSet}(x_n))]$
$Start \in [0, MaxAngularRange] \in N$
$End \in [0, MaxAngularRange] \in N$
$Event = \{simple\_obstacle, glass\_or\_mirror, sound\_reflection, SensorType\_error\}$

An example of a situation of a general robot with a common distance sensor inside a common squared room is shown in Figure 2.2.

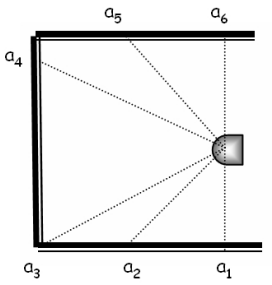



Figure 2.2: An example of a situation of the robot in a general squared room.

The general $FDP_t(sensor)$ corresponding to the example situation in Figure 2.2 is shown next. This $FDP_t(sensor)$ is made up of five $Zone$s ($K = 5$). The starting angular position of $Zone_1$ is $a_1{}^\circ$ and its ending angular position is $a_2{}^\circ$ and its angular amplitude is determined by $a_2 - a_1$. The fuzzy distances related to the first zone are determined by the distance names ($\{x_1, ..., x_n\}$) and grades of membership $\{\mu_{FdSet}(x_1), ..., \mu_{FdSet}(x_n)\}$ contained by $FdSet_{Zone_1}$. The sort of obstacle the robot is facing is determined by $Event$. The remaining zones of the pattern are described in the same way.

$FDP_t(sensor) = ([\text{Zone}_1, \ldots, \text{Zone}_5])$
$=$([ [[$a_1$, $a_2$], [ $FdSet_{Zone_1}$ ], Event],
[($a_2$, $a_3$], [ $FdSet_{Zone_2}$ ], Event],
[($a_3$, $a_4$], [ $FdSet_{Zone_3}$ ], Event],
[($a_4$, $a_5$], [ $FdSet_{Zone_4}$ ], Event],
[($a_5$, $a_6$], [ $FdSet_{Zone_5}$ ], Event]]).

As an example, if a Fuzzy distance Set is defined as: $(FdSet, \mu_{FdSet})$ where, $FdSet$ is a finite set $FdSet = \{at, very\_close, close, quite\_near, near, medium, quite\_far, far, very\_far, too\_far, extremely\_far, out\_of\_range\}$ and $\mu_{FdSet} : FdSet \rightarrow [0, 1] \in \Re$. The $FDP_t(sensor)$ corresponding to the example situation in Figure 2.2 could take the following values:

$FDP_t(sensor)$([
[[0, 46], [[close, 0.70], [quite_near, 0.30]], Simple_obstacle],
[(46, 58], [[quite_near, 0.56], [near, 0.44]], Simple_obstacle],
[(58, 123], [[near, 0.80], [medium, 0.20]], Simple_obstacle],
[(123, 137], [[quite_near, 0.56], [near, 0.44]], Simple_obstacle],
[(137, 179], [[close, 0.50], [quite_near, 0.50]], Simple_obstacle]
]).

### 2.3.2 Defining a Dissimilarity Factor ($DF$) between $FdSet$s

In order to compare *Fuzzy distance Sets* ($FdSet$s), a *Dissimilarity Factor* ($DF$) has been defined. This $DF$ compares both the qualitative distances and their corresponding grades of membership.

Given two general $FdSet$s, $FdSet_A$ and $FdSet_B$, containing $n$ and $m$ elements respectively:

$$FdSet_A = \{A_1, A_2, ..., A_n\} \text{ where } A_i = (x_{A_i}, \mu_A(x_{A_i}))$$
$$FdSet_B = \{B_1, B_2, ..., B_m\} \text{ where } B_j = (x_{B_j}, \mu_B(x_{B_j}))$$

The dissimilarity ($dSim$) between two elements, each corresponding to a different fuzzy set, $A_i$ and $B_j$, is defined by:

$$dSim(A_i, B_j) = dSimQd(x_{A_i}, x_{B_j}) \cdot \mu_A(x_{A_i}) \cdot \mu_B(x_{B_j}) \tag{2.1}$$

Note that $x_{A_i}$ corresponds to the fuzzy set name associated with fuzzy distance $A_i$, while $\mu_A(x_{A_i})$ is the grade of membership associated with the fuzzy set name $x_{A_i}$.

The dissimilarity between labels of qualitative distances $(x_{A_i}, x_{B_j})$ or $dSimQd$ in Eq. 2.1 is solved by analyzing the conceptual neighborhood relations between the concepts defined. The term *conceptual neighborhood* was introduced by Freksa [1991] in his analysis of the 13 interval relations defined in Allen's temporal logic Allen [1981]: *"Two relations between pairs of events are conceptual neighbors if they can be directly transformed one into another by continuous deformation (i.e., shortening or lengthening) of the events"*.

Conceptual neighborhood relations can be found between the qualitative labels defining distances. For example, the distances $x_i$ (i.e. *far*) and $x_{i+1}$ (i.e *very_far*) can be considered conceptual neighbors since a quantitative extension of the distance $x_i$ leads to a direct transition to the distance $x_{i+1}$. However, the distances $x_i$ and $x_{i-2}$ (i.e. *close*) are not conceptual neighbors, since a transition between them must go through the distance $x_{i-1}$ (i.e. *near*) first. Therefore, let us define the dissimilarity value between two qualitative distances that are conceptual neighbors as: (1) one positive unit if the first compared qualitative name represents a smaller distance than the second qualitative name (an increase in distance is noticed), and (2) one negative unit otherwise, as a decrease in distance is noticed. This definition can be represented in general as:

$$x_1 \xrightarrow[]{+1} x_2 \xrightarrow[]{+1} x_3 \cdots x_i \xrightarrow[]{+1} x_{i+1} \cdots x_{n-2} \xrightarrow[]{+1} x_{n-1} \xrightarrow[]{+1} x_n$$
$$x_1 \xleftarrow[-1]{} x_2 \xleftarrow[-1]{} x_3 \cdots x_i \xleftarrow[-1]{} x_{i+1} \cdots x_{n-2} \xleftarrow[-1]{} x_{n-1} \xleftarrow[-1]{} x_n$$

All the possible dissimilarity values between general qualitative distances are calculated in Table 2.1.

The dissimilarity value between the same qualitative distances is zero (that is $dSimQd(x_{A_i}, x_{A_i}) = 0$) and the dissimilarity between the fuzzy distances defined as the limits of the set is the maximum dissimilarity, which is the cardinality of the defined $FdSet$ ($dSimQd(x_{A_i}, x_{A_n}) = card(FdSet_A) - 1$).

Finally, the number of dissimilarities among all the elements that is needed to calculate in order to obtain the final dissimilarity between two fuzzy sets, $FdSet_A$ and $FdSet_B$, are given by the Cartesian product of their elements $(n \cdot m)$. Therefore, the *Dissimilarity Factor* ($DF$) between two fuzzy sets is

Table 2.1: Dissimilarity matrix for qualitative distances in general. Each cell corresponds to the formula: $(x_i, x_j) = j - i$.

| | $x_1$ | $x_2$ | $x_3$ | .. | $x_j$ | .. | $x_k$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 0 | 1 | 2 | .. | $j-1$ | .. | $k-1$ |
| $x_2$ | -1 | 0 | 1 | .. | $j-2$ | .. | $k-2$ |
| $x_3$ | -2 | -1 | 0 | .. | $j-3$ | .. | $k-3$ |
| .. | .. | .. | .. | .. | .. | .. | .. |
| $x_i$ | $1-i$ | $2-i$ | $3-i$ | .. | $j-i$ | .. | $k-i$ |
| .. | .. | .. | .. | .. | .. | .. | .. |
| $x_k$ | $1-k$ | $2-k$ | $3-k$ | .. | $j-k$ | .. | 0 |

obtained by accumulating the dissimilarity value between each pair of elements that composes each relation obtained by the Cartesian product of the two sets involved, as Eq. 2.2 shows.

$$DF(FdSet_A, FdSet_B) = \sum_{i,j=1}^{n \cdot m} dSim([A_1..A_n] \times [B_1..B_m]) \tag{2.2}$$

Considering the following $FdSet$s with cardinalities $n = 2$ and $m = 3$, that is,

$FdSet_A = [A_1, A_2] = [[x_{A_1},\ \mu_A(x_{A_1})], [x_{A_2},\ \mu_A(x_{A_2})]]$
$FdSet_B = [B_1, B_2, B_3] = [[x_{B_1},\ \mu_B(x_{B1})], [x_{B2},\ \mu_B(x_{B_2})], [x_{B_3},\ \mu_B(x_{B_3})]]$

the Equation 2.2 that calculates the $DF$ between both $FdSet$s corresponds to:

$$\begin{aligned} DF(FdSet_A, FdSet_B) &= dSimQd(x_{A_1}, x_{B_1}) \cdot \mu_A(x_{A_1}) \cdot \mu_B(x_{B_1}) \\ &+ dSimQd(x_{A_1}, x_{B_2}) \cdot \mu(x_{A_1}) \cdot \mu(x_{B_2}) \\ &+ dSimQd(x_{A_1}, x_{B_3}) \cdot \mu(x_{A_1}) \cdot \mu(x_{B_3}) \\ &+ dSimQd(x_{A_2}, x_{B_1}) \cdot \mu(x_{A_2}) \cdot \mu(x_{B_1}) \\ &+ dSimQd(x_{A_2}, x_{B_2}) \cdot \mu(x_{A_2}) \cdot \mu(x_{B_2}) \\ &+ dSimQd(x_{A_2}, x_{B_3}) \cdot \mu(x_{A_2}) \cdot \mu(x_{B_3}) \end{aligned}$$

Following the previous example, if a $FdSet$ is defined as: $(FdSet, \mu_{FdSet})$ where, $FdSet$ is a finite set $FdSet = \{$*at*, *very_close*, *close*, *quite_near*, *near*, *medium*, *quite_far*, *far*, *very_far*, *too_far*, *extremely_far*, *out_of_range*$\}$ and $\mu_{FdSet}$ : $FdSet \rightarrow [0, 1] \in \Re$, then the matrix of values built in order to obtain the dissimilarity between qualitative distance names ($dSimQd$) is shown in Table 2.2.

In order to exemplify the calculus of the Dissimilarity Factor ($DF$) defined, let us consider the instances of the previous example of $FdSet$:

$FdSet_A = [[at, 0.9], [very\_close, 0.1]]$
$FdSet_B = [[near, 0.75], [medium, 0.25]]$
$FdSet_C = [[near, 0.3], [medium, 0.8]]$
$FdSet_D = [[too\_far, 1.0]]$
$FdSet_E = [[extremely\_far, 1.0]]$

Table 2.2: Dissimilarity matrix for qualitative distances belonging to a particular $FdSet$.

| | *at* | *v.close* | *close* | *q.near* | *near* | *med.* | *q.far* | *far* | *v.far* | *t.far* | *ex.far* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *at* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *v.close* | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| *close* | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *q.near* | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *near* | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| *med.* | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| *q.far* | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
| *far* | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| *v.far* | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 |
| *t.far* | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 |
| *ex.far* | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 |

The $DF$s obtained for these $FdSet$s are the following:

(a) DF(FdSet$_A$, FdSet$_B$) = $dSimQd$(*at*, *near*) · (0.9 · 0.75) + $dSimQd$(*at*, *medium*) · (0.9 · 0.25) + $dSimQd$(*very_close*, *near*) · (0.1 · 0.75) + $dSimQd$(*very_close*, *medium*) · (0.1·0.25) = 4 · 0.675 + 5 · 0.225 + 3 · 0.075 + 4 · 0.025 = 4.15

(b) DF (FdSet$_B$, FdSet$_A$) = $dSimQd$(*near*, *at*) · (0.75 · 0.9) + $dSimQd$(*medium*, *at*) · (0.25·0.9) + $dSimQd$(*near*, *very_close*) · (0.75·0) + $dSimQd$(*medium*, *very_close*) · (0.25·0) = (-4) · 0.675 + (-5) · 0.225 + (-3) · 0.075 + (-4) · 0.025 = -4.15

(c) DF (FdSet$_A$, FdSet$_C$) = $dSimQd$(*at*, *near*) · (0.9 · 0.3) + $dSimQd$(*at*, *medium*) · (0.9 · 0.8) + $dSimQd$(*very_close*, *near*) · (0.1 · 0.3) + $dSimQd$(*very_close*, *medium*) · (0.1 · 0.8) = 4 · 0.27 + 5 · 0.72 + 3 · 0.03 + 4 · 0.08 = 5.09

(d) DF (FdSet$_C$, FdSet$_C$) = $dSimQd$(*near*, *near*) · (0.3 · 0.3) + $dSimQd$(*near*, *medium*) · (0.3 · 0.8) + $dSimQd$(*medium*, *near*) · (0.8 · 0.3) + $dSimQd$(*medium*, *medium*) · (0.8· 0.8) = 0 · 0.9 + 1 · 0.24 + (-1) · 0.24 + 0 · 0.64 = 0

(e) DF (FdSet$_D$, FdSet$_E$) = $dSimQd$(*too_far*, *extremely_far*) · (1.0 · 1.0) = 1 · 1.0 = 1.0

(f) DF ([at, 1.0],[extremely_far, 1.0]) = $dSimQd$(*at*, *extremely_far*) · (1.0 · 1.0) = 10 · 1.0 = 10

### 2.3.3 Comparing FDPs for detecting unreliable sensor readings

In order to detect sensor malfunctions, the environment is scaned three times while the robot is static and the three fuzzy distance patterns obtained are compared ($FDP_t, FDP_{t-1}, FDP_{t-2}$). Assuming that all the objects in the robot environment are also static, if a sensor obtains different readings depending on time and not on the situation, it is possible that this sensor has a technical problem.

The static comparison of patterns consists of comparing the current pattern ($FDP_t$) with the two previous ones ($FDP_{t-1}, FDP_{t-2}$) and determining which is the most new and reliable.

In order to compare fuzzy distance patterns, first a measure of similarity between the *Zones* that build them must be defined. Therefore, let us consider that two *Zones* are qualitatively similar ($Qsimilar$) if the fuzzy distance sets ($FdSet$s) contained in them have a $DF = 0$ and that they are close in meaning ($CloseMeaning$) if they have a $|DF| < Threshold$. As conceptual neighbour distances in our approach have a $|DF| = 1$, this *Threshold* can be given the value 2.

Therefore, two fuzzy distance patterns ($FDP_t, FDP_{t-1}$) are considered qualitatively similar ($QsimilarPatterns$) if they are composed of the same number of *Zones* which are qualitatively similar ($Qsimilar$).

As Algorithm 1 shows, if the current pattern ($FDP_t$) and at least one of the two previous ones are $QsimilarPatterns$, the current pattern is selected as the final pattern ($FinalFDP_t$). Otherwise, the two previous patterns are compared ($FDP_{t-1}, FDP_{t-2}$) and if they are $QsimilarPatterns$, the most recent pattern ($FDP_{t-1}$) is selected as the final one ($FinalFDP_t$). However, if none of the patterns are completely $Qsimilar$, a new pattern is built.

**Algorithm 1** Description of the static comparison of fuzzy distance patterns

From $Sensor(x)$ obtaining: $FDP_t$, $FDP_{t-1}$ and $FDP_{t-2}$
**if** $QsimilarPatterns(FDP_t, FDP_{t-1})$ or $QsimilarPatterns(FDP_t, FDP_{t-2})$ **then**
  $FinalFDP_t \leftarrow FDP_t$
**else if** $QsimilarPatterns(FDP_{t-1}, FDP_{t-2})$ **then**
  $FinalFDP_t \leftarrow FDP_{t-1}$
**else**
  $FinalFDP_t \leftarrow$ Building_a_New_$FDP(FDP_t, FDP_{t-1}, FDP_{t-2})$
**end if**

As Algorithm 2 shows, in order to build a new $FDP$s from the most reliable zones of the previous ones, for each angular position ($a$), the most current $FdSet$ that is $Qsimilar$ or have $CloseMeaning$ to the others is selected to build the $FinalFDP$. If there is any angular position ($a$) where all the $FdSet$s are very different from each other, nothing about the real distance can be known. Therefore this reading is characterized as *none*, meaning no distance, with the maximum grade of membership. A value of *none* suggests technical problems with the sensor located in the angular position where the reading was taken.

**Algorithm 2** Building_a_New_$FDP$

**for** each Angular position($a$) in $FDP_t$, $FDP_{t-1}$ and $FDP_{t-2}$ **do**
  $FdSet_t \leftarrow Extract\_FdSet(FDP_t, a)$
  $FdSet_{t-1} \leftarrow Extract\_FdSet(FDP_{t-1}, a)$
  $FdSet_{t-2} \leftarrow Extract\_FdSet(FDP_{t-2}, a)$
  **if** $Qsimilar(FdSet_t, FdSet_{t-1})$ or $Qsimilar(FdSet_t, FdSet_{t-2})$ **then**
    $FdSet_{final} \leftarrow FdSet_t$
  **else if** $Qsimilar(FdSet_{t-1}, FdSet_{t-2})$ **then**
    $FdSet_{final} \leftarrow FdSet_{t-1}$
  **else if** $CloseMeaning(FdSet_t, FdSet_{t-1})$ or
  $CloseMeaning(FdSet_t, FdSet_{t-2})$ **then**
    $FdSet_{final} \leftarrow FdSet_t$
  **else if** $CloseMeaning(FdSet_{t-1}, FdSet_{t-2})$ **then**
    $FdSet_{final} \leftarrow FdSet_{t-1}$
  **else**
    $FdSet_{final} \leftarrow (none, 1.00)$
    $Event \leftarrow SensorType\_error$
  **end if**
  $FinalFDP_t \leftarrow Add\_FdSet(FdSet_{final}, a)$
**end for**

## 2.4 Integration of Sonar and Laser FDPs and Detection of Static Special Obstacles

As sonar and laser sensors have problems in different situations of the robot environment, their readings can be integrated to overcome these problems and also to identify the specific situation that the robot is facing.

The main problems of sonar sensors are: multiple reflections in corners; uncertainty in locating the target due to the cone-shaped beam; and external ultrasound sources or crosstalk. Although laser sensors usually provide accurate readings, they may also present some drawbacks related to the nature of the target surfaces. Low reflectance surfaces, like dark colors or soft materials, absorb the laser beam and return it with a feeble intensity; while high reflectance surfaces present more serious problems: mirrors reflect the laser beam in any direction, while glasses can react to the laser beam as transparent, partial mirrors or perfect mirrors, depending on the glass type, thickness and angle of incidence. Therefore, as these sensors fail in different situations, a method to extract the advantages of both of them can be developed.

In Algorithm 3, the integration of sonar and laser $FDPs$ is described. After obtaining a robust $FinalFDP$ from both sonar and laser sensors (Section 2.3.3), we check if any of the sensors has technical problems, that is, if they obtain distances defined as *none* or *out_of_range*. If both kinds of sensors show technical problems for the same *Zone*, the corresponding *Zone* of the $FinalFDP$ will indicate *Sonar_Laser_error* as the *Event*. If, for the same *Zone*, one sensor has technical problems while the other one works well, the distance obtained by the second one is included in the $FinalFDP$ and the type of sensor that has technical problems (*sonar_error* or *laser_error*) is indicated as the *Event* for that *Zone*. If none of the sensors has technical problems in a *Zone*, then the

sonar $FdSet$ and the laser $FdSet$ corresponding to that $Zone$ are compared by calculating a *Dissimilarity Factor* ($DF$). If there is a large $DF$ between both $FdSet$s, we can determine that:

- If the $DF$ is larger than a threshold and positive, the distance obtained by the laser sensor is much larger than the sonar one. Therefore the robot could be facing a *glass window or mirror* that reflects the laser beam in any direction and this will be the *Event* determined.
- If the $DF$ is larger than a threshold and negative, the distance obtained by the sonar sensor is much larger than the laser one. Therefore the robot could be facing a corner that could have made the sound waves rebound and not return to the receiver. Then the *Event* determined would be *sound reflection* in corner.

If there is not a large $DF$ between the $FdSet$s obtained by each type of sensor, then the final $FdSet$ for any angular position is that obtained by the laser sensor, since it is the most accurate sensor, and the *Event* determined is *simple obstacle*.

**Algorithm 3** Description of the integration of sonar and laser $FDP$s

**for all** $FdSet(sonar)$ in $FDP(Sonar)$ and $FdSet(laser)$ in $FDP(Laser)$ **do**
  **if** $FdSet(sonar)$ and $FdSet(laser)$ are *none* **then**
    $FdSet(final), Event \leftarrow (none, 1.00), Sonar\_Laser\_error$
  **else if** $FdSet(sonar)$ and $FdSet(laser)$ are $out\_of\_range$ **then**
    $FdSet(final), Event \leftarrow (out\_of\_range, 1.00), Sonar\_Laser\_out\_of\_range$
  **else if** $FdSet(laser)$ is *none* or $out\_of\_range$ **then**
    $FdSet(final), Event \leftarrow FdSet(sonar), Laser\_error$
  **else if** $FdSet(sonar)$ is *none* or $out\_of\_range$ **then**
    $FdSet(final), Event \leftarrow FdSet(laser), Sonar\_error$
  **else**
    $DF(FdSet(sonar), FdSet(laser))$
    **if** $|DF| \geq Threshold$ and $DF > 0$ **then**
      $FdSet(final), Event \leftarrow FdSet(sonar), glass\_or\_mirror$
    **else if** $|DF| \geq Threshold$ and $DF < 0$ **then**
      $FdSet(final), Event \leftarrow FdSet(laser), sound\_reflection$
    **else**
      $FdSet(final), Event \leftarrow FdSet(laser), simple\_obstacle$
    **end if**
  **end if**
  $FinalFDP \leftarrow FdSet(final), Event$
**end for**

Next, let us consider the following examples of $FDP$s corresponding to the situation in Figure 2.3 in order to give an example of integration:

$FDP(sonar)$=([(...)
[ [50, 69 ], [(close, 0.56), (quite_near, 0.44)] ],
[ [70, 149 ], [(quite_near, 0.7),(near, 0.31)] ], (...) ]).

$FDP(laser)$=([(...)
[ [66, 86 ], [(quite_near, 0.64),(near, 0.36)] ],
[ [87, 106 ], [(extremely_far,1)] ],
[ [107, 122 ], [(quite_near,0.63), (near,0.37)], (...) ]).

DF( [(quite_near,0.7)(near,0.31)], [(quite_near,0.64)(near,0.36)] )= 0.05
DF( [(quite_near,0.7)(near,0.31)], [(extremely_far,1)] )= 6.76
DF( [(quite_near,0.7)(near,0.31)], [(quite_near,0.63)(near,0.37)] )= 0.06

$FDP(final)$=([(...)
[ [66, 86 ], [(quite_near, 0.64),(near, 0.36)], Simple Obstacle],
**[ [87, 106 ], [(quite_near, 0.7),(near, 0.31)], Glass or mirror],**
[ [107, 122 ], [(quite_near, 0.63)(near, 0.37)] Simple Obstacle], (...) ]).

By analyzing sonar and laser $FDPs$, we can see that the laser distances for the angular positions 87 to 106 are larger than the sonar ones because the obtained $DF$ is 6.76, therefore the $Final\_FDP$ includes the sonar $Zone$ as the right one and points to *glass or mirror* in that angular position.

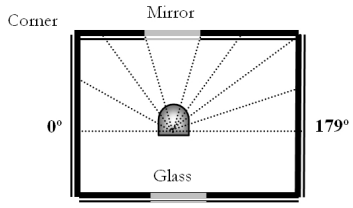



Figure 2.3: Example of a robot located inside a general room with a mirror and a glass wall.

## 2.5 Characterizing Reference Systems (RS) using the defined FDPs and DF

In this section, a $DF$ is obtained between the pairs of $FdSet$s that compose the final $FDP$ in order to discover if there are large discontinuities in the distances obtained. If these exist, it is supposed that there is an area that cannot be seen by the current position of the robot. This information combined with the corner information provided by the approach defined by Peris and Escrig [2005], enables the robot to characterize Reference Systems (RS) as *closed* or *open*.

### 2.5.1 Overview of Peris and Escrig's approach for building RSs

In the approach by Peris and Escrig [2005], the corners in a room that are detected by a robot are defined as the main landmarks of that room. Two consecutive corners (which can be concave or convex) in a robot scan define a Reference System ($RS$) and a set of $RS$s determines the final map of the room. Two kinds of $RS$s can compose a hybrid map defined by Peris and Escrig [2005]: *closed RS* and *open RS*. *Closed RS*s are those in which a new landmark cannot appear between those landmarks, which are the limits of the $RS$ and define it. However, *open RS*s are those in which new landmarks can appear between the two landmarks that define the $RS$ and, as a consequence, a more accurate exploration is needed in order to define clearly all the main landmarks in the room.

As an example, let us consider the robot situation shown in Figure 2.4, where the robot has detected four corners: C1 and C4 are *concave*, while C2 and C3 are *convex*. By joining these consecutive corners, three $RS$s are obtained, $RS_{12}$, $RS_{23}$ and $RS_{34}$. As shown in Figure 2.4 $RS_{12}$ and $RS_{34}$ are *open* while $RS_{23}$ is *closed*. This characterization can also be inferred from the $FDPs$ obtained by our approach, as explained in Section 2.5.3.

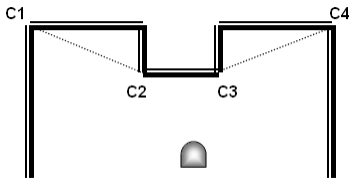



Figure 2.4: Example of a robot situation inside a room

### 2.5.2 Calculating Discontinuities in the Final FDP

The final fuzzy distance pattern ($FinalFDP$) resulting from the integration of the sonar and laser $FDPs$ (see Section 2.4) can be used to classify $RSs$, defined by Peris and Escrig [2005], as *open* or *closed*.

By calculating the $DissimilarityFactor$ ($DF$) between the $FdSets$ of each $Zone_i$ (simplified by $Z_i$) of the $FinalFDP$, large dissimilarities between $Zones$

can be detected, which correspond to discontinuities in the robot environment:

- If the $DF$ is larger than a threshold ($|DF| >> T$) and negative ($DF < 0$), it corresponds to a change from a $Zone$ containing large fuzzy distances to a $Zone$ with small fuzzy distances, which is called an *approaching_discontinuity* in our approach.
- If the $DF$ is larger than a threshold ($|DF| >> T$) and positive ($DF > 0$), it corresponds to a change from a $Zone$ containing small fuzzy distances to a $Zone$ with large fuzzy distances, which is called a *moving_away_discontinuity* in our approach.

The possible discontinuities that our approach could obtain from the $FinalFDP$ in the general situation shown in Figure 2.4 are the following:

| | |
|---|---|
| $FDP(final)$=([ | |
| [[$a_1$, $a_2$],[$FdSet_{Z_1}$]], | |
| [[$a_2$, $a_3$],[$FdSet_{Z_2}$]], | DF($FdSet_{Z_1}$, $FdSet_{Z_2}$) |
| ..., | ... |
| [[..., $a_{C2}$],[$FdSet_{Z_i}$]], | DF($FdSet_{Z_{i-1}}$, $FdSet_{Z_i}$) |
| [[ $a_{C2}$, $a_{C3}$],[$FdSet_{Z_{i+1}}$]], | DF($FdSet_{Z_i}$, $FdSet_{Z_{i+1}}$) $\rightarrow |DF| >> T$ and $DF < 0$ |
| [[$a_{C3}$, $a_j$],[$FdSet_{Z_{i+2}}$]], | DF($FdSet_{Z_{i+1}}$, $FdSet_{Z_{i+2}}$) $\rightarrow |DF| >> T$ and $DF > 0$ |
| ..., | ... |
| [[$a_k$, $a_{k+1}$],[$FdSet_{Z_k}$]]). | DF($FdSet_{Z_{k-1}}$, $FdSet_{Z_k}$) |

If we analyze the previous general distance pattern obtained by the robot when it is placed in the situation described in Figure 2.4, we observe that two large dissimilarities can be found:

- One between $\text{Zone}_i$ and $\text{Zone}_{i+1}$, reflected by a $DF$ large and negative, which corresponds to an *approaching* discontinuity in the angular position where corner $C_2$ is located approximately ($a_{C2}$); and
- One between $\text{Zone}_{i+1}$ and $\text{Zone}_{i+2}$, reflected by a $DF$ large and positive, which corresponds to a *moving away* discontinuity in the angular position where corner $C_3$ is located approximately ($a_{C3}$).

For example, if the $FinalFDP$ obtained is the following:

| | | |
|---|---|---|
| $FDP(final)$= | | |
| ([[0, 17,[(near, 0.18), (medium, 0.82)]], | = $\text{Zone}_0$ | |
| [[18, 35] ,[(medium, 0.58), (quite_far, 0.42)]], | = $\text{Zone}_1$ | DF($\text{Zone}_0$, $\text{Zone}_1$) = 0.6 |
| [[36, 48] ,[(quite_far, 0.71), (far, 0.29)]], | = $\text{Zone}_2$ | DF($\text{Zone}_1$, $\text{Zone}_2$) = 0.87 |
| [[49, **66**],[(medium, 0.57), (quite_far, 0.43)]], | = $\text{Zone}_3$ | DF($\text{Zone}_2$, $\text{Zone}_3$) = -0.86 |
| [[**67**,**111**], [(close, 0.30), (quite_near, 0.70)]], | = $\text{Zone}_4$ | DF($\text{Zone}_3$, $\text{Zone}_4$) = **-2.73** |
| [[**112**,125], [(medium, 0.52), (quite_far, 0.48)]], | = $\text{Zone}_5$ | DF($\text{Zone}_4$, $\text{Zone}_5$) =**2.78** |
| [[126,132], [(quite_far, 0.85), (far, 0.15)]], | = $\text{Zone}_6$ | DF($\text{Zone}_5$, $\text{Zone}_6$) = 0.67 |
| [[133, 146], [(medium, 0.55), (quite_far, 0.45)]], | = $\text{Zone}_7$ | DF($\text{Zone}_6$, $\text{Zone}_7$) = 0.7 |
| [[147, 179], [(near, 0.53), (medium, 0.47)]]]). | = $\text{Zone}_8$ | DF($\text{Zone}_7$, $\text{Zone}_8$) = 0.98 |

Two large dissimilarities are found:

- One between $\text{Zone}_3$ and $\text{Zone}_4$, reflected by a $DF$ of -2.73, which corresponds to an *approaching* discontinuity in the angular position where corner $C_2$ is located approximately at 67 degrees;
- One between $\text{Zone}_4$ and $\text{Zone}_5$, reflected by a $DF$ of 2.78, which corresponds to a *moving away* discontinuity in the angular position where corner $C_3$ is located approximately at 112 degrees.

### 2.5.3 Characterizing RSs as *open* or *closed*

By relating the approximate angular location of the discontinuities obtained from the final FDP with the approximate angular location and the type of the corners (*concave* or *convex*) detected, an algorithm to classify the reference systems (*RS*s) of the robot environment as *open* or *closed* can be defined.

The main situations in which the robot can find *consecutive* corners are shown in Figure 2.5 and detailed next:

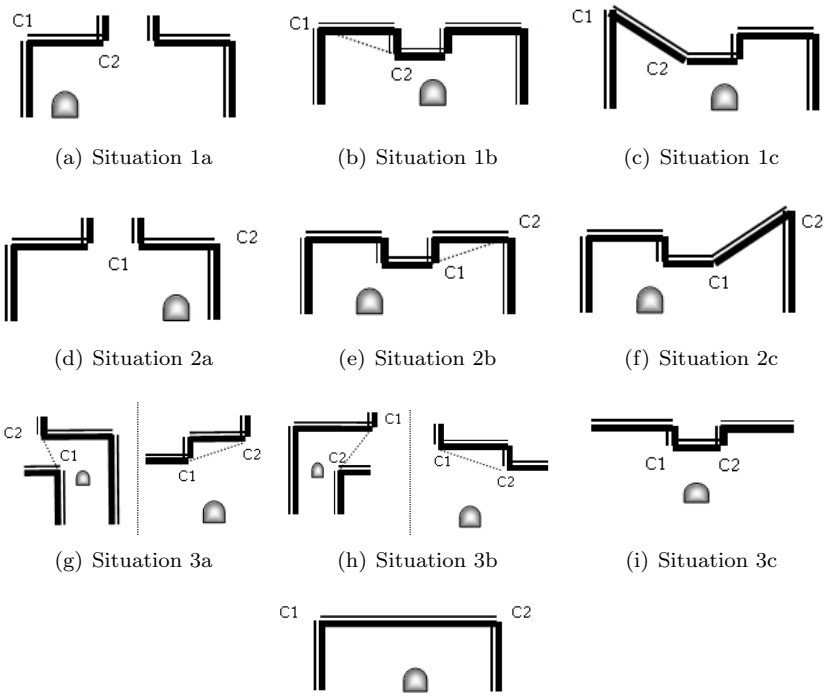
- **Situation 1**: a *concave* corner (C1) and a *convex* corner (C2),
  i. If a *moving away* discontinuity coincides with the *convex* corner (C2), if it happens after the *convex* corner, and then RS defined is *closed* (Situation 1a), but if it happens before the *convex* corner, and then RS defined is *open* (Situation 1b).
  ii. If there are no discontinuities that coincide with these corners, they define a *closed* RS (Situation 1c).
- **Situation 2**: a *convex* corner (C1) and a *concave* corner (C2),
  i. If an *approaching* discontinuity coincides with the *convex* corner (C1), it happens after the *convex* corner, and then the RS defined is *closed* (Situation 2a), but if it happens before the *convex* corner, and then the RS defined is *open* (Situation 2b).
  ii. If there are no discontinuities that coincide with these corners, they define a *closed* RS (Situation 2c). This situation is symmetrical to the previous one.
- **Situation 3**: two *convex* corners,
  i. If there is a *moving away* discontinuity coinciding with the first *convex* corner, the discontinuity happens after the corner, therefore the RS defined is *open* (Situation 3a).
  ii. If there is an *approaching* discontinuity coinciding with the second *convex* corner, the discontinuity happens before the corner and the RS defined is *open* (Situation 3b).
  iii. If an *approaching* discontinuity coincides with the first *convex* corner it happens before this corner and it corresponds to a previous RS. Similarly, if a *moving away* discontinuity coincides with the second *convex* corner, it happens after this corner and it corresponds to the following RS. This is the reason that although both convex corners coincide with discontinuities in situation 3c the RS determined by them is *closed*.

iv. If no discontinuity coincides with both convex corners, the RS defined by them is *closed*.

- **Situation 4**: Two *concave* corners. These corners always define a *closed* RS, since there cannot be a discontinuity of the robot environment between them, as a discontinuity involves the detection of another corner between the original ones.

Finally, it is important to consider that the same discontinuity in the distance pattern cannot be related to more than one open reference system.

Figure 2.5: Situations in which a robot can find different open or closed RS.




(a) Situation 1a (b) Situation 1b (c) Situation 1c

(d) Situation 2a (e) Situation 2b (f) Situation 2c

(g) Situation 3a (h) Situation 3b (i) Situation 3c

(j) Situation 4

As Algorithm 4 shows, by analysing the previous situtations, it can be deduced that if the first corner is *convex* and coincides with a *moving away* discontinuity, this discontinuity takes place before the second corner and, therefore, they define an *open* RS. Similarly, if the second corner is a *convex* corner and coincides with an *approaching* discontinuity, this discontinuity also takes place before the second corner and, therfore, they define an *open* RS. In other situation, the RS determined by the two consecutive corners is *closed*.

**Algorithm 4** Classification of *RS* as *open* or *closed*

**if** (C1.type = *convex*) and (discontinuity_in(C1) = *moving_away*) **then**
  SR(C1,C2).type ← *open*
**else if** (C2.type = *convex*) and (discontinuity_in(C2) = *approaching*) **then**
  SR(C1,C2).type ← *open*
**else**
  SR(C1,C2).type ← *closed*
**end if**

### 2.5.4 Integrating Final $FDP$ with RS information

Because of its high precision, the corners in the robot world are obtained by analyzing the distances provided by the laser sensor.

However, when a glass surface is located in front of the robot, corners corresponding to the obstacles on the other side of the glass are detected. These corners are *false* corners that do not correspond to the real world. Therefore, in this situation, an integration of the information provided by the final $FDP$ with the information of the RSs obtained is needed.

The integration done in our approach consists in not considering those corners whose angular distance coincides with the location of a *glass or mirror*, as Algorithm 5 shows.

**Algorithm 5** Integration of the final $FDP$ with the Corners information to build the real RS.

**for** $Corner(id, angle, type)$ in $CornersVector$[0 .. N] **do**
  **if** $angle$ not included in a $Zone_i$ with $Event \leftarrow Glass\_or\_mirror$ **then**
    $NewCornersVector \leftarrow Corner(id, angle, type)$
  **end if**
**end for**
$BuildNewRSs(NewCornersVector$[0 .. M])

## 2.6 Defuzzification of FDPs to Obtain a Smooth Robot Speed

As our approach for distance sensor data integration obtains patterns of fuzzy distances, fuzzy set theory can be used in order to control the speed of the robot and to obtain smooth movements while the robot is approaching an obstacle.

Therefore, a fuzzy controller has been designed in order to define the robot speed depending on the frontal *Zone* (90° aproximately) of the $FinalFDP$ resulting from the integration. This controller is composed of:

- the fuzzy distance set $FdSet$ representing distance to the obstacles:

  $(FdSet, \mu_{FdSet})$ where $FdSet$ is a finite set $FdSet = \{x_1, ..., x_n\}$ and $\mu_{FdSet} : FdSet \rightarrow [0, 1] \in \Re$

- the fuzzy speed set $FSpeed$ representing the robot speed:

  $(FSpeed, \mu_{FSpeed})$ where $FSpeed$ is a finite set $FSpeed = \{y_1, ..., y_p\}$ and $\mu_{FSpeed} : FSpeed \rightarrow [0, 1] \in \Re$

- a set of rules that relate each element of the fuzzy distance set $FdSet$ with the corresponding fuzzy speed set $FSpeed$:

  *if distance is* $x_i \in FdSet$ *then speed is* $y_r \in FSpeed$

- a defuzzification method which obtains the corresponding numerical speed from the obtained fuzzy distances. Our approach uses the Centre of Gravity ($CoG$) weighted by height for defuzzification (2.3), although other methods could also be used (more details are given by Galindo [2007]).

$$\widehat{x} = \frac{\sum_{i=1}^{n} H_i \cdot CoG_i}{\sum_{i=1}^{n} H_i} \tag{2.3}$$

  The rules defined relate the $FdSet$s obtained with the corresponding fuctions that define the $FSpeed$ sets. The selected functions of the $FSpeed$ set are truncated by height according to the grade of membership of the corresponding $FdSet$s and finally, the Centre of Gravity ($CoG$) of the area defined by those functions is obtained as the final numerical robot speed.

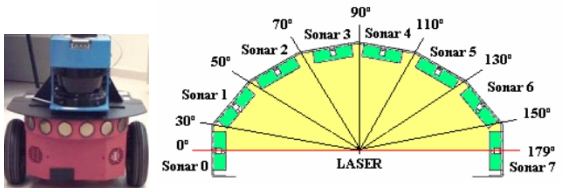
  The formula of the $CoG$ is shown in Eq. 2.4, where $FSpeed_i(x)$ are the functions representing speed selected by the rules; and the formula of the $CoG$ weighted by height is shown in Eq. 2.3, where $H_i$ is the height to truncate $FSpeed$ functions and which correspond to the grade of membership obtained by the $FdSet$s in the frontal $Zone$ of the $FDP$ (90° aproximately).

$$CoG_i = \frac{AreaX_i}{Area_i} = \frac{\int FSpeed_i(x) \cdot x \; dx}{\int FSpeed_i(x) \; dx} \tag{2.4}$$

## 2.7 Experimentation

The physical robotic platform used is an ActiveMedia Pioneer 2 dx mobile robot[1] containing eight sonar sensors and a SICK LMS-200 laser range scanner[2]. As Figure 2.6 shows, the laser sensor is mounted on the top of the robot and it does a 180 degree rotational scan, providing one reading per degree, whereas the eight sonar sensors are arranged in a half circle around it, providing only one reading per sensor for each scan. The sonar sensors incorporated by Pioneer 2 dx have a maximum range of 4 meters, while the SICK laser scanner can reach a maximum of 50 meters, but our approach determined the maximum in 8 meters since this range is enough for indoor robot navigation.

Figure 2.6: Robotic platform used to test our approach.




(a) Pioneer 2 (b) Distribution of the sensor readings

The software application used to carry out the experimentation of our approach is Player/Stage [3] as the network server for robot control, which provides a simple interface to the robot's sensors and actuators.

Finally, the testing scenarios used are those available in our University building:

- Scenario I: Pioneer 2 dx located at the end of a corridor, which is closed with a glass window (see Figure 2.7 (a)).
- Scenario II: Pioneer 2 dx located in front of the back doors of our building, which also have glass windows (see Figure 2.7 (b)).

Figure 2.7: Scenarios to exemplify the results of our approach.

(a) Scenario I (b) Scenario II

[1] http://www.mobilerobots.com
[2] http://www.sick.com
[3] http://playerstage.sourceforge.net

### 2.7.1 Parameters Selection

For parameterizing our approach, the $FdSet$ is defined as:

$(FdSet, \mu_{FdSet})$ where,
$FdSet$ is a finite set $FdSet = \{$*at*, *very_close*, *close*, *quite_near*, *near*, *medium*, *quite_far*, *far*, *very_far*, *too_far*, *extremely_far*, *out_of_range*$\}$ and
$\mu_{FdSet} : FdSet \rightarrow [0, 1] \in \Re$ defined by the triangular membership functions shown in Figure 2.8.

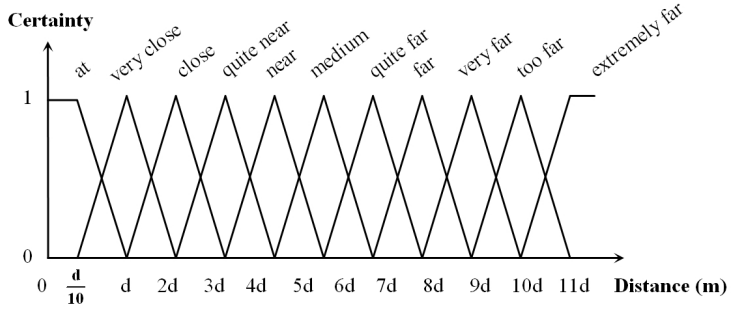



Figure 2.8: Fuzzy distance Sets ($FdSet$s) defined for our approach.

The membership functions of our $FdSet$ ($\mu_{FdSet}$) have been defined by experimentation for indoor robot navigation: each limit depends on the diameter of the robot ($d$) (e.g. 60 cm for a Pioneer 2 dx mobile robot, 40 cm for an ERRATIC[4] mobile robot , etc.). These divisions are motivated by the relation between the amount of distance in metres to an obstacle and the size of the object moving towards it. However, other distance labels and limits could be established according to the application.

The membership functions for our robot speed, named as $FSpeed$, are defined by experts according to our application as:

$(FSpeed, \mu_{FSpeed})$ where,
$FSpeed$ is a finite set $FSpeed = \{$*stopped*, *very_slow*, *slow*, *quite_slow*, *medium*, *fast*, *very_fast*, *real_fast*$\}$ and
$\mu_{FSpeed} : FSpeed \rightarrow [0, 1] \in \Re$ defined by the triangular membership functions shown in Figure 2.9.

Then the rules defined by our approach for relating the distances of our $FdSet$ with the speeds of our $FSpeed$ are shown in Algorithm 6. By activating them and applying the defuzzification method presented in (2.3), a smooth robot speed is obtained according to the distance to obstacles measured at the robot front (90° in its reading sensor distribution).

[4] http://www.videredesign.com

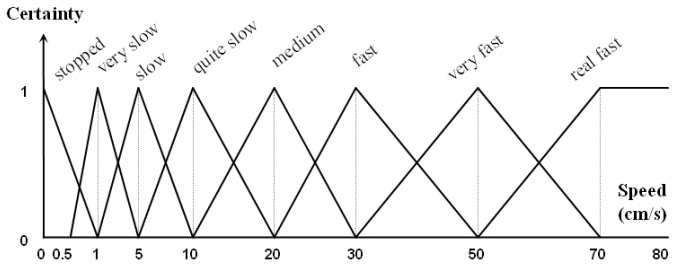



Figure 2.9: Fuzzy set *FSpeed* that defines the robot speed in centimetres per second.

**Algorithm 6** Rules for defuzzification

**if** *distance is extremely_far* **then** speed is real_fast
**if** *distance is too_far* **then** speed is very_fast
**if** *distance is very_far* **then** speed is very_fast
**if** *distance is far* **then** speed is fast
**if** *distance is quite_far* **then** speed is fast
**if** *distance is medium* **then** speed is medium
**if** *distance is near* **then** speed is quite_slow
**if** *distance is quite_near* **then** speed is quite_slow
**if** *distance is close* **then** speed is slow
**if** *distance is very_close* **then** speed is very_slow
**if** *distance is at* **then** speed is stopped

In our approach, as the membership functions for distance and speed are triangular (*FdSet* and *FSpeed*), the calculus of the centre of gravity (CoG) used in (2.3) is obtained from an extended trapezoid as illustrated by Figure 2.10 (a) and (b). Therefore, the final equation for defuzzification is defined as:

$$\widehat{x} = \frac{\sum_{i=1}^{n} H_i \cdot CoG_i}{\sum_{i=1}^{n} H_i}$$

$$CoG_i = \frac{AreaX_i}{Area_i} = \frac{H_i \cdot \frac{x_3^2+x_2^2-x_1^2-x_0^2+x_3 \cdot x_2-x_1 \cdot x_0}{6}}{H_i \cdot \frac{x_3+x_2-x_1-x_2}{2}}$$

$$CoG_i = \frac{x_3^2 + x_2^2 - x_1^2 - x_0^2 + x_3 \cdot x_2 - x_1 \cdot x_0}{3 \cdot (x_3 + x_2 - x_1 - x_2)}$$

The following example is given in order to explain how our fuzzy controller obtains the robot speed from the pattern of fuzzy distances obtained after the sensor data integration. Supposing the following distance pattern is obtained, the fuzzy distances in the frontal area of the robot are *close* with 0.68 of certainty and *quite_near* with 0.32 of certainty:

FDP(final) ([[0, 145], [[***close*, 0.68**], [***quite_near*, 0.32**]], *Simple_obstacle*], [[146, 179] , [[*very_close*, 0.43], [*close*, 0.57]], *Simple_obstacle*]).

Therefore, the rules for defuzzification that are activated are the following ones:

if *distance* is *quite_near* then *speed* is *quite_slow* in 0.32 of certainty and
if *distance* is *close* then *speed* is *slow* in 0.68 of certainty.

As a consequence, the extended trapezoid obtained is that illustrated in Figure 2.10 (c), which is determined by two heights $H_x = 0.68$ and $H_y = 0.32$ and the following points: $x_0 = 1$, $x_1 = 3.72$, $x_2 = 6.6$, $x_3 = 10$, $y_0 = 5$, $y_1 = 6.6$, $y_2 = 16.8$ and $y_3 = 20$. Finally, after applying 2.3 to the situation described, we obtain a speed of 7.5 centrimetres per second:

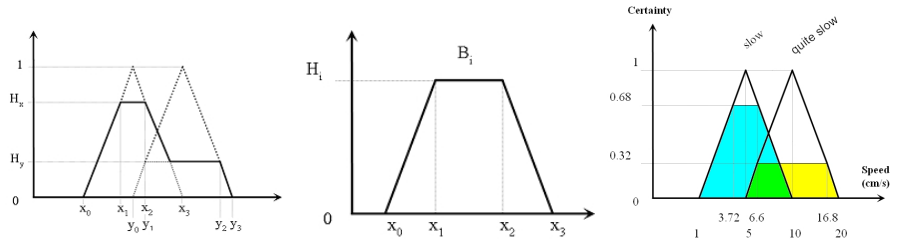
$$\widehat{x} = \frac{\sum_{i=1}^{n} H_i \cdot CoG_i}{\sum_{i=1}^{n} H_i} = \frac{H_x \cdot CoG_{slow} + H_y \cdot CoG_{quite\_slow}}{H_x + H_y}$$

$$CoG_{slow} = \frac{10^2 + 6.6^2 - 3.72^2 - 1^2 + 10 \cdot 6.6 - 3.72 \cdot 1}{3 \cdot (10 + 6.6 - 3.72 - 1)} = 5.36$$

$$CoG_{quite\_slow} = \frac{20^2 + 16.8^2 - 6.6^2 - 5^2 + 20 \cdot 16.8 - 6.6 \cdot 5}{3 \cdot (20 + 16.8 - 6.6 - 5)} = 12.16$$

$$\widehat{x} = \frac{0.68 \cdot 5.36 + 0.32 \cdot 12.16}{0.68 + 0.32} \approx 7.5$$

Figure 2.10: Relation of the parameters in the Equations 2.3 and the graphical representation of the fuzzy sets.




(a) Extended trapezoid

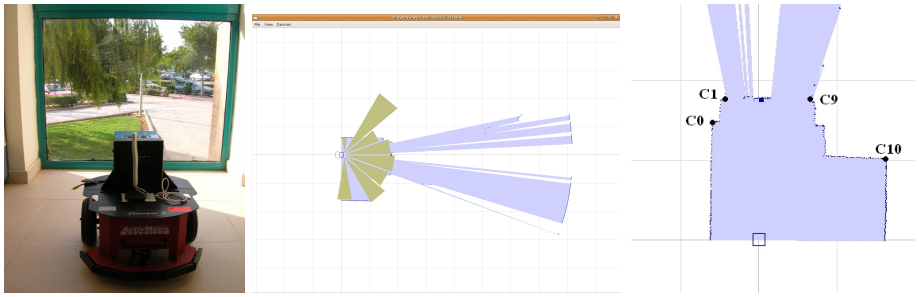
(b) A general trapezoid defined by the four points $x_0$, $x_1$, $x_2$, $x_3$ and its height ($H_i$)

(c) Extended trapezoid obtained from the example by applying the CoG weighed by height

### 2.7.2 Tests and Results in Scenario I

In Scenario I, our Pioneer 2 dx robot is located at the end of a corridor, which is closed with a glass window, as shown in Figure 2.11 (a). The distances obtained by the sonar and laser sensors in this scenario are those shown in the Player Viewer screenshot in Figure 2.11 (b).

Figure 2.11: Results obtained from the tests in Scenario I.




(a) Scenario I (b) Player Viewer application showing laser and sonar distances (c) Corners building *RS*s

The $FDP$s built from the distances captured by the distance sensors in this scenario are the following ones:

$FDP_t(sonar)$=
[ [0, 30), [(at, 0.11), (very_close, 0.89)] ],
[ [30, 50), [( near, 0.8), (medium, 0.2)] ],
[ [50, 70), [(very_close, 0.04), (close, 0.96)] ],
[ [70, 150), [(close, 0.66), (quite_near, 0.35)] ],
[ [150, 180), [(quite_near, 1)] ].

$FDP_t(laser)$=
[ [0, 54), [(very_close, 0.6), (close, 0.4)] ],
[ [54, 72), [(close,0.69), (quite_near, 0.31)] ],
[ [72, 110), [(extremely_far, 1)] ],
[ [110, 160), [(close, 0.6), (quite_near, 0.4)] ],
[ [160, 180), [(quite_near, 0.68), (near, 0.32)] ].

Both sonar and laser fuzzy distance patterns ($FDP(sonar)$ and $FDP(laser)$) are compared calculating the $DF$ between the corresponding $FdSet$s from (2.2). After determining a dissimilarity threshold by experimentation ($|T| > 2.5$), two important dissimilarities are found:

$DF(FdSet(sonar), FdSet(laser)) =$
$DF$([(near, 0.8),(medium, 0.2)],[(very_close, 0.6),(close, 0.4)]) $= -2.8$
$DF$([(close, 0.66), (quite_near, 0.35)],[(extremely_far, 1)]) $= 7.73$

According to the Algorithm 3, these dissimilarities are interpreted by our approach as a *Sound Reflection* in the angular positions [30,50) and a *Glass or mirror* in the

angular positions [72,110). Therefore, the final $FDP$ obtained is shown next:

$FDP_t(final)$=
[ [0, 30), [(very_close,0.6), (close,0.4)], Simple Obstacle],
[ [30, 50), [(very_close,0.6), (close,0.4)], Sound Reflection], $DF = -2.8$
[ [50, 54), [(very_close,0.6), (close,0.4)], Simple Obstacle],
[ [54, 72), [(close, 0.69), (quite_near,0.31)], Simple Obstacle],
[ [72, 110), [(close,0.66), (quite_near,0.35)], Glass or mirror], $DF = 7.73$
[ [110, 160), [(close, 0.6), (quite_near,0.4)], Simple Obstacle],
[ [160, 180), [(quite_near,0.68), (near,0.32)], Simple Obstacle].

The $FdSet$ of the final $FDP$ at time $t$ used to obtain the robot speed is that corresponding to the robot front (90°): [(close, 0.66),(quite_near,0.35)]. And, from (2.3), the speed obtained at time $t$ is: 0.08 cm/s. This speed decreases smoothly as time passes by and the robot is approaching the obstacle[5].

There are no large discontinuities of distance in the final $FDP_t$ because the $DF$s obtained from (2.2) when comparing the $FdSet$s that compose it are: 0, 0.91, 0.04, 0.05 and 0.93, respectively. Therefore, applying Algorithm 4, the Reference Systems ($RS$s) built from the corners obtained are all closed. Next, the pairs of corners extracted by the approach by Peris and Escrig [2005] are shown (each corner is related to its angular position and its geometrical convexity) and the result of Algorithm 4 is presented:

Corners[(47,Concave), (72,Convex)] → RS(C0, C1): Closed
Corners[(72,Convex), (75,Convex)]→ RS(C1, C2): Closed
Corners[(75,Convex), (78,Convex)]→ RS(C2, C3): Closed
Corners[(78,Convex), (79,Convex)]→ RS(C3, C4): Closed
Corners[(79,Convex), (83,Convex)]→ RS(C4, C5): Closed
Corners[(83,Convex), (84,Convex)]→ RS(C5, C6): Closed
Corners[(84,Convex), (87,Convex)]→ RS(C6, C7): Closed
Corners[(87,Convex), (95,Convex)]→ RS(C7, C8): Closed
Corners[(95,Convex), (111,Convex)]→ RS(C8, C9): Closed
Corners[(111,Convex), (148,Concave)]→ RS(C9, C10): Closed

From Figure 2.11 (c), it is important to note that: (i) the *Sound Reflection* detected in the final $FDP$ coincides with the angular position of the first corner (C0), that is 47°; and (ii) the angular positions that are classified as *Glass or mirror* coincide with *convex* corners: C1 at 72° and C9 at 109°.

As the *Glass or mirror* detected is situated between the angular positions 72 and 109, the following $RS$s are discarded: RS(C1,C2), RS(C2,C3), RS(C3,C4), RS(C4,C5), RS(C5,C6), RS(C6,C7), RS(C7,C8) and RS(C8,C9). Therefore, the $RS$s considered by the robot are RS(C0,C1) and RS(C9,C10). A new RS(C1,C9) is built as result of the integration of the final $FDP$ with the $CornersVector$, as shown in Algorithm 5. The angular position and type of these corners correspond to reality, as can be seen by comparing them with the image in Figure 2.11(c):

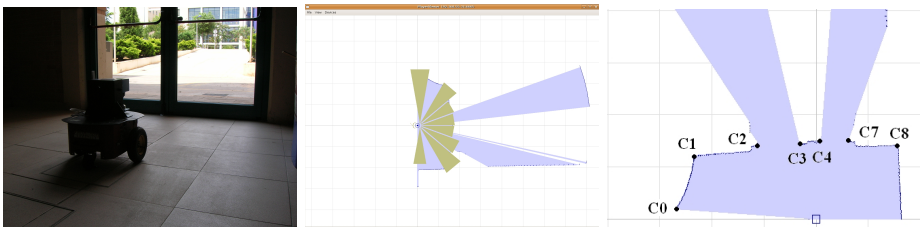
Corners[(47,Concave),(72,Convex)] → RS(C0,C1): Closed
Corners[(72,Convex), (111,Convex)] → New RS(C1,C9): Closed
Corners[(111,Convex),(148,Concave)]→ RS(C9,C10): Closed

[5]A video showing the execution of our application in this scenario can be downloaded from our website: http://www.c-robots.com/IJUFKS/scenario1.MOV

### 2.7.3 Tests and Results in Scenario II

In Scenario II, our Pioneer 2 dx robot is located in front of the back doors of our building, which also have glass windows, as shown in Figure 2.12 (a). The distances obtained by the sonar and laser sensors in this scenario are those shown in the Player Viewer screenshot in Figure 2.12 (b).

Figure 2.12: Corner detection and *RS*s building in Scenario II.




(a) Scenario II (b) Laser and sonar distances (c) Corners building *RS*s

The $FDP$s built from the distances captured by the distance sensors in this scenario are the following ones:

$FDP_t(sonar)$=
[ [0, 30), [(too_far, 0.18), (extremely_far, 0.82)] ],
[ [30, 50), [(medium, 0.68), (quite_far, 0.32)] ],
[ [50, 130), [(quite_near, 0.73), (near, 0.28)] ],
[ [130, 150), [(near, 0.84), (medium, 0.16)] ],
[ [150, 180), [(close, 0.14), (quite_near, 0.86)] ].

$FDP_t(laser)$=
[ [0, 35), [(medium, 0.69), (quite_far, 0.31)] ],
[ [35, 59), [(near, 0.5), (medium, 0.5)] ],
[ [59, 79), [(extremely_far, 1)] ],
[ [79, 92), [(quite_near, 0.75), (near, 0.25)] ],
[ [92, 110), [(extremely_far, 1)] ],
[ [110, 130), [(quite_near,0.37), (near, 0.63)] ],
[ [130, 144), [(near, 0.79), (medium, 0.21)] ],
[ [144, 180), [(quite_near, 0.51), (near, 0.49)] ].

As explained before, both sonar and laser fuzzy distance patterns ($FDP(sonar)$ and $FDP(laser)$) are compared calculating the $DF$ between the corresponding $FdSet$s from (2.2). In this scenario, three large dissimilarities are found ($-4.51$, 6.69 and 6.69, respectively), which are interpreted by our approach as a *Sound Reflection* in the angular positions [0,30) and a *Glass or mirror* in the angular positions [59,79) and [92,110). The final $FDP$ obtained is shown next:

$FDP_t(final)$=
[ [0, 30), [(medium, 0.69), (quite_far,0.31)], Sound Reflection], $DF = -4.51$
[ [30, 35), [(medium, 0.69), (quite_far, 0.31)], Simple Obstacle],
[ [35, 59), [(near,0.5), (medium,0.5)], Simple Obstacle],

[ [59, 79), [(quite_near,0.73), (near,0.28)], Glass or mirror], $DF = 6.69$
[ [79, 92), [(quite_near,0.75), (near,0.25)], Simple Obstacle],
[ [92, 110), [(quite_near,0.73), (near,0.28)], Glass or mirror], $DF = 6.69$
[ [110, 130), [(quite_near,0.37), (near,0.63)], Simple Obstacle],
[ [130, 144), [(near, 0.79), (medium, 0.21)], Simple Obstacle],
[ [144, 180), [(quite_near, 0.51), (near, 0.49)], Simple Obstacle].

The $FdSet$ of the final $FDP$ at time $t$ used to obtain the robot speed is that corresponding to the robot front (90°): [(quite_near, 0.75), (near, 0.25)]. And, from (2.3), the speed obtained at time $t$ is: 0.12 cm/s. This speed decreases smoothly as time passes by and the robot is approaching the obstacle[6].

There are no large discontinuities of distance in the final $FDP_t$ because the $DF$s obtained from (2.2) when comparing the $FdSet$s that compose it are: $-0.81$, $-1.24$, $-0.03$, 0.03, 0.36, 0.58 and $-0.72$, respectively. Therefore, applying Algorithm 4, the Reference Systems ($RS$s) built from the corners obtained are all closed. Next, the pairs of corners extracted by the approach by Peris and Escrig [2005] are shown (each corner is related to its angular position and its geometrical convexity) and the result of Algorithm 4 is presented:

Corners[(3,Convex),(32,Concave)] → RS(C0,C1): Closed
Corners[(32,Concave),(59,Convex)] → RS(C1,C2): Closed
Corners[(59,Convex),(79,Convex)] → RS(C2,C3): Closed
Corners[(79,Convex),(92,Convex)] → RS(C3,C4): Closed
Corners[(92,Convex),(93,Convex)] → RS(C4,C5): Closed
Corners[(93,Convex),(94,Convex)] → RS(C5,C6): Closed
Corners[(94,Convex),(110,Convex)] → RS(C6,C7): Closed
Corners[(110,Convex),(132,Concave)] → RS(C7,C8): Closed

From Figure 2.12(c), it is important to note that the *Sound Reflection* detected in the $FDP_t(final)$ coincides with the left round wall limited by corners C0 at 3° and C1 at 32°. It is also interesting to note that the angular positions of the first *Glass or mirror* coincides with two *convex* corners, C2 at 59° and C3 at 79°, and also the second *Glass or mirror* coincides with two *convex* corners, C4 at 92° and C7 at 110°.

As the *Glass or mirror* detected is situated between the angular locations 72-109 and 93-109, the following corners, located at the other side of the glass, are discarded: C5 at 93° and C6 at 94°. Therefore, the following $RS$s are discarded: RS(C4,C5), RS(C5,C6) and RS(C6,C7). And a new RS(C4, C7) is build following the Algorithm 5. Finally, the $RS$s considered by the robot are RS(C0,C1), RS(C1,C2), RS(C2,C3), RS(C3,C4), RS(C4,C7) and RS(C7,C8) and the angular location and type of these corners correspond to reality as it can be seen when comparing them with the situation in Figure 2.12 (c):

Corners[(3,Convex),(32,Concave)] → RS(C0,C1): Closed
Corners[(32,Concave),(59,Convex)] → RS(C1,C2): Closed
Corners[(59,Convex),(79,Convex)] → RS(C2,C3): Closed
Corners[(79,Convex),(92,Convex)] → RS(C3,C4): Closed

[6]A video showing the execution of our application in this scenario can be downloaded from our website: http://www.c-robots.com/IJUFKS/scenario2.MOV

Corners[(92,Convex),(110,Convex)] → New RS(C4,C7): Closed
Corners[(110,Convex),(132,Concave)] → RS(C7,C8): Closed

## 2.8 Conclusions

An approach to distance sensor data integration that provides a robust interpretation of the robot environment has been presented in this chapter. Our approach consists in obtaining patterns of fuzzy distance zones from sonar and laser sensor readings; comparing these patterns in order to detect non-working sensors; and integrating the patterns obtained to detect obstacles of any sort. A dissimilarity factor between fuzzy sets has been defined and applied to this approach. And a method for defuzzifying the obtained fuzzy distances into a fuzzy robot speed has also been used.

In order to test our approach, an ActivMedia Pioneer 2 dx mobile robot incorporating a SICK LMS-200 laser range scanner and the Player/Stage control interface have been used. However, our approach is extensible to other types of distance sensors (such as infrared sensors) and other kinds of mobile robots containing distance sensors.

The results obtained show that this approach enables the robot to: (1) detect non-working sensors, such as laser sensor disconnection, and avoiding crashing into obstacles when lacking the information provided by them, (2) detect mirrors and glass windows as obstacles; (3) obtain the real distance to corners, since the sonar sound reflections are detected; (4) approach obstacles at a smooth speed and avoid crashing into them; and (5) properly classify reference systems ($RS$s) as *open* or *closed*, including those that incorporate glass surfaces.

As future work, we intend to test our approach in other robotic platforms that incorporate the same or other kind of distance sensors.

# Chapter 3

# A Model for Qualitative Image Description

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 retrieve 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 or when they attempt to answer questions on the basis of visual memories.

Psycho-linguistic researchers have studied which language human beings use to describe from memory what they have seen: usually nouns are used to refer to objects, adjectives to express properties of these objects and prepositions to express relationships between them [Landau and Jackendoff, 1993]. These nouns, adjectives and prepositions are qualitative labels that can be used to extract knowledge from images and that can communicate image content.

Here, we propose an approach that obtains a qualitative description of any image composed of the visual and spatial features of the objects/regions within it. We consider the shape and colour of each region as visual features, which are absolute properties only depending on the region itself. We consider the topology and orientation of each region as spatial features, which are properties defined with respect to other regions (i.e. containers and neighbours of the regions).

This Chapter is organized as follows: Section 3.1 presents the related work; and Section 3.2 outlines our approach for qualitative image description. Then the qualitative models applied by our approach are explained: the qualitative

model for shape description in Section 3.3, the qualitative model for colour description in Section 3.4, the topological model in Section 3.5 and the qualitative orientation models in Section 3.6. In Section 3.8 the implementation of the proposed approach is explained. Section 3.9 details the two scenarios where the experimentation has been carried out and the obtained results. Finally, Section 3.10 explains our conclusions.

## 3.1 Related Work on Qualitative Description of Images

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] proposed 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] proposed an approach for classifying images of indoor scenes in semantic categories such as bookstore, 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 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 describing qualitatively 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 process. 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 in an artificial number of parts that depend on the size of the cell, while our approach extracts complete objects, which we think is more cognitive. The approach of Oliva and Torralba [2001] is useful for distinguish-

ing 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 developed by Quattoni and Torralba [2009] performs well for 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.

## 3.2 Outlining the Model for Qualitative Image Description (QID)

Some studies on how people describe images can be found in the literature [Jörgensen, 1998; Laine-Hernandez and Westman, 2006; Greisdorf and O'Connor, 2002; Wang et al., 2008]. The main objective of most of these studies is analysing people's image descriptions for image retrieval in databases.

Research by Jörgensen [1998] investigated 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 [2006] analysing how people describe journalistic photographs.

Psychological studies by Greisdorf and O'Connor [2002] showed that seven basic attributes generally ascribed to images when computer users look at them are: objects, colour, shape, texture, location, actions and affects.

Research by Wang et al. [2008] 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.

Figure 3.1: Examples of possible images to describe.

(a) Image containing a tree (b) Image containing a tree and a penguin (c) Image containing a tree, a penguin and a ball

In addition, we can also think about how we describe images. If we consider

Figure 3.1(a), as the image contains only one object, what we usually describe is the object and its orientation with respect to (wrt) the point of view of the observer. For example, a possible description of Figure 3.1(a) could be *an image containing a tree on the right and some clouds in a blue sky.*

If we consider Figure 3.1(b), as the image contains two objects, we can also describe the orientation of an object wrt the other object. For example, a possible description of Figure 3.1(b) could be *an image containing a tree on the right, a penguin in the lower left corner and some clouds in a blue sky. The tree is on the right hand side of the penguin.*

Finally, if we consider Figure 3.1(c), as the image contains more than two objects, then more relations of relative orientation between objects can be described. For example, a possible description of the image Figure 3.1(c) could be *an image containing a tree on the right, a penguin in the lower left corner, a football and some clouds in a blue sky. The ball is above the penguin and between the penguin and the tree.*

If we analyse these simple descriptions, we can observe that concepts of: shape (*football*) and colour (*blue*) are used to identify the objects and concepts of topology (*in*) and orientation (*left, right*) are used to locate the objects in the image.

All the studies in literature and our common sense gave us an idea of the concepts people pay attention to when they look at an image. In order to simplify the amount of information to extract from an image, we decided to tackle the problem of qualitative image description by describing objects within an image visually and spatially.

Therefore, the approach presented in this chapter first extracts the relevant objects/regions within a digital image and then describes them qualitatively by describing their visual and spatial features. 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 3.2.

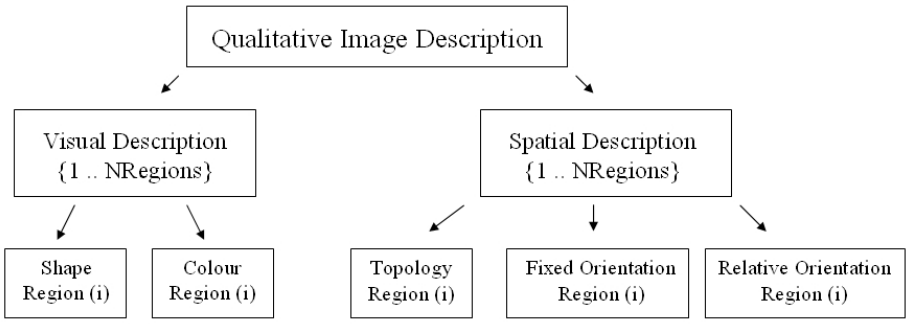



Figure 3.2: Structure of the qualitative image description obtained by our approach.

Exactly, in order to describe the visual content of the image, our approach uses the qualitative models of:

1. shape, described in Section 3.3, and
2. colour, described in Section 3.4.

And to describe the content of the image spatially, our approach uses the qualitative models of:

1. topology [Egenhofer and Franzosa, 1991], described in Section 3.5, and
2. orientation [Hernández, 1991; Freksa, 1992], described in Section 3.6.

## 3.3 Qualitative Shape Description

Wilson and Keil [1999] define shape as an aspect of a stimulus that remains invariant despite changes in size, position and orientation.

Therefore, shape is probably the single most significant property we perceive about objects. By knowing the shape of an object, the perceiver 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) [Palmer, 1999].

Moreover, shape is a visual and a spatial property of an object: it is first perceived by the sense of sight, but it can also be perceived spatially by the sense of touch.

Section 3.3.1 describes related work on qualitative shape description. Section 3.3.2 explains our approach for qualitative shape description and Section 3.3.3 describes how our approach characterizes 2D objects from the description of their sides and angles. Finally, Section 3.3.4 compares the approach of Museros and Escrig [2004] with our approach for qualitative shape description and it explains how the latter solves some problems presented by the former.

### 3.3.1 Related Work on Qualitative Shape Description

In general, approaches dealing with qualitative shape description can be classified as:

- **Axial representations:** these approaches are based on a description of the axes of an object, describing the shape qualitatively by reducing it to a "skeleton" or "axis". The "axis" is a planar arc reflecting some global or local symmetry or regularity within the shape. The shape can be generated from the axis by moving a geometric figure (called a "generator") along the axis and sweeping out the boundary of the shape. The generator is a constant shape and keeps a specified point (i.e. its centre) but can change its size and its inclination with respect to the axis. Some works inside this group correspond to the ones by Leyton [2004, 1988]; Brady [1983]; Gottfried [2005, 2008].
- **Primitive-based approaches:** in these approaches complex objects are described as combinations of more primitive and simple objects. Here we can distinguish two schemes:

  - **Generalized cylinder and geon-based representations**, which describe an object as a set of primitives plus a set of spatial connectivity relations among them [Biederman, 1985; Flynn and Jain, 1991];
  - **Constructive representations**, which describe an object as the Boolean combination of primitive point sets or half planes [Damski and Gero, 1996; Gero, 1999; Requicha, 1980; Brisson, 1993]

- **Topological and logic-based representations:** these approaches rely on topology and/or logics representing shapes [Cohn, 1995; Randell et al., 1992a; Clementini and Di Felice, 1997].

- **Cover-based representations:** in these approaches the shape of an object is described by covering it with simple figures, such as rectangles and spheres [Pobil and Serna, 1995].

- **Ordering and projection-based representations:** in these approaches different aspects of the shape of an object are represented by either looking at it from different angles or by projecting it onto different axes [Wang and Freeman, 1990; Schlieder, 1994; Damski and Gero, 1996]. In Museros and Escrig [2004] the vertices and curves of the shape of the objects are used to give a unique and complete qualitative description of each shape.

We will focus on the Museros and Escrig [2004] qualitative model for shape description, which describes objects qualitatively by naming the main qualitative features of the vertices and the maximal points of curvature detected in the shape of the object. This model is oriented to describe the shape of the edges of tiles that are automatically assembled into a ceramic mosaic by a robot arm. This qualitative approach deals with the uncertainty introduced by the fact that two tiles manufactured for a cell of a ceramic mosaic are never exactly identical but any one of them fits on that cell. It is also focused on the shape that real manufactured tiles can have. Not all imagined 2D objects can be made as tiles because sharp curves or very acute angles in the shape would cause the tile to break.

Our approach for qualitative shape description is an extension of the approach of Museros and Escrig [2004] but focused on obtaining a unique and complete qualitative description of any 2D object appearing in a digital image. Our extension of that approach consists in (1) qualitatively describing not only the maximal points of curvature of each curve, but also the qualitative features of its start and end points; (2) identifying the kind of edges connected by each vertex (such as two straight lines, a line and a curve or two curves); (3) adding the feature of qualitative compared length to the description of the points of maximum curvature; (4) defining the angle, the type of curvature and the relative length of the edges of the object at a fine level of granularity, and (5) naming the colour of the object according to a qualitative model defined in Section 3.4, instead of describing it using RGB coordinates. Finally, our approach also characterizes each object by naming it, according to its number of edges and kind of angles that its shape has, and by describing its convexity and regularity.

### 3.3.2 A New Qualitative Shape Description (QSD)

It is well known that 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 the most relevant ones, from a cognitive point of view, for describing shapes and this is the main reason why we use them.

Given a digital image containing a two-dimensional object, our approach for Qualitative Shape Description (QSD) automatically extracts the closed boundary of this object applying an image segmentation method. From all the points that define the boundary ($N$), a set of relevant points ($RPSet$) of the shape is extracted as described in Algorithm 7. The points of a boundary that are considered consecutive are those separated by a pre-established *granularity step(k)*. If the slope between a point $P_i$ and its consecutive point $P_{i+k}$, denoted by $s_1$,

**Algorithm 7** Extraction of the relevant points of the shape from all the pixels of the boundary of a 2D object.

**for** $i = 0$ to $N - 2k$; $i = i + k$ **do**
  $s_1 \leftarrow slope(P_i, P_{i+k})$
  $s_2 \leftarrow slope(P_i, P_{i+2k})$
  **if** $s_1 = s_2$ **then**
    $P_i, P_{i+k}, P_{i+2k} \in SameSegment$
  **else**
    $P_{i+k} \in RPSet$
  **end if**
**end for**

and the slope between $P_i$ and $P_{i+2k}$, termed $s_2$, are equal, then $P_i$, $P_{i+k}$ and $P_{i+2k}$ belong to the same straight segment. If $s_1$ and $s_2$ are not equal, $P_i$, $P_{i+k}$ and $P_{i+2k}$ belong to different straight segments or to a curved segment. This process is repeated for a new point $P_{i+3k}$ and the process stops when all the consecutive points of the boundary are visited. $P$ is considered a relevant point if it is the point at which the slope stops being constant or it is the point at which the slope changes its sign. Note that the *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* has a larger value; if they are short, the *granularity step* has a smaller value.

Finally, a set of relevant points, denoted by $\{P_0, P_1, ..., P_N\}$, determines the shape of the object. Each of those relevant points $P$ is described by a set of four features, which are defined below:

$$<\mathrm{KEC}_p,\ \mathrm{A}_p \mid \mathrm{TC}_p,\ \mathrm{L}_p,\ \mathrm{C}_p>$$

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 = {°, $A_{LAB}$, $A_{INT}$} where, degrees (°) 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 (°) 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]}

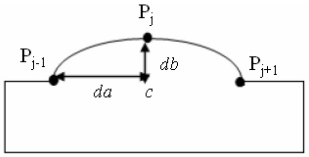



Figure 3.3: Obtaining the Type of Curvature of $P_j$.

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 = {°, $TC_{LAB}$, $TC_{INT}$}, where ° refers to the amplitude in degrees of the angle given by the relation between the distances *da* and *db* (see Figure 3.3 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 (°) 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]}

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.

$L_{LAB}$ = {*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, ∞)}

It is important to 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. However, they can be adjusted for a different application.

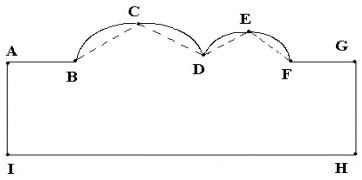



Figure 3.4: Describing how to obtain the approximate compared length between each pair of consecutive relevant points: vertices or points of curvature.

According to the kind of relevant point described, our approach calculates the compared length feature as follows:

- If point $P_j$ connects two straight lines (such as vertex A in Figure 3.4), the length of the first edge is the Euclidean distance between points $P_{j-1}$ and $P_j$(that is the length of the segment IA in Figure 3.4) and the length of the second edge is the Euclidean distance between points $P_j$ and $P_{j+1}$ (that is the length of the segment AB in Figure 3.4).
- If point $P_j$ connects a line with a curve, it is the starting point of a curve (such as vertex B in Figure 3.4). The length of the first edge is the Euclidean distance between points $P_{j-1}$ and $P_j$ (that is the length of the segment AB in Figure 3.4) and the approximate length of the second edge is the Euclidean distance between $P_j$ and the point of curvature $P_{j+1}$ (that is the length of the dashed line BC in Figure 3.4).
- If point $P_j$ connects a curve with a line, it is the ending point of a curve (such as vertex F in Figure 3.4). The approximate length of the first edge is the Euclidean distance between the point of curvature $P_{j-1}$ and the point $P_j$ (that is the length of the dashed line EF in Figure 3.4) and the length of the second edge is the Euclidean distance between points $P_j$ and $P_{j+1}$ (that is the length of the segment FG in Figure 3.4).
- If point $P_j$ connects two curves, it is the ending point of a curve and the starting point of another curve (such as vertex D in Figure 3.4). The approximate length of the first edge is the Euclidean distance between the point of curvature $P_{j-1}$ and the point $P_j$ (that is the length of the dashed line CD in Figure 3.4), and the approximate length of the second edge is the Euclidean distance between the point $P_j$ and the point of curvature $P_{j+1}$ (that is the length of the dashed line DE in Figure 3.4).
- If point $P_j$ is the point of curvature of a curve (such as C in Figure 3.4), the approximate length of the first edge is the Euclidean distance between the starting point of the curve $P_{j-1}$ and the point of curvature $P_j$ (that is the length of the dashed line BC in Figure 3.4), and the approximate length of the second edge is the Euclidean distance between the point of curvature $P_j$ and the ending point of the curve $P_{j+1}$ (that is the length of the dashed line CD in Figure 3.4).

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 relevant points of the shape clockwise. If point $P_j$ is on the left of the segment defined by $\mathrm{P}_{j-1}$ and $\mathrm{P}_{j+1}$, then $P_j$ is *convex*; otherwise $P_j$ is *concave*. For example, as Figure 3.5 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.

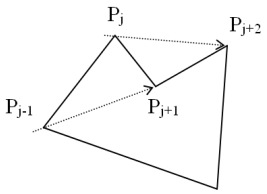



Figure 3.5: Obtaining the Convexity of $P_j$.

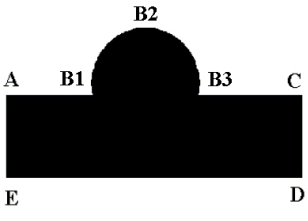
Therefore, the complete description of the shape of a 2D object is given from a set of qualitative tags as follows:

$$[\mathrm{KEC}_0,\ \mathrm{A}_0 \mid \mathrm{TC}_0,\ \mathrm{L}_0,\ \mathrm{C}_0],\ [\mathrm{KEC}_1,\ \mathrm{A}_1 \mid \mathrm{TC}_1,\ \mathrm{L}_1,\ \mathrm{C}_1],\ldots\ ,\ [\mathrm{KEC}_{n-1},\ \mathrm{A}_{n-1} \mid \mathrm{TC}_{n-1},\ \mathrm{L}_{n-1},\ \mathrm{C}_{n-1}]$$

where n is the total number of relevant points of the object, $\mathbf{KEC}_i$ describes the Kind of Edges Connected by the relevant point of the shape of the object, $\mathbf{A}_i \mid \mathbf{TC}_i$ describes the Angle or the Type of Curvature defined by the relevant point of the shape of the object, $\mathbf{L}_i$ describes the compared length of the edges connected by the relevant point of the shape of the object and finally, $\mathbf{C}_i$ describes the convexity of the relevant point of the shape of the object. 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.

Finally, an example of the QSD of an object is shown in Table 3.1.

Table 3.1: Qualitative description of a 2D object containing straight segments and curves.

| Object | Qualitative Description | |
|---|---|---|
|   | | QualitativeShapeDesc(Object)=[ |
| | A: | [line-line, right, sl, convex], |
| | B1: | [line-curve, obtuse, sl, concave], |
| | B2: | [curvature-point, acute, sl, convex], |
| | B3: | [curve-line, obtuse, absh, concave], |
| | C: | [line-line, right, abl, convex], |
| | D: | [line-line, right, msh, convex], |
| | E: | [line-line, right, ml, convex]]. |

### 3.3.3 Characterizing Objects by their Shapes

According to geometric principles, a characterization of the shape of the objects can be defined by using the qualitative features described for each relevant point. This characterization consists in: (1) giving a name to the object shape, (2) describing the regularity of its edges and (3) defining the convexity of the whole object. This charaterization can be divided into: objects with curves and objects without curves.

First, objects without curves can be characterized by a set of three elements:

$$[Name, Regularity, Convexity], \text{ where}$$
$$Name \in \{triangle,\ quadrilateral,\ pentagon,\ hexagon,\ heptagon,\ octagon,\ polygon\}$$
$$Regularity \in \{regular,\ irregular\}$$
$$Convexity \in \{convex,\ concave\}$$

$Name$ is the description given to the shape of the object depending on its number of relevant points and it can take values from *triangle* (3 relevant points) to *polygon* (more than 8 relevant points).

*Regularity* indicates if the object has all the same qualitative angles and all the edges of similar length (then it is *regular*), or not (then it is *irregular*).

$Convexity$ indicates if the object has a concave angle (then it is *concave*) or not (then it is *convex*).

However, for triangular and quadrilateral objects a more accurate characterization can be made. Triangular objects can be characterized as *right*, *obtuse* or *acute triangles* according to the kind of angles they have, and as *equilateral*, *isosceles* or *scalene triangles* according to the relation of length between the edges. Therefore, the $Name$ of a triangle is made up of three elements:

$$triangle\text{–}Kind\_of\_angles\text{–}Edges\_relation, \text{ where}$$
$$Kind\_of\_angles \in \{right,\ obtuse,\ acute\}$$
$$Edges\_Relation \in \{equilateral,\ isosceles,\ scalene\}$$

$Kind\_of\_angles$ indicates if the triangle has a right angle (then it is a *right-angled triangle*), an obtuse or very obtuse angle (then it is an *obtuse triangle*), or if all its angles are acute or very acute (then it is an *acute triangle*).

*Edges_relation* indicates if the edges of the triangle are all equal (then it is an *equilateral* triangle), or two are equal (then it is an *isosceles* triangle), or none are equal (then it is a *scalene* triangle).

Quadrilateral objects can also be characterized more accurately as *square, rectangle* or *rhombus* depending on the compared length between the edges and on the kind of angles. Therefore, the element $Name$ for a quadrilateral is made up of two elements:

$$quadrilateral\text{–}Type, \text{ where}$$
$$Type \in \{square,\ rectangle,\ rhombus\}$$

$Type$ specifies if the quadrilateral is a *square* (if all the angles are right and the edges have similar length), a *rectangle* (if all the angles are right and the opposite edges have similar length), or a *rhombus* (if all the edges have similar length and two opposite angles are obtuse or very obtuse and the other two are acute or very acute).

Secondly, objects with curves can be also characterized by a set of three elements:

[*Name*, *Regularity*, *Convexity*], where
*Name* ∈ {*circle*, *ellipse*, *polycurve*, *mixed-shape*}
*Regularity* ∈ {*regular, irregular*}
*Convexity* ∈ {*convex, concave*}

*Name* is the description given to the shape of the object depending on its properties: *mixed-shape* (if the shape has at least one *curvature-point* and at least one *line-line* relevant point), *polycurve* (if all the relevant points of the shape are *curvature-points*, *curve-curve*, *curve-line* or *line-curve* points), *circle* (if the shape of the object is a polycurve with only four relevant points, two of them defined as *semicircular* points of curvature) and *ellipse* (if the shape of the object is a polycurve with only four relevant points, two of them defined as points of curvature with the same type of curvature, that is, both veryplane, plane, acute or very acute).

*Regularity* of curves is not defined by our approach from the point of view of geometry. We consider 2D objects with circular or elliptical shapes to be *regular* and other objects with curvaceous shapes to be *irregular*.

*Convexity* of objects with curvaceous shapes is defined in the same way as for objects containing only straight edges: if an object has a *concave* relevant point, this object is defined as *concave*; otherwise it is defined as *convex*.

Thus, including the characterization of the objects, the complete description of a shape of an object defined by our approach is a set of qualitative tags such:

[[*Name*, *Regularity*, *Convexity*], *Qualitative_Colour*, [$\mathrm{KEC}_0$, $\mathrm{A}_0$ | $\mathrm{TC}_0$, $\mathrm{L}_0$, $\mathrm{C}_0$], ..., [$\mathrm{KEC}_{n-1}$, $\mathrm{A}_{n-1}$ | $\mathrm{TC}_{n-1}$, $\mathrm{L}_{n-1}$, $\mathrm{C}_{n-1}$]]

where *Name*, *Regularity* and *Convexity* describes the qualitative name given to the shape defined by those relevant points and *Qualitative_Colour* qualitatively describes the colour of the object as will be explained in Section 3.4.

The object shown in Table 3.1 as an example, it is characterized by our approach as a *black* object with a *mixed*, *irregular* and *concave* shape.

### 3.3.4 Comparing our QSD with the Approach of Museros and Escrig [2004]

In this Section, the approach of Museros and Escrig [2004] is summarized and some of the problems it has when describing some objects are presented. Moreover, how our QSD solves these problems is also explained.

#### 3.3.4.1 Outline of the Approach of Museros and Escrig [2004]

The complete description of a 2D object by Museros and Escrig [2004] is defined as a set of qualitative tags as:
[ [*Type*, [R,G,B], [$\mathrm{A}_0$, $\mathrm{C}_0$, $\mathrm{L}_0$] | [*curve*, $\mathrm{TC}_0$, $\mathrm{C}_0$], ..., [$\mathrm{A}_{n-1}$, $\mathrm{C}_{n-1}$, $\mathrm{L}_{n-1}$] | [*curve*, $\mathrm{TC}_{n-1}$, $\mathrm{C}_{n-1}$] ]

where $n$ is the total number of relevant points of the object, *Type* belongs to the set {*without-curves, with-curves*}, [R,G,B] describes the Red, Green and Blue coordinates of the colour of the object and $A_i$, $C_i$, $L_i$ and $TC_i$ describes the angles/curvatures, edges and convexity of the relevant points of the shape of an object, but with the following granularity and qualitative labels:

$$\mathrm{A}_i \in \{right,\ acute,\ obtuse\};$$
$$\mathrm{C}_i \in \{convex,\ concave\} \text{ and}$$
$$\mathrm{L}_i \in \{smaller,\ equal,\ bigger\}$$
$$\mathrm{TC}_i \in \{acute,\ semicircular,\ plane\}$$

As an example, Table 3.2 shows the qualitative description provided by Museros and Escrig's approach of a 2D object containing straight edges and curves. Note the differences between this description and the description provided by our New QSD of the same object presented in Table 3.1:

i. the kind of edges connected by the relevant points of the shape are described by our New QSD ($A$ is a *line-line*, $B1$ is a *line-curve*, $B2$ is a *curvature-point*, $B3$ is a *curve-line* and $C$, $D$ and $E$ are *line-line* relevant points), while missed by Museros and Escrig [2004];

ii. the start and end points of the curve ($B1$ and $B3$, respectively, see Table 3.1) are described by our New QSD, while missed by Museros and Escrig [2004];

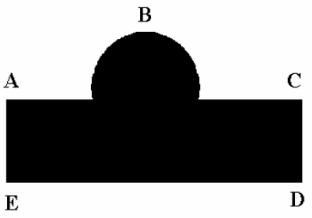
iii. the compared length at the *curvature-point* is described by our New QSD (the compared length at $B2$ is *similar length* or (*sl*)), while missed by Museros and Escrig [2004];

iv. the angle, type of curvature and the compared length of the edges of the object are defined at a finer level of granularity by our New QSD;

v. the colour of the object is described qualitatively by our New QSD instead of using its RGB colour coordinates as Museros and Escrig [2004] do;

vi. our New QSD also gives a name to the shape of the object and describes its regularity and convexity, while Museros and Escrig [2004] only differenciates if the object has curves (*with-curves*) or not (*without-curves*).

Table 3.2: Qualitative description of a 2D object containing straight segments and curves described by Museros and Escrig [2004].

| Object | Qualitative Description |
|---|---|
|  | QualitativeShapeDesc(Object)=<br>[ with-curves, [0, 0, 0],<br>[<br>A: [right, convex, smaller],<br>B: [curve, convex, acute],<br>C: [right, convex, bigger],<br>D: [right, convex, smaller],<br>E: [right, convex, bigger]<br>],<br>]. |

As the approach of Museros and Escrig [2004] was focused on describing manufactured tiles that could be assembled in a mosaic, it did not consider

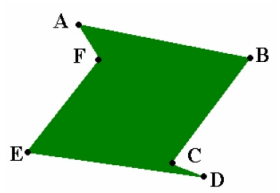
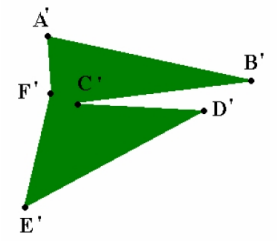
2D objects with sharp curves or very acute angles, which are very fragile and hardly ever used in mosaics. However, as our current purpose is to describe any 2D object contained in a digital image, no kind of shape can be discarded and we have found some situations where the qualitative description obtained by Museros and Escrig [2004] when describing two different objects is ambiguous.

#### 3.3.4.2 How Our New QSD Solve Problems of the Approach of Museros and Escrig [2004]

In this section, two ambiguous QSDs obtained by Museros and Escrig [2004] are shown and solved by our New QSD approach.

The **first ambiguous description** is shown in Table 3.3, in which two objects that appear very different to the human eye have the same qualitative shape description.

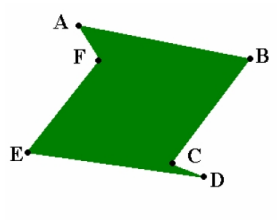
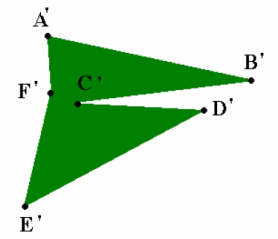
Table 3.3: Example of two 2D objects which are described by Museros and Escrig [2004] using exactly the same qualitative features.

| Object | Qualitative Description |
|---|---|
|  | QualitativeShapeDesc(Object 1)=<br>[ without-curves, [0, 128, 0],<br>[<br>A: [acute, convex, smaller],<br>B: [acute, convex, bigger],<br>C: [acute, concave, bigger],<br>D: [acute, convex, smaller],<br>E: [acute, convex, bigger],<br>F: [obtuse, concave, bigger],<br>],<br>]. |
|  | QualitativeShapeDesc(Object 2)=<br>[ without-curves, [0, 128, 0],<br>[<br>A': [acute, convex, smaller],<br>B': [acute, convex, bigger],<br>C': [acute, concave, bigger],<br>D': [acute, convex, smaller],<br>E': [acute, convex, bigger],<br>F': [obtuse, concave, bigger],<br>],<br>]. |

The first ambiguous description is solved by our New QSD because it uses a reference system for compared length with a finer level of granularity. Note that, in Table 3.3, the compared lengths for vertices C and C' are the same (*bigger*) and for D and D' are also the same (*smaller*). However, note that in the description provided by our New QSD, shown in Table 3.4, the compared length for C is *much longer* or *ml* while for C' is *a bit longer* or *abl* and that

the compared length for D is *much shorter* or *msh* while for D' is *a bit shorter* or *absh*. Therefore, both QSDs are not ambiguous.

Table 3.4: Qualitative description of 2D objects obtained by our New QSD, which solve the ambiguous situation presented in Table 3.3.

| Object | Qualitative Description |
|---|---|
|  | QualitativeShapeDesc(Object 1)=<br>[ [hexagon, irregular, concave], green,<br>[<br>A: [line-line, acute, convex, msh],<br>B: [line-line, acute, convex, abl],<br>C: [line-line, acute, concave, **ml**],<br>D: [line-line, acute, convex, **msh**],<br>E: [line-line, acute, convex, abl],<br>F: [line-line, acute, concave, ml],<br>]]. |
|  | QualitativeShapeDesc(Object 2)=<br>[ [hexagon, irregular, concave], green,<br>[<br>A': [line-line, acute, convex, msh],<br>B': [line-line, acute, convex, abl],<br>C': [line-line, acute, concave, **abl**],<br>D': [line-line, acute, convex, **absh**],<br>E': [line-line, acute, convex, abl],<br>F': [line-line, acute, concave, ml],<br>]]. |

The **second ambiguous description** produced by the approach by Museros and Escrig [2004] is shown in Table 3.5: all three objects have the same qualitative description because curves (its start, end and point of curvature) are described only by one relevant point. As a result, the straight edge between the two curves in Object 1 and also in Object 2 is not described, and both objects have the same number of relevant points as Object 3. Moreover, as the start and end points of the curves are not described, the approach by Museros and Escrig [2004] cannot establish a relation of size between the curves in Objects 1 and 2 and therefore curves B and B' and curves B" and C" in Table 3.5 cannot be distinguished.

Our New QSD approach solves the ambiguous description because:

i. it describes the start and end points of every curve as any other relevant point. Therefore, our New QSD approach can describe segments between two curves (i.e. DE and D'E' in Objects 1 and 2 in Table 3.6) and vertices connecting two curves (such as vertex D" in Object 3 in Table 3.6).

ii. it defines a reference system for the type of curvature (TCRS) at a fine level of granularity. Therefore, it can distinguish very plane curves (i.e. C" in Object 3 of Table 3.6) from plane curves (i.e. E" in Object 3 of Table 3.6).

Finally, Objects 1 and 2 in Table 3.6 can be distinguished by their New QSDs, which have different compared length descriptions for vertices B and B', D and D', E and E' and G and G', respectively. Object 3 can be distinguished from Objects 1 and 2 because its New QSD has one vertex less to describe. Moreover, the New QSD can difference the type of curvature of both curves in Object 3 (C" and E"), which are described by distinct qualitative tags (*plane* and *very-plane*).

Table 3.5: Example of three 2D objects containing different kind of curves in different positions whose qualitative description generated by the model by Museros and Escrig [2004] is the same.

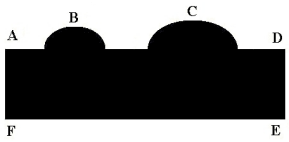
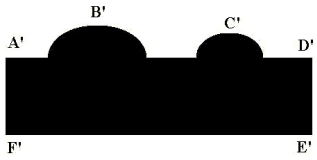
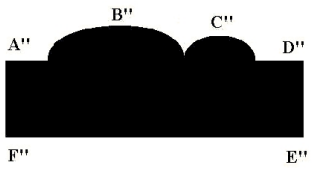
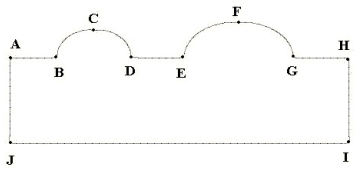
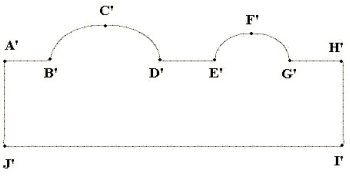
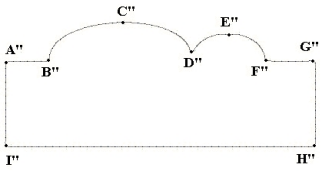
| Object | Qualitative Description |
|---|---|
|  | QualitativeShapeDesc(Object 1)=<br>[ without-curves, [0, 0, 0],<br>[<br>A: [right, convex, bigger],<br>B: [curve, convex, plane],<br>C: [curve, convex, plane],<br>D: [right, convex, smaller],<br>E: [right, convex, smaller],<br>F: [right, convex, bigger],<br>],<br>]. |
|  | QualitativeShapeDesc(Object 2)=<br>[ without-curves, [0, 0, 0],<br>[<br>A': [right, convex, bigger],<br>B': [curve, convex, plane],<br>C': [curve, convex, plane],<br>D': [right, convex, smaller],<br>E': [right, convex, smaller],<br>F': [right, convex, bigger],<br>],<br>]. |
|  | QualitativeShapeDesc(Object 3)=<br>[ without-curves, [0, 0, 0],<br>[<br>A": [right, convex, bigger],<br>B": [curve, convex, plane],<br>C": [curve, convex, plane],<br>D": [right, convex, smaller],<br>E": [right, convex, smaller],<br>F": [right, convex, bigger],<br>],<br>]. |

Table 3.6: Qualitative description of objects obtained by our approach, which solve the ambiguous situation presented in Table 3.5.

| Object | Qualitative Description |
|---|---|
|  | QualitativeShapeDesc(Object 1)=<br>[ [mixshape, irregular, concave], black,<br>[<br>A: [line-line, right, dl, convex],<br>B: [line-curve, obtuse, **absh**, concave],<br>C: [curvature-point, plane, sl, convex],<br>D: [curve-line, obtuse,**sl**, concave],<br>E: [line-curve, obtuse,**absh**, concave],<br>F: [curvature-point, plane, sl, convex],<br>G: [curve-line, obtuse,**abl**, concave],<br>H: [line-line, right, hl, convex],<br>I: [line-line, right, msh, convex,<br>J: [line-line, right, ml, convex]<br>]]. |
|  | QualitativeShapeDesc(Object 2)=<br>[ [mixshape, irregular, concave], black,<br>[<br>A': [line-line, right, dl, convex],<br>B': [line-curve, obtuse, **hl**, concave],<br>C': [curvature-point, plane, sl, convex],<br>D': [curve-line, obtuse,**abl**, concave],<br>E': [line-curve, obtuse,**sl**, concave],<br>F': [curvature-point, plane, sl, convex],<br>G': [curve-line, obtuse,**sl**, concave],<br>H': [line-line, right, hl, convex],<br>I': [line-line, right, msh, convex],<br>J': [line-line, right, ml, convex]<br>]]. |
|  | QualitativeShapeDesc(Object 3)=<br>[ [mixshape, irregular, concave], black,<br>[<br>A": [line-line, right, dl, convex],<br>B": [line-curve, obtuse, hl, concave],<br>C": [curvature-point,**very_plane**,sl,convex],<br>D": [curve-curve, obtuse, **ml**,concave],<br>E": [curvature-point,**plane**,sl,convex],<br>F": [curve-line, obtuse, sl, concave],<br>G": [line-line, right, hl, convex],<br>H": [line-line, right, msh, convex],<br>I": [line-line, right, ml, convex]<br>]]. |

## 3.4 Qualitative Colour

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 [Conway, 1992].

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 have their own particular hues, saturations and lightness [Palmer, 1999].

### 3.4.1 Related Work on Qualitative Colour

Colour naming models are designed to relate a numerical colour space with semantic colour names used in natural language. Therefore, they are an effective and widely used way to support semantic-based image retrieval [Liu et al., 2007].

In the literature, different colour naming models have been defined based on different colour spaces or even combinations of some of these.

Some of the colour spaces found in the literature are: RGB (red, green and blue), HSL (hue, saturation and lightness), HSV/HSB (hue, saturation and value or brightness), HSI (hue, saturation and intensity), CIE (*Commission Internationale de l'Eclairage*) L*ab* or L*uv* (luminance L and chrominance *uv* or *ab*), L*C*H* (lightness, chroma and hue) or Munsell colour space [Nickerson, 1976], CIECAM02 (*CIE colour appearance model*) [Moroney et al., 2002], HCL (hue, chroma and luminance) inspired by HSL and L*ab* [Sarifuddin and Missaoui, 2005].

Some of the colour naming models related to some of the above-mentioned colour spaces are cited here. Menegaz et al. [2007] presented a model for computational colour categorization and naming based on CIE L*ab* colour space and fuzzy partitioning. Weijer and Schmid [2007] presented a colour name descriptor based on CIE L*ab* colour space. Mojsilovic [2005] presented a computational model for colour categorization and naming and extraction of colour composition based on CIE L*ab* and HSL colour spaces. Seaborn et al. [2005] defined fuzzy colour categories based on Musell colour space (L*C*H). Liu et al. [2004] converted the dominant colour of a region (in HSV space) to a set of 35 semantic colour names some of them related to natural scene images like *sky blue* or *grass green*. Stanchev et al. [2003] defined 12 fundamental colours based on the L*uv* colour space and used Johannes Itten's colour theory to define light-dark contrast, warm-cool contrast, etc. Corridoni et al. [1998] presented a model for colour naming based on the HSL colour space and also introduced some semantic connotations as *warm/cool* or *light/dark* colours. Lammens [1994] presented a computational model for colour perception and colour naming based on CIE XYZ, CIE L*ab* and NPP colour spaces. Berk et al. [1982] defined the well-known colour naming system or CNS which divides HSL space into 627 distinct colours: the hue (H) value is divided into 10 basic colours and saturation (S) and lightness (L) are adjectives signifying the richness and brightness of the colour.

All these studies have inspired our model for qualitative colour description. However, we have defined our own model based on the requirements of our application.

### 3.4.2 A New Qualitative Colour Description (QCD)

Our approach translates the Red, Green and Blue (RGB) colour channels of the centroid of each segmented object in an image into coordinates of Hue, Saturation and Lightness (HSL) colour space (shown in Figure 3.6) in order to give a name to the perceptual colour of the object.

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 colors and corresponds to dominant color as perceived by a human and takes values between 0 and 360; saturation corresponds to the relative purity or the quantity of white light that is mixed with hue and takes values between 0 and 100; and luminance refers to the amount of light in a colour and takes values between 0 and 100. Furthermore, as W3C mentions[1], 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 HSV, considering the value colour coordinate (V) at the maximum, it goes from saturated colour to white, which is not intuitive, whereas in HSL, the saturation colour coordinate (S) takes values from fully saturated colour to the equivalent grey; and (ii) in HSV, the value colour coordinate (V) only goes from black to the chosen hue, while in HSL, the lightness colour coordinate (L) always spans the entire range from black through the chosen hue to white.

Therefore, HSL colour space is suitable for dividing into intervals of values corresponding to colour names and also intuitive for adding semantic labels to these names in order to refer to the richness (saturation) or the brightness of the colour (lightness) [Sarifuddin and Missaoui, 2005].

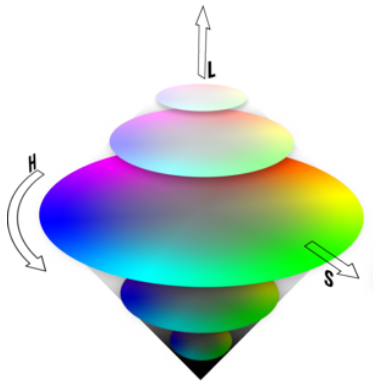



Figure 3.6: HSL colour space

From the HSL colour coordinates obtained, a reference system for qualitative colour description is defined as: QCRS = {UH, US, UL, $QC_{LAB_{1..M}}$, $QC_{INT_{1..M}}$} where UH is the Unit of Hue; US is the Unit of Saturation; UL is the Unit of Lightness; $QC_{LAB_{1..M}}$ refers to the qualitative labels related to colour distributed in M colour sets; and $QC_{INT_{1..M}}$ refers to the three intervals

[1]See the CSS3 specification from the W3C (http://www.w3.org/TR/css3-color/#hsl-color)

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 Figure 3.6). 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 and as a consequence of chosing HSL as our colour space, our model defines $M = 5$ colour sets: (1) grey colours, (2) rainbow colours, (3) pale rainbow colours, (4) light rainbow colours and (5) dark rainbow colours.

For our approach, the $\mathrm{QC}_{LAB_{1..M}}$ and $\mathrm{QC}_{INT_{1..M}}$ are the following:

$\mathrm{QC}_{LAB_1}$ = {*black, dark_grey, grey, light_grey, white*}
$\mathrm{QC}_{INT_1}$ = {[0, 20), [20, 30), [30, 40), [40, 80), [80, 100) ∈ UL / [0, 360] ∈ UH ∧ [0, 20] ∈ US}

The colour name set for the grey scale is defined by $\mathrm{QC}_{LAB_1}$ whose corresponding intervals of values in HSL are determined by $\mathrm{QC}_{INT_1}$. All the colours in this set can take any value of hue, values of saturation between 0 and 20 and different values of lightness that determine the colour names defined for this set. Note that the saturation coordinate of the HSL colour space (US) determines if the colour corresponds to the grey scale or to the rainbow scale.

$\mathrm{QC}_{LAB_2}$ = {*red, yellow, green, turquoise, blue, purple, pink*}
$\mathrm{QC}_{INT_2}$ = {(335, 360] ∧ [0, 40], (40, 80], (80, 160], (160, 200], (200, 260], (260, 297], (297, 335] ∈ UH / (50, 100] ∈ US ∧ (40, 55] ∈ UL }

The colours in the rainbow scale are defined by the names in $\mathrm{QC}_{LAB_2}$ and they are considered the more saturated ones or the strong ones. In $\mathrm{QC}_{INT_2}$, their saturation can take values between 50 and 100, their lightness between 40 and 55 and the different values of hue are those that determine the colour names defined for this set.

$\mathrm{QC}_{LAB_3}$ = {*pale_* + $\mathrm{QC}_{LAB_2}$}
$\mathrm{QC}_{INT_3}$ = { (335, 360] ∧ [0, 40], (40, 80], (80, 160], (160, 200], (200, 260], (260, 297], (297, 335] ∈ UH / (20, 50] ∈ US ∧ (40, 55] ∈ UL }

The pale colour name set ($\mathrm{QC}_{LAB_3}$) is defined by adding the prefix *pale* to the colours defined for the rainbow scale ($\mathrm{QC}_{LAB_2}$). These colour names are defined in $\mathrm{QC}_{INT_3}$ by the same hue and lightness intervals. They differ from rainbow colours by their saturation, which can take values between 20 and 50.

$\mathrm{QC}_{LAB_4}$ = {*light_* + $\mathrm{QC}_{LAB_2}$}
$\mathrm{QC}_{INT_4}$ = { (335, 360] ∧ [0, 40], (40, 80], (80, 160], (160, 200], (200, 260], (260, 297], (297, 335] ∈ UH / (50, 100] ∈ US ∧ (55, 100] ∈ UL }

$\mathrm{QC}_{LAB_5}$ = {*dark_* + $\mathrm{QC}_{LAB_2}$}

$QC_{INT_5}$ = { (335, 360] ∧ [0, 40], (40, 80], (80, 160], (160, 200], (200, 260], (260, 297], (297, 335] ∈ UH / (50, 100] ∈ US ∧ (20, 40] ∈ UL}

The lightness coordinate (UL) determines the luminosity of the colour: dark and light colours are distinguished and given an explicit descriptor 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 55 and 100 and dark colours between 20 and 40. Note that 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.

Therefore, in our QCD approach, 10 basic colours are defined (*black, grey, white, red, yellow, green, turquoise, blue, purple, pink*) and adding the semantic descriptors *pale, light* and *dark*, a total of $5+7\cdot4 = 33$ colour names are obtained. Research by Conway [1992] showed that, although it may be strictly accurate, people tend not to describe a colour as *dark pale blue* and may even consider this a contradiction. Conway [1992] 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 and also, if a lightness and saturation modifier appear equally applicable to a particular colour, the saturation modifier should be chosen. This aspect is reflected in our model.

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

Finally, as an example, according to the previous definitions, if the colour of the object in Table 3.1 has as HSL colour coordinates [0, 0, 0], the colour name assigned to it would be one defined in the grey scale ($QC_{LAB_1}$) as *black*.

## 3.5 Topology

Topological relations are spatial relations that are invariant under topological transformations, such as translations, rotation and scaling [Egenhofer, 1989].

### 3.5.1 Topology in the Literature

In the literature, methods for representing topological relationships can be classified as based on intersections or connections.

The models that characterize topological relations as **intersections** are: the 4-Intersection model [Egenhofer and Franzosa, 1991], which represents eight topological relations (*disjoint, contains, inside, equal, meet, covers, coveredBy, overlap*) between two spatial objects in two dimensions by considering the intersections of the objects' interiors and exteriors; and the 9-Intersection model

[Egenhofer and Herring, 1991], which also considers the intersections of the boundaries of the objects.

The 4-Intersection model was extended by Clementini et al. [1993] to define the Dimension Extended Method (DEM), which considers the dimension of the intersections, and also in [Egenhofer et al., 1994] to model topological relations between two-dimensional objects with arbitrary holes.

The 9-Intersection model was extended [Clementini and Felice, 1997] to consider objects with broad or inderterminate boundaries. A calculus based method (CBD) was also defined for this intersection model [Clementini et al., 1993], based on an Object-Calculus [Clementini and Di Felice, 1993], which takes into account the dimension of the result of the intersection of points, lines and areas. The 9-Intersection model is also renamed as the $9^+$-Intersection model [Kurata and Egenhofer, 2007; Kurata, 2008] for representing topological relations between a directed line segment and a region in a 2D space, which is interesting for characterizing movement patterns of an agent with respect to a region. Finally, the 9-intersection model is also extended [Egenhofer, 2005] to include three new topological relations in a sphere (*attaches, embraces, entwined*) [Egenhofer and Vasardani, 2007] to model the 23 topological relations that exist between a hole-free and a single-holed region and to represent the topological relations which exist between a hole-free and a multi-holed region [Vasardani and Egenhofer, 2009].

The model that characterizes topological relations as **connections** (e.g. regions are connected if they share a point) is the Region Connection Calculus ($\mathcal{RCC}$-8) [Randell et al., 1992b; Cohn et al., 1994], which describes eight jointly exhaustive and pairwise disjoint (JEPD) topological relations between two non-empty spatial regions: *disconnected*, *externally connected*, *partially overlapping*, *equal*, *tangential proper part*, *non-tangential proper part*, *tangential proper part inverse*, and *non-tangential proper part inverse*.

The RCC-8 has been widely studied and extended in the literature: to consider regions with indeterminate boundaries [Cohn and Gotts, 1995]; to describe the topology of spherical regions [Gotts, 1996]; to create a coarser version of RCC-8, RCC-5, considering only the relations: *discrete from*, *partially overlapping*, *equal*, *proper part* and *proper part inverse* [Cohn et al., 1997]; to show that each RCC model leads to a Boolean algebra [Stell, 1999; Düntsch et al., 2001]; to give an extensional RCC-8 composition table [Li and Ying, 2003]; to generalize RCC so that it can admit not only continuous representations of space but also discrete ones by representing each region in space by finite steps from basic regions [Li and Ying, 2004; Xia and Li, 2006]; to generalize RCC so that the spatial relations can be fuzzy relations and connections can be defined in terms of closeness between fuzzy sets [Schockaert et al., 2008]; to determine how to reason with fuzzy RCC [Schockaert et al., 2009], etc.

### 3.5.2 Topological Model applied to QID Approach

In order to represent the topological relationships of the objects in the image, we have used the intersection models defined by Egenhofer and Franzosa [1991] and Egenhofer and Herring [1991]. For region configurations in $R^2$, the 4-intersection and the 9-intersection model provide the same eight relations: *disjoint, contains, inside, equal, meet, covers, coveredBy, overlap.*

Our main reasons for choosing this model are: (i) our approach considers

intersections of regions in the image (which are areas in $R^2$) and (ii) there is a relevant connection between human interpretation of spatial relationships and this model, as was concluded by Mark and Egenhofer [1992] after testing on human subjects. Approaches dealing with regions with holes are interesting [Egenhofer and Vasardani, 2007; Vasardani and Egenhofer, 2009]. However our approach does not distinguish regions inside other regions from regions with holes. Therefore, the simplest model for describing the topological regions contained in an image is used.

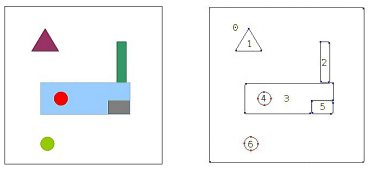
In order to represent the topological relationships of the objects in the image, the intersection model defined by Egenhofer and Franzosa Egenhofer and Franzosa [1991] for region configurations in $R^2$ is used. However, as information on depth cannot be obtained from digital images, the topological relations *overlap*, *coveredBy*, *covers* and *equal* defined by that model cannot be distinguished by the QID-approach and are all substituted by *touching*. For example, as Table 3.7 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, we express the location in space (invariant under translations, rotation and scaling) of an object A with respect to (wrt) the location in space of another object B (A wrt B), by using the following Topology Reference System (TRS):

$$T_{LAB} = \{disjoint,\ touching,\ completely\_inside,\ container\}$$

The QID-approach determines if an object is *completely_inside* another object (applying Jordan's curve theorem [Courant and Robbins, 1996]) or vice versa, if it is the *container* of another object. It also defines the *neighbours of level* of an object as all the other objects with 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.

Table 3.7: Drawing for exemplifying the topological relations between object 3 and the other objects within the image described by the QID-approach.

| Image | Topology Description |
|---|---|
|  | ...<br>[ 3, [Container, 0], [touching, 2, 5],<br>[disjoint, 1, 6], [completely_inside, 4]<br>],<br>... |

Finally, as an example, the topological situation of the blue object in the drawing in Table 3.7 is described as having: (i) one *container* (the image, Object 0); (ii) an object located *completedly 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).

## 3.6 Qualitative Orientation

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, since 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.

### 3.6.1 Qualitative Orientation in Literature

Qualitative orientation approaches for reasoning about locations appearing in the literature can be classified as based on projections [Guesgen, 1989; Jungert, 1993; Mukerjee and Joe, 1990a], or not based on projections [Freksa, 1992; Hernández, 1991; Frank, 1991; Ligozat, 1993; Moratz et al., 2000, 2005]. In projection-based models, the relative orientation of objects is obtained by using (orthogonal or non-orthogonal) projections of objects onto external axes, and then reasoning in one-dimension, by using Allen's temporal logic. However, when people orientate themselves, we never naturally think that our orientation or position is somewhere with respect to an external reference system and we usually include ourselves in the reference system. Therefore, as models not based on projections are more cognitive, they have been applied more often in applications dealing with a human-user interface. In these models, the space is divided into qualitative regions by means of reference systems, which are centred on the reference objects (i.e. the reference system is local and egocentric).

From orientation models not based on projections, some important and widely applied calculi have been developed in order to reason about orientations and point configurations: the FlipFlop calculus [Ligozat, 1993] extended by Scivos and Nebel [2001], which describes the position of a point C (the referent) in the plane with respect to two other points A (the origin) and B (the relatum); the Double Cross Calculus (DCC) [Freksa, 1992], which describes the direction of a point C (the referent) with respect to a point B (the relatum) as seen from a third point A (the origin); the Dipole Calculus (DC) [Moratz et al., 2000; Schlieder, 1995], which deals with the orientation of line segments of concrete length; and the Oriented Point Relation Algebra (OPRA) [Moratz et al., 2005], which deals with the orientation of line segments with infinitely small length at different levels of granularity.

### 3.6.2 Qualitative Orientation Models applied to QID

The QID-approach applies two kinds of qualitative orientation models: (i) the model by Hernández [1991] 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 double cross orientation model by Freksa [1992] 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 Hernández [1991] and point-based Freksa [1992] 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 organized as follows. The fixed orientation model applied is explained in Subsection 3.6.2.1 and its relative orientation model used is described in Subsection 3.6.2.2. Moreover, in Section 3.6.2.4, a comparison of the reference frames used by each model of orientation is given. Finally, as the orientation relation of each region in an image depend on the total number of regions contained within this image, Section 3.6.2.3 explains which orientation relations can be computed according to the number of objects with the same container.

#### 3.6.2.1 Qualitative Model of Fixed Orientation

A Fixed Orientation Reference System (FORS) is defined using the model by Hernández [1991], 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 is fixed to the upper edge of the image and divides the space into eight regions (Figure 3.7) which are labelled as:

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

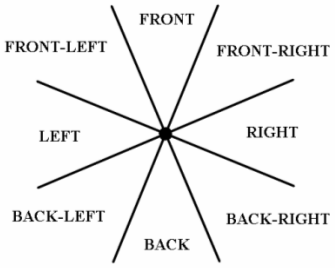
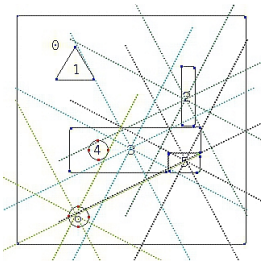



Figure 3.7: The qualitative orientation model by Hernández [1991].

In order to obtain the fixed orientation of each object wrt another object, our approach locates the centre of the FORS on the centroid of the reference object. 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.

Table 3.8: Drawing for exemplifying the fixed orientation relations of object 1 described by the QID-approach.

| Image | Fixed Orientation Description |
|---|---|
|   | ...<br>[ 1, [Container, 0], [Orientation wrt 0: front_left],<br>[disjoint, 2, 3, 5, 6],<br>[Orientation wrt Neighbours: [2, left], [3, front_left],<br>[5, front_left], [6, front]],<br>...<br>[Vertices_Orientation, front, back_right, back_left], ],<br>],<br>... |

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 3.8 is described as located: *front-left* wrt its container (the image); *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).

#### 3.6.2.2 Qualitative Model of Relative Orientation

A Relative Orientation Reference System (RORS) is defined using the double cross orientation model by Freksa [1992]. 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 3.8). This model divides the space into 15 regions, which are labelled as:

$$RO_{LAB} = \{\textit{left\_front (lf), straight\_front (sf), right\_front (rf), left (l), identical\_front (idf), right (r), left\_middle (lm), same\_middle (sm), right\_middle (rm), identical\_back\_left (ibl), identical\_back (ib), identical\_back\_right (ibr), back\_front (bf), same\_back (sb), back\_right (br)}\}$$

In order to obtain the relative orientation of an object, 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.

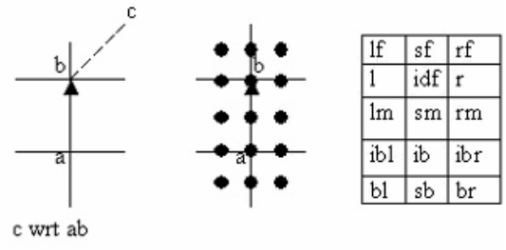
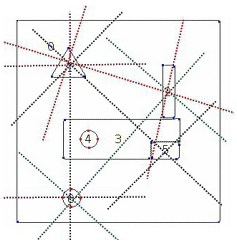



Figure 3.8: The qualitative orientation model by Freksa [1992] and its iconical representation: *l is left, r is right, f is front, s is straight, m is middle, b is back and i is identical.*

Note that the RO information would change if there is a significant translation of any of the objects in the image, whereas would remain invariant to image and object rotations.

Table 3.9: 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]],<br>[ 2, [RO: [[1,5], lm], [[1,6], lm], [[5,6], br]]],<br>[ 3, [RO: [[1,6], lm, rm]]<br>[ 4, [RO: - ],<br>[ 5, [RO: [[1,2], rm, rf], [[1,6], lm], [[2,6], lm]],<br>[ 6, [RO: [[1,2], rm], [[1,3], rf], [[1,5], rm], [[2,3], lf], [[2,5], lf], [[3, 5], rm, br]], |

As an example, some relative orientations (RO) that can be extracted from the objects in the drawing in Table 3.9 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 3.9 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).

#### 3.6.2.3 Organization of Orientation Relations

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 3.10.

Table 3.10: Spatial features described depending on the number of objects at each level.

| ***Spatial Features Described*** | | ***Objects within the same container*** | | |
|---|---|---|---|---|
| | | **1** | **2** | **> 2** |
| Wrt its Container | Topology | x | x | x |
| | FORS | x | x | x |
| Wrt its Neighbours | Topology | - | x | x |
| | FORS | - | x | x |
| | RORS | - | - | x |

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 simplified description of the image.

#### 3.6.2.4 Reference Frames of the FORS and the RORS

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 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

(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).

## 3.7 Structure of the Qualitative Image Description (QID)

Summarizing, the structure of the QID obtained for any digital image is that presented in the diagram in Figure 3.9. For each region in the image, its visual and spatial features are described qualitatively. As explained before, the shape and colour of each region are considered absolute properties that depend on the region itself, whereas the topology and orientation of each region is described with respect to other regions: their containers, neighbours touching, neighbours disjoint, etc.

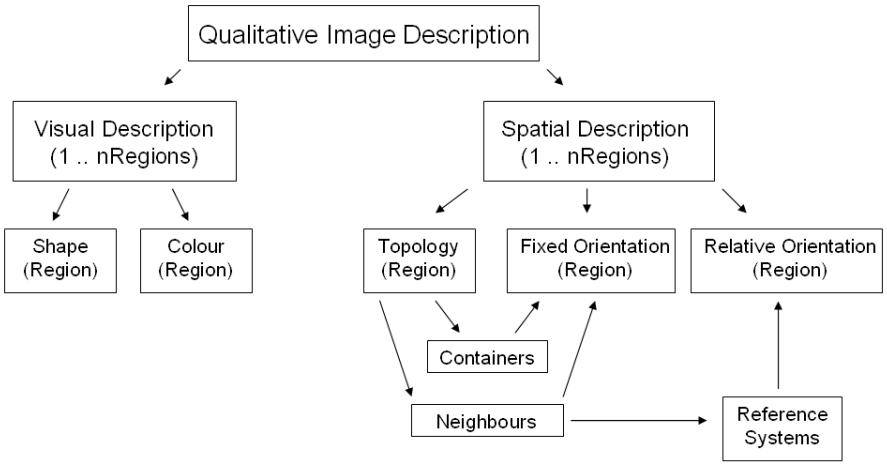



Figure 3.9: Structure of the QID obtained by our approach.

Therefore, the general structure of the qualitative image description (QID) provided by our approach is defined as a set of qualitative tags such that:

**QID** (IdImage) = [ SpatialDescription(Regions), VisualDescription(Regions) ]

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:

**SpatialDescription** (1 .. NRegions) = [ IdRegion, Topology(Container), FixedOrientation(Container), Topology(Region), FixedOrientation(Neighbours), RelativeOrientation(RSs) ]

Topology(Container) = [ *Container*, IdContainer ]
FixedOrientation(Container) = [ *Orientation wrt* IdContainer: FixedOrientationTags ]
Topology(Region) = [ *touching*(IdRegions), *disjoint*(IdRegions), *completedly_inside*(IdRegions) ]
FixedOrientation (1 .. NNeighbours) = [ *Orientation wrt Neighbours*: [ IdNeighbour, FOs]]
RelativeOrientation (1 .. NRSs) = [*Relative Orientation wrt Neighbours Disjoint*: [RSs, ROs]]
ReferenceSystem = [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:

**VisualDescription** (1.. NRegions) = [ IdRegion, QCD, QSD(RPs), Orientation(RPs) ]

QSD (1 .. nRP) = [KEC, A or TC, L, C] where,
KEC ∈ {*line-line, line-curve, curve-line, curve-curve, curvature-point*};
A ∈ {*very_acute, acute, right, obtuse, very_obtuse*};
TC ∈ {*very_acute, acute, semicircular, plane, very_plane*};
C ∈ {*convex, concave*};
L ∈ {*msh, hl, qsh, sl, ql, dl, ml*}
QCD ∈ {*black, dark_grey, grey, light_grey, white, red, yellow, green, turquoise, blue, purple, pink, pale_red, ..., pale_pink, light_red, ..., light_pink, dark_red, ..., dark_pink*}
Orientation (1 .. nRP) = [FOs]

## 3.8 Our Computational Model for the QID

As Figure 3.10 illustrates, our model for QID has been implemented in a SW application that first obtains the relevant regions of any image by applying image processing algorithms and then describes the visual and spatial features of these regions by applying qualitative models. The visual features of the regions in the image are described by the qualitative models of shape and colour given in Sections 3.3 and 3.4, respectively. The spatial features of the regions in the image are described by the qualitative models of topology and orientation given in Sections 3.5 and 3.6, respectively.

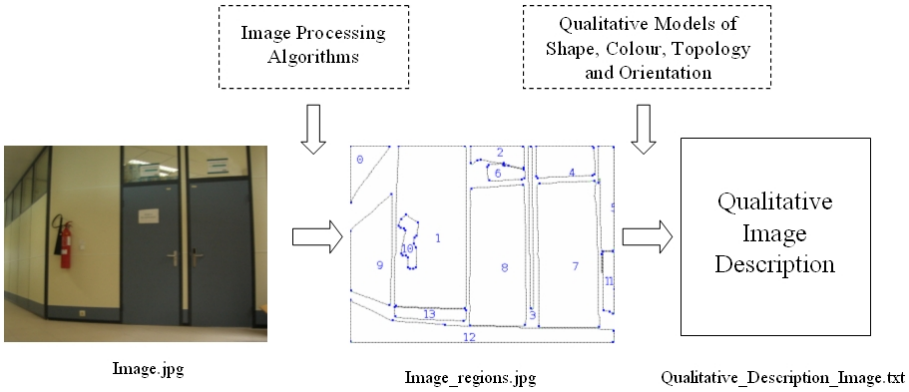



Figure 3.10: Schema of our approach for qualitative image description.

Section 3.8.1 discusses the approaches used to extract the main regions of any digital image. Section 3.8.2 outlines the algorithm followed by our computational approach and Section 3.7 explains the structure of the qualitative description obtained.

### 3.8.1 Obtaining the Relevant Regions of Any Digital Image

*Region segmentation* is defined by Palmer [1999] 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:

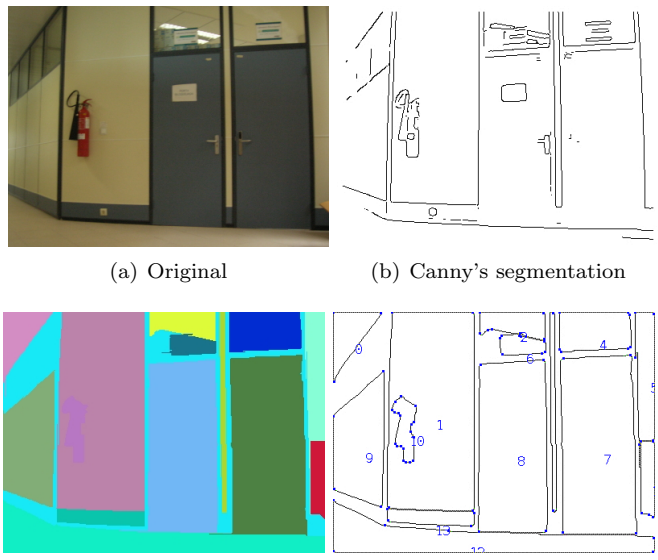
- *boundary-based approaches*, in which the visual system detects differences (or gradients) in local visual properties that divide one region from another. The approaches that first detect the edges or boundaries in a digital image and then, from the obtained boundaries, extract the regions within it are included in this group. An example is the well-known segmentation approach by Canny [1986].
- *region-based approaches*, which consider that in an image, different colours or textures usually indicate different regions of interest to the human eye. The approaches 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 [2004].

As mentioned by Felzenszwalb and Huttenlocher [2004], the problems of image segmentation and grouping remain great challenges for computer vision because, to be 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 nearly linear time in the number of image pixels.

The segmentation method by Felzenszwalb and Huttenlocher [2004] achieves the above goals and it also preserves detail in low-variability image regions while ignoring detail in high-variability regions by adjusting its segmentation parameters: $\sigma$, 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.

Figure 3.11: Image from the robot environment segmented by (b) the boundary based method by Canny [1986] and (c) the region based method by Felzenszwalb and Huttenlocher [2004].



(a) Original (b) Canny's segmentation

(c) Felzenszwalb's segmentation: $\sigma = 0.4$, $k = 500$ and $min = 1000$ (d) Extraction of boundaries from Felzenszwalb's

Generally, image region-based segmentation approaches are considered more cognitive than boundary-based segmentation approaches because the extracted edges are defined by the boundaries between colour regions and all of these regions are closed. In Figure 3.11 the results of both approaches applied to the segmentation of the same image can be compared. While the segmentation approach by Canny [1986] obtains open edges, the boundaries extracted from

the approach by Felzenszwalb and Huttenlocher [2004] are all closed.

However, our approach is not dependent on the region segmentation approach used. Therefore, the most convenient approach from the literature can be selected depending on the application.

### 3.8.2 The SW Application Algorithm

The computational approach for the QID-approach is outlined in Algorithm 8 which is described next.

**Algorithm 8** 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 aforementinoned, 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 characterize its shape are extracted;
- its container ($R.Container$) is obtained and, as an inverse relationship, the current region is located *completely inside* of 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 of its boundary ($R.Points$);
- the fixed orientation of each of the relevant points of the region ($R.Points$) is described with respect to its centroid ($R.Centroid$);
- its fixed orientation of each region with respect to the centroid of its container ($R.Container.Centroid$) is obtained;
- its neighbours 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 of level (touching or disjoint) is obtained;
- for each pair of neighbours disjoint ($R.Neighbours\_Disjoint$) of the current region a reference system ($RS$) is build and the relative orientations of the current region wrt all these reference systesm are obtained.

The computational cost of the algorithm is $O(PR^3)$ where $P$ is the biggest 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-Algoritm peaks when a lot of regions are extracted in the image and those regions have irregular boundaries described by a high number of relevant points.

## 3.9 Experimentation and Results

Our approach for qualitative description of images is applied to: (i) Scenario I: the description of images of the world captured by a webcam located on a mobile robot (Section 3.9.1), and (ii) Scenario II: the description of mosaic images captured by an industrial camera located on a platform, which is used by a robot arm to assemble mosaics automatically (Section 3.9.2).

### 3.9.1 Scenario I

Scenario I is the description of images of the world, mainly images of visual landmarks, captured by a webcam Logitech Quickcam Pro 9000[2] with a Carl Zeiss optic lense and 2 Megapixel resolution located on the top of ActiveMedia Pioneer 2 dx mobile robot [3], shown in Figure 3.12.

Figure 3.12: Our approach applied to Scenario I: describing an image of our University corridor taken by a webcam located on a Pioneer 2 mobile robot.

(a) Webcam located on a mobile robot

(b) Image obtained 400x300

(c) Felzenszwalb's segmentation: $\sigma = 0.4$, $k = 500$ and $min = 1000$

For this scenario, a region-based segmentation method is used, specifically the one by Felzenszwalb and Huttenlocher [2004]. This method extracts the regions of interest in an image by a graph-based region segmentation method based on intensity differences. This segmentation method is used by our approach because it captures the perceptually important regions in an image, it is highly efficient, running in time nearly linear in the number of image pixels and it also preserves detail in low-variability image regions while ignoring detail in high-variability regions by adjusting its segmentation parameters: $\sigma$, used to smooth the input image before segmenting it; $k$, the value for the threshold function in segmentation, the larger the value, the larger components in the result; and $min$, minimum size of the extracted regions in pixels enforced by post-processing.

Table 3.11 presents an excerpt of the qualitative description of the digital image captured by the Pioneer 2 robot camera, shown in Figure 3.12. Specifically, this table gives the qualitative spatial description of regions 0, 1 and 10 and the qualitative visual description of regions 7 and 10.

The spatial description of region 1 can be intuitively read as follows: its *container* is the Image and it is located wrt the Image at *front, front_left, back,*

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

Table 3.11: An excerpt of the qualitative description obtained for the image captured by the Pioneer 2 webcam in Figure 3.12.

[**SpatialDescription**,
[ **0**, [Container, Image], [Orientation wrt Image: front_left],
[disjoint, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13],
[Orientation wrt Neighbours: [1, front_left], [2, left], [3, front_left], [4, left],
(...) [9, front], [11, front_left], [12, front_left, front], [13, front, front_left]]
[Relative Orientation wrt Neighbours Disjoint: [[1, 2], lm, bl], [[1, 3], bl], [[1, 4], bl],
(...) [[7, 8], rf], [[7, 9], rm, rf ], [[7, 11], bl] (...) [[12, 13], rf]]
]
[ **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], [3, back_right], [4, left, back_left], [5, left, back_left], (...)
[Relative Orientation wrt Neighbours Disjoint: [[0, 4], rm], (...) [[4, 7], br, rf],
(...) [[7, 8], rf, lf], [[7, 9], rm, lm], [[7, 11], bl, br], (...) [[11, 12], rm, rf]]
]
(...)
[ **10**, [[Container, 1] [Orientation wrt 1: left, back_left, back],
[None Neighbours of Level] ]
(...)
]
[**VisualDescription**,
[ **7**, dark_grey, quadrilateral,
[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, mixed-shape,
[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 (...) ]],
],
(...)
]

*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 our 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

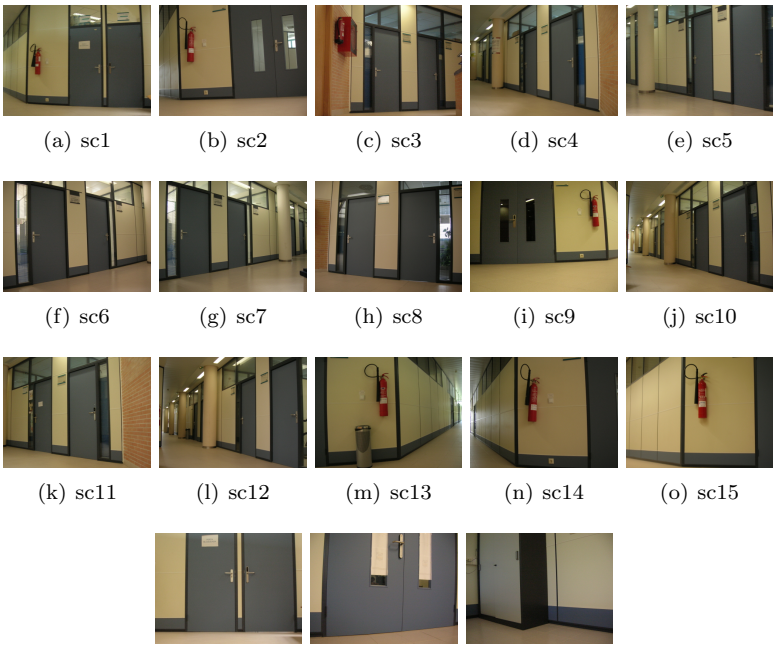
region 2 it is *left, back, back_left*, wrt region 3 it is *back_right* and in a similar way, the fixed orientation of region 1 is described wrt all its *neighbours of level*. Finally, the relative orientation wrt the disjoint neighbours of region 1 is given: from region 0 to region 4, region 1 is located *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)*. Note that region 0 is described similarly.

The spatial description of region 10 is also given in Table 3.11: 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 3.11 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 the computational time, it should be noted that, for the image shown in Figure 3.12, the time of execution for the extraction 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 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.

Figure 3.13: Images of indoor scenes used for testing our approach.



(a) sc1 (b) sc2 (c) sc3 (d) sc4 (e) sc5

(f) sc6 (g) sc7 (h) sc8 (i) sc9 (j) sc10

(k) sc11 (l) sc12 (m) sc13 (n) sc14 (o) sc15

(p) sc16 (q) sc17 (r) sc18

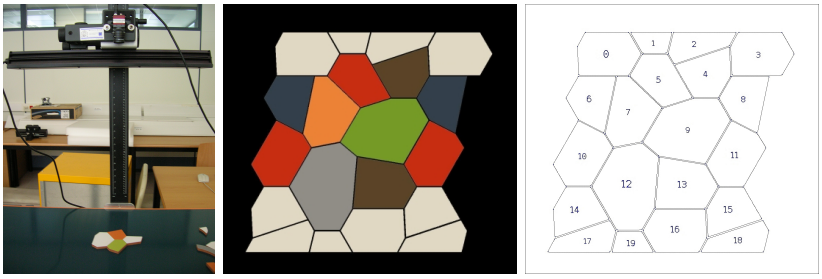
The images of indoor scenes captured by a webcam located on a Pioneer 2 robot shown in Figure 3.13 have been used to test our approach. The segmentation result of each image, the execution times and the QID obtained are available on-line[4]. Note that the QID time depends on the computational cost of the algorithm is $O(PR^3)$ where $P$ is the biggest number of relevant points that define a region in the image and $R$ is the total number of regions in the image.

### 3.9.2 Scenario II

The proposed approach is also applied to Scenario II, that is, to process mosaic images captured by an industrial camera AVT-Guppy F033C located on a platform from which a robot arm picks and places tile pieces (Figure 3.14).

For this scenario, a boundary-based segmentation approach is used, specifically the one presented by Canny [1986]. This segmentation method is chosen because it is fast and obtains good results in this scenario because the boundaries of the objects are clearly defined and distinguished as they are made usually by straight edges or simple curves.

Figure 3.14: Our approach applied to Scenario II: describing an image of a tile mosaic taken by an industrial camera located in a platform which is used by a robot arm to assemble tile mosaics.



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

(b) Image obtained 500x500

(c) Canny's segmentation

Table 3.12 presents an excerpt of the qualitative description of the digital image of a tile mosaic taken by an industrial camera used by a robot arm to assemble tile mosaics, shown in Figure 3.14. Specifically, this table shows the qualitative spatial description of regions 0 and 9 and the qualitative visual description of regions 0 and 1.

The spatial description of region 0 can be intuitively read as follows: its *container* is the Image and it is located wrt to the Image at *front_left, front*. Its *touching neighbours* are the regions 1, 2, 6 and 7. (Note that some of these are not technically touching but are closer to region 0 than the threshold determined for our application). Its *disjoint neighbours* are the rest of the pieces of the mosaic and it does not have any region completely inside. The *fixed orientation*

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

Table 3.12: An excerpt of the qualitative description obtained for the mosaic image captured by the industrial camera in Figure 3.14.

[**SpatialDescription**,
[ **0**, [Container, Image], [Orientation wrt Image: front_left, front],
[touching, 1, 5, 6, 7], [disjoint, 2, 3, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19],
[Orientation wrt Neighbours:[1, left, back_left], [2, left], [3, left], [4, left, front_left],
(...) [15, front_left], [16, front, front_left], [17, front], [18, front_left], [19, front]],
[Relative Orientation wrt Neighbours Disjoint: [[2, 3], br, bl] , (...) [[3, 8], rm, rf],
(...) [[10, 15], bl, lm] (...) [[16, 17], rm] (...) [[18, 19], rf, rm]]
]
(...)
[ **9**, [Container, Image], [Orientation wrt Image:front_right, right, back_right, back_left,
left, front],
[touching, 4, 5, 7, 8, 11, 12, 13], [disjoint, 0, 1, 2, 3, 6, 10, 14, 15, 16, 17, 18, 19],
[Orientation wrt Neighbours: [0, back_right, right, back], [1, back_right, back],
[2, back, back_left], [3, back_left, back], [4, back_left, back, back_right],
(...) [10, front_right, right], [11, front_left, front, left], [12, front_right, front, right],
(...) [16, front], [17, front_right], [18, front, front_left], [19, front, front_right]],
[Relative Orientation wrt Neighbours Disjoint: [[0, 1], rf, rm], (...) [[3, 16], rm, lm]
(...) [[6, 16], lm] (...) [[10, 14], bl], [[10, 15], lm] (...) [[17, 19], lf], [[18, 19], rm]]
]
(...)
]
[**VisualDescription**,
[ **0**, white, hexagon,
[Boundary_Shape,
[line-line, obtuse, much_shorter, convex],
[line-line, obtuse, a_bit_longer, convex],
[line-line, obtuse, similar_length, convex],
[line-line, obtuse, half_length, convex],
[line-line, right, a_bit_longer, convex],
[line-line, very_obtuse, a_bit_longer, convex]],
[Vertices_Orientation, front_left, front_right, right, back_right, back_left, left],
],
[ **1**, white, quadrilateral,
[Boundary_Shape,
[line_line, acute, half_length, convex],
[line_line, acute, double_length, convex],
[line_line, obtuse, similar_lenght, convex],
[line_line, obtuse, similar_lenght, convex]],
[Vertices_Orientation,
left, right, back_right, back_left],
],
(...)
]

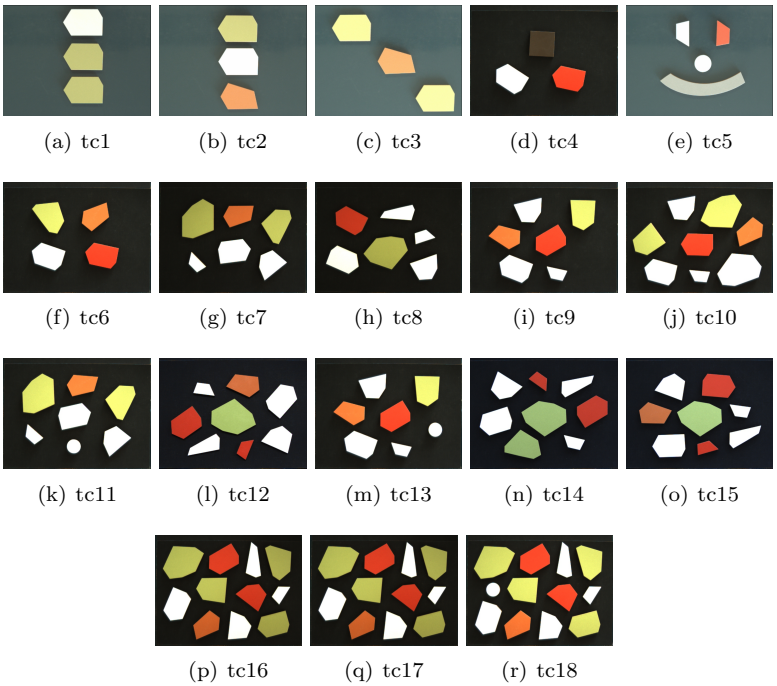
of region 0 wrt region 1 is *left, back_left*, wrt region 2 it is *left*, wrt region 3 it is also *left* and in a similar way, the fixed orientation of region 0 is described

wrt all its *neighbours of level*. Finally, the relative orientation wrt the disjoint neighbours of region 0 is given: from region 2 to region 3, region 0 is located *back_right (br), back_left (bl)*; from region 3 to region 8, region 0 is located *right_middle (rm), right_front (rf))*, from region 18 to region 19, region 0 is located *right_front (rf), right_middle (rm)*. Note that the spatial description of region 9, also given in Table 3.12, is explained similarly.

The visual description of region 0 in Table 3.12 shows that its colour is *white* and that the shape of its boundary is a hexagon, qualitatively described as composed of four *line-line* segments whose angles are four *obtuse* and *convex*, one *right* and *convex* and one *very_obtuse* and whose compared distances are respectively *much_shorter, a_bit_longer, similar_length, half_length, a_bit_longer, a_bit_longer*. Finally, the orientation of its vertices with respect to the centroid of the region is in a clockwise direction: *front_left, front_right, right, back_right, back_left, left*. Note that the visual description of region 2 is explained similarly.

With respect to the computational time, it should be noted that, for the image shown in Figure 3.14, 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 around 1.66 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.

Figure 3.15: Images of tile compositions used for testing our approach.



(a) tc1 (b) tc2 (c) tc3 (d) tc4 (e) tc5

(f) tc6 (g) tc7 (h) tc8 (i) tc9 (j) tc10

(k) tc11 (l) tc12 (m) tc13 (n) tc14 (o) tc15

(p) tc16 (q) tc17 (r) tc18

The images of tile compositions captured by an industrial camera located above the table where a robot arm assemble tile mosaics shown in Figure 3.15

have been used to test our approach. The segmentation result by Canny's method of each image, the execution times and the QID obtained are available on-line[5]. Note that the QID time depends on the computational cost of the algorithm is $O(PR^3)$ where $P$ is the biggest number of relevant points that define a region in the image and $R$ is the total number of regions in the image.

## 3.10 Conclusions

In this chapter, a computational approach for qualitative description of any digital image is presented. Our approach gives a visual and 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.

The QID approach is independent of the image segmentation method used. Successful results have been obtained using a boundary-based segmentation method [Canny, 1986] and a region-based segmentation method [Felzenszwalb and Huttenlocher, 2004], depending on the scenario to which our approach is applied.

The QID approach has been tested in two scenarios: (i) the description of images of the world, mainly images of visual landmarks, captured by a webcam located on the top of a Pioneer 2 mobile robot, and (ii) the description of mosaic images captured by an industrial camera located on a platform, which is used by a robot arm to assemble mosaics automatically. In both scenarios promising results have been obtained.

As future work, we intend to: (1) combine our model, which can describe unknown objects in the robot world, with an invariant feature detector, such as SIFT [Lowe, 2004] or Harris-Affine, Hessian-Affine or MSER [Mikolajczyk et al., 2005], for describing known objects; (2) compare the descriptions obtained by the QID-approach with those produced by human beings in order to study how to improve the cognitive perspective of our approach; (3) define a grammar and translate the QID into natural language; (4) define a three-dimensional QID incorporating the depth information of the images which is easily obtained by new devices such as a Kinnect sensor.

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

# Chapter 4

# An Ontology for Qualitative Image Description

Extracting semantic information from images as human beings can do is still an unsolved problem in computer vision. The approach presented in Chapter 3 can describe any digital image using qualitative information that is both visual (e.g. shape, colour) and spatial (e.g. topology, orientation). However, describing any image qualitatively and interpreting it in a meaningful way remains a challenge, and the association of meaning with representations obtained by robotic systems, also known as the *symbol-grounding problem*, is still a prominent issue within the field of Artificial Intelligence (AI) [Williams, 2008].

In this chapter, we present a small step forward in this area. We use ontologies to give a formal meaning to the qualitative labels used to describe images captured by a robot camera in indoor environments.

An ontology is defined as *"a formal specification of a shared conceptualization"* [Borst et al., 1997] that provides 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 provide a common vocabulary and meaning to allow these applications to communicate with each other [Guarino, 1998].

In our scenario, ontologies will provide our system with: (i) an explicit representation of knowledge inside the robot, (ii) a standard language to represent and communicate knowledge between agents and (iii) new knowledge inferred by the reasoners from the ontology facts. The semantic information extracted from these images will be used later to support robot self-localization and navigation.

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 2[1] [Horrocks et al., 2003; Cuenca Grau et al., 2008], which is based on the description logic $\mathcal{SROIQ}$ [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.

[1]Ontology Web Language: `http://www.w3.org/TR/owl2-syntax/`

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 the knowledge of the world is incomplete. In this chapter, the suitability of the OWA for our domain is analyzed and the cases where 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 chapter is organized as follows. Section 4.1 describes the related work. Section 4.2 describes the ontology schema, and the ontology facts provided by our approach, and it also deals with the OWA problem. Section 4.3 details our approach and it also presents the results of its application to the description of digital images of our robot environment. Finally, Section 4.4 explains our conclusions and future work.

## 4.1 Related Work

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 Moller, 2008; Dasiopoulou and Kompatsiaris, 2010]. Neumann and Moller [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 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 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 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.

## 4.2 Giving Meaning to Qualitative Description of Images

Our QID approach (presented in Chapter 3) 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.

The aim of using a 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 at University Jaume I 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[2], a DL-based ontology to represent qualitative description of images and how we have dealt with the Open World Assumption (OWA) [Hustadt, 1994] in order to infer the expected knowledge (see Section 4.2.2).

### 4.2.1 Ontology Constructors and Three-Layer Representation

For our approach, we adopted OWL 2.0 [3] [Horrocks et al., 2003; Cuenca Grau et al., 2008] as the ontology language for representing the knowledge of QImageOntology and we selected Protégé 4[4] as the ontology editor for developing QImageOntology. Additionally we chose to use the DL reasoner HermiT[5] for classifying new captured images and for inferring new knowledge.

Table 4.1 gives a subset of the main OWL 2.0 axioms and concept constructors. Note that the description logic $\mathcal{SROIQ}$ [Horrocks et al., 2006] provides the logical underpinning for OWL 2.0. OWL Axioms define the relations between entities (classes, properties and individuals), whereas OWL concept constructors characterize the set of individuals that belongs to a specific class.

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).

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

[2]Available at: `http://krono.act.uji.es/people/Ernesto/qimage-ontology/`

[3]Ontology Web Language: `http://www.w3.org/TR/owl2-syntax/`

[4]Protégé: `http://protege.stanford.edu`

[5]HermiT: `http://hermit-reasoner.com/`. A custom complied version of Protégé4.1 that includes HermiT 1.1: `http://hermit-reasoner.com/download/protege4_1/CustomProtege4_1.zip` is suggested to the interested reader

Table 4.1: Some OWL 2 Axioms and Concept Constructors. $C, D$ are complex concepts, $R$ denotes an *atomic role*, $A$ represents an *atomic concept* and $a, b$ *individuals*.

| **OWL Axioms** | | |
|---|---|---|
| Global Concept Inclusion | $C \sqsubseteq D$ | $Right\_Triangle \sqsubseteq Triangle$ |
| Equivalence | $C \equiv D$ | $WhiteObject \equiv Object\_type \sqcap \ni has\_colour.\{white\}$ |
| Disjointness | $C \sqsubseteq \neg D$ | $Circle \sqsubseteq \neg Triangle$ |
| Property Domain | $\exists R.\top \sqsubseteq C$ | $\exists has\_point.\top \sqsubseteq Shape\_type$ |
| Property Range | $\top \sqsubseteq \forall R.C$ | $\top \sqsubseteq \forall has\_point.Point\_type$ |
| Transitive Property | $\mathsf{Trans}(R)$ | $\mathsf{Trans}(is\_container\_of)$ |
| Inverse property | $\mathsf{Inv}(R)$ | $\mathsf{Inv}(is\_disjoint)$ |
| Class Assertion | $C : a$ | $Colour\_type : red$ |
| Property Assertion | $R(a, b)$ | $isContainerOf(img1, obj1)$ |
| Negative Property Assertion | $\neg R(a, b)$ | $\neg isContainerOf(img2, obj1)$ |
| Extensional class definition | $C \sqsubseteq \{a, b, c\}$ | $Colour\_type \sqsubseteq \{red, white\}$ |
| Different Individuals | $a \neq b$ | $white \neq red$ |
| **OWL Concept Constructors** | | |
| Top Concept | $\top$ | $Thing$ |
| Atomic Class | $A$ | $Circle, Colour\_type$ |
| Negation | $\neg C$ | $\neg Triangle$ |
| Intersection | $C \sqcap D$ | $Triangle \sqcap \forall has\_angle.RightAngle$ |
| Union | $C \sqcup D$ | $Obtuse \sqcup Very\_Obtuse$ |
| Existential | $\exists R.C$ | $\exists isContainerOf.Object$ |
| Value Restriction | $\ni R.\{a\}$ | $\ni has\_colour.\{white\}$ |
| Universal | $\forall R.C$ | $\forall is\_left.Object\_type$ |
| Min Cardinality | $\geqslant n\, R.C$ | $\geqslant 3\, has\_point.Point\_type$ |
| Max Cardinality | $\leqslant n\, R.C$ | $\leqslant 2\, has\_angle.Angle$ |
| Exact Cardinality | $= n\, R.C$ | $= 4\, has\_point.Line\_Line$ |
| Nominals | $\{a, b, c\}$ | $\{point1, point2, point3\}$ |

- 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*[6] 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.

[6]We 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]

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 $\mathcal{SHOIQ}$ 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 4.2: 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 4.2: An excerpt of the Reference Conceptualization of QImageOntology

| | |
|---|---|
| $\alpha_1$ | $\mathsf{Image\_type} \sqsubseteq \exists \mathsf{is\_container\_of}.\mathsf{Object\_type}$ |
| $\alpha_2$ | $\mathsf{Object\_type} \sqsubseteq \exists \mathsf{has\_colour}.\mathsf{Colour\_type}$ |
| $\alpha_3$ | $\mathsf{Object\_type} \sqsubseteq \exists \mathsf{has\_fixed\_orientation}.\mathsf{Object\_type}$ |
| $\alpha_4$ | $\mathsf{Object\_type} \sqsubseteq \exists \mathsf{is\_touching}.\mathsf{Object\_type}$ |
| $\alpha_5$ | $\mathsf{Object\_type} \sqsubseteq \exists \mathsf{has\_shape}.\mathsf{Shape\_type}$ |
| $\alpha_6$ | $\mathsf{Shape\_type} \sqsubseteq \geqslant 3\, \mathsf{has\_point}.\mathsf{Point\_type}$ |
| $\alpha_7$ | $\mathsf{Quadrilateral} \sqsubseteq \mathsf{Shape\_type} \sqcap = 4\, \mathsf{has\_point}.\mathsf{line\_line}$ |
| $\alpha_8$ | $\mathsf{is\_left} \sqsubseteq \mathsf{has\_fixed\_orientation}$ |
| $\alpha_9$ | $\mathsf{Colour\_type} : \mathsf{red}$ |

Table 4.3 represents an excerpt from the *contextualized knowledge* of QImageOntology, where four objects are characterized:

- 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;
- 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;
- the definition of an office door (*UJI_Office_Door*) as a *grey* or *dark_grey* quadrilateral object located inside the image;
- the definition of 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 4.3: An excerpt of the Contextualized Descriptions of QImageOntology

| | | | |
|---|---|---|---|
| $\beta_1$ | UJI_Wall | $\equiv$ | Object_type $\sqcap$ $\exists$has_shape.Quadrilateral $\sqcap$ $\exists$is_completely_inside.Image_type $\sqcap$ ($\ni$ has_colour.{pale_yellow} $\sqcup$ $\ni$ has_colour.{dark_yellow} $\sqcup$ $\ni$ has_colour.{pale_red} $\sqcup$ $\ni$ has_colour.{light_grey}) |
| $\beta_2$ | UJI_Floor | $\equiv$ | Object_type $\sqcap$ $\exists$is_completely_inside.Image_type $\sqcap$ ($\ni$ has_colour.{pale_red} $\sqcup$ $\ni$ has_colour.{light_grey}) $\sqcap$ $\exists$is_back.Image $\sqcap$ $\neg$($\exists$is_front.Image) |
| $\beta_3$ | UJI_Office_Door | $\equiv$ | Object_type $\sqcap$ $\exists$has_shape.Quadrilateral $\sqcap$ $\exists$is_completely_inside.Image_type $\sqcap$ ($\ni$ has_colour.{grey} $\sqcup$ ($\ni$ has_colour.{dark_grey}) |
| $\beta_4$ | UJI_Fire_Extinguisher | $\equiv$ | Object_type $\sqcap$ $\exists$is_completely_inside.UJI_Wall $\sqcap$($\ni$ has_colour.{red} $\sqcup$ $\ni$ has_colour.{dark_red}) |

### 4.2.2 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 $\alpha_7$ from Table 4.2), 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 4.4 should be classified as *Quadrilateral* according to axioms $\gamma_1 - \gamma_7$ and $\alpha_7$ from Table 4.2. 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.

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., 2007; 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

Table 4.4: An excerpt of the basic image facts for a shape

| | |
|---|---|
| $\gamma_1$ | Object_type : object1 |
| $\gamma_2$ | Shape_type : shape1 |
| $\gamma_3$ | has_shape(object1, shape1) |
| $\gamma_4$ | has_point(shape1, point1) |
| $\gamma_5$ | has_point(shape1, point2) |
| $\gamma_6$ | has_point(shape1, point3) |
| $\gamma_7$ | has_point(shape1, point4) |

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 nominals[7]. For example, the class Point_type is defined as a set of all points recognized within the image (see axiom $\gamma_8$ from Table 4.5 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 4.5).

- *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 $\gamma_{12}$) in order to be interpreted as four different points for the *quadrilateral* individual.

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

[7]It 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

Table 4.5: An excerpt of facts of QImageOntology for closing the world for a shape.

| | |
|---|---|
| $\gamma_8$ | $\mathsf{Point\_type} \sqsubseteq \{\mathsf{point1}, \mathsf{point2}, \mathsf{point3}, \mathsf{point4}, \mathsf{point5}, \mathsf{point6}, \mathsf{point7}\}$ |
| $\gamma_9$ | $\neg\mathsf{has\_point}(\mathsf{shape1}, \mathsf{point5})$ |
| $\gamma_{10}$ | $\neg\mathsf{has\_point}(\mathsf{shape1}, \mathsf{point6})$ |
| $\gamma_{11}$ | $\neg\mathsf{has\_point}(\mathsf{shape1}, \mathsf{point7})$ |
| $\gamma_{12}$ | $\mathsf{point1} \not\approx \mathsf{point2} \not\approx \mathsf{point3} \not\approx \mathsf{point4}$ |

QImageOntology. The number of *extra* axioms to add is reasonable for our setting 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.

## 4.3 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 4.3.1, we present how our approach has been implemented and we also describe the results obtained. In Section 4.3.2, the tests done in different situations within our robot scenario and the evaluation method used are explained. In Section 4.3.3, the results obtained are analysed.

### 4.3.1 Implementation and Testing

Figure 4.1 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 3.11) and also as a set of OWL ontology facts.

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 reasoner (e.g. FaCT++[8] or Pellet[9]) 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 4.1 (shown in Table 3.11) and the contextualized descriptions shown in Table 4.3 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.

[8]FaCT++: `http://owl.man.ac.uk/factplusplus/`
[9]Pellet: `http://clarkparsia.com/pellet/`

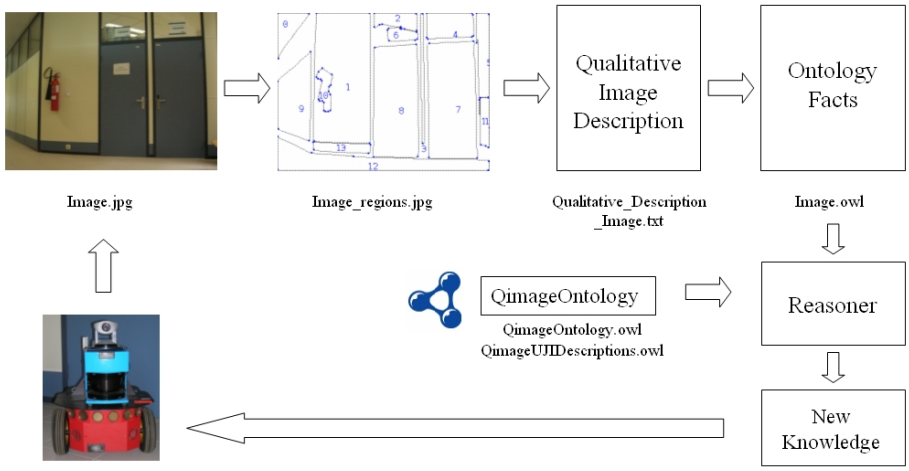



Figure 4.1: Overview of the approach for obtaining an ontology from a QID.

### 4.3.2 Evaluation of Results

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. Tables 4.6 and 4.7 present a selection of the tests. According to the kind of objects detected inside the captured images, our system can classify them as: (i) images containing *UJI Office Doors* (see Table 4.6); (ii) images containing *UJI Fire Extinguishers* (see Table 4.7); and (iii) images containing both *UJI Office Doors* and *UJI Fire Extinguishers* (see Table 4.7).

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.

- 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.

Table 4.6: Some images of the corridors of our building containing *UJI Office Doors*, *UJI Lab Walls* and *UJI Floor*.

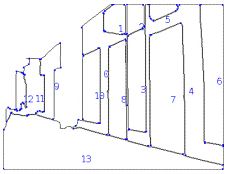
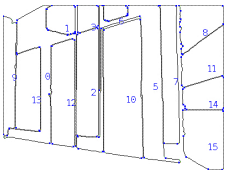
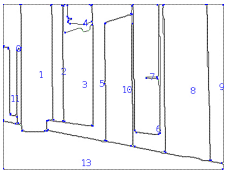
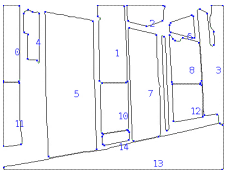
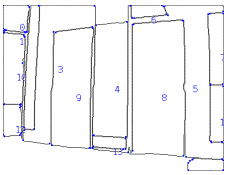
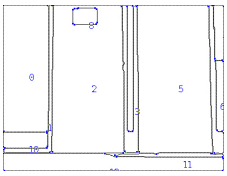
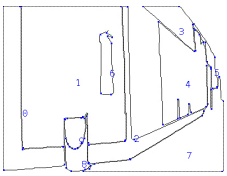
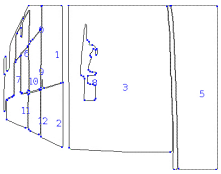
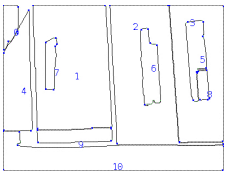
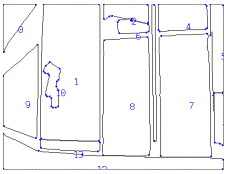
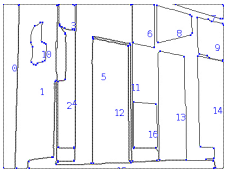
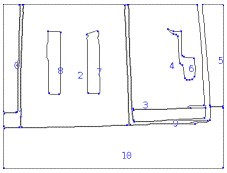
| Image | Described Objects | Inferred Information |
|---|---|---|
|  |  | Objects 7 and 8 are UJI_Office_Doors. Objects 0, 1, 2, 3, 5, 6, 9, 10, 11 and 12 are UJI_Walls. Object 13 is a UJI_Floor. |
|  |  | Objects 10 and 12 are UJI_Office_Doors. Objects 0, 2, 3, 4, 7, 8, 9, 11, 13, 14 are UJI_Walls. |
|  |  | Objects 8 and 10 are UJI_Office_Doors. Objects 0, 1, 2, 3, 4, 6, 7, 12 are UJI_Walls. Object 13 is a UJI_Floor. |
|  |  | Objects 5 and 7 are UJI_Office_Doors. Objects 0, 1, 3, 6, 8, 9, 10 are UJI_Walls. Object 13 is a UJI_Floor. |
|  |  | Objects 8 and 9 are UJI_Office_Doors. Objects 0, 1, 4, 6 are UJI_Walls. Object 13 is UJI_Floor |
|  |  | Objects 2 and 4 are UJI_Office_Doors. Objects 0, 1, 3, 6, 8, 9 are UJI_Walls. Object 10, 11, 12 are UJI_Floors. |

Table 4.7: 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 UJI_Fire_Extinguisher. Objects 1, 4, 5, 7, 8, 9 are UJI_Walls. |
|  |  | Object 8 is UJI_Fire_Extinguisher. Objects 0, 1, 2, 3, 7, 9, 10, 11 are UJI_Walls. |
|  |  | Object 7 is a UJI_Fire_Extinguisher. Object 1 is a UJI_Wall. Object 3 is a UJI_Office_Door. |
|  |  | Object 10 is a UJI_Fire_Extinguisher. Objects 0-6, 9, 11, 12 and 13 are UJI_Walls. Objects 8 and 7 are UJI_Office_Doors. Object 12 is a UJI_Floor. |
|  |  | Object 10 is a UJI_Fire_Extinguisher. Objects 1-4, 6 and 16 are UJI_Walls. Objects 12 and 13 are UJI_Office_Doors. |
|  |  | Objects 6 is a UJI_Fire_Extinguisher. Objects 0-5 and 8 are UJI_Walls. Object 2 is a UJI_Office_Door. Object 9 and 10 are UJI_Floors. |

### 4.3.3 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. laboratories/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 corresponding to similar or possibly the same visual landmarks and those landmarks could be used to localize the robot specifically in the world.

## 4.4 Conclusions

This chapter presented a novel approach to represent qualitative descriptions 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.

The QID approach presented in the previous chapter obtains the qualitative shape, colour, topology, and fixed and relative orientation of all the characteristic regions/objects within an image. 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. Although a contextualized ontology definition for every possible object detected in an image (e.g. printer, office desk or chair) is not provided, our approach can automatically process any random image and obtain a set of DL-axioms that describe it visually and spatially.

The approach presented in this chapter has been tested using digital images of the corridors of our building at University Jaume I (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 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.

# Chapter 5

# A Similarity Measure Between Qualitative Image Descriptions

When visiting a new city, walking through a new path on the mountains or driving a car in a new country, sign pictures provide knowledge to us if we can match them to a guide. If we cannot match the pictures, they are not useful to orientate ourselves. In the same way, images captured by a robot webcam can help the robot to situate itself in a map if it can match them to descriptions of places or landmarks it has visited or seen before.

The problem of image similarity has been widely studied in literature for years. The application of image retrieval in data bases has been the most used for testing, using as the query of the searches: first a complete image, but now evolving more and more to human-language descriptions. Moreover, since the image processing algorithms have been speed up to provide response in real time, the problem of identifying natural visual landmarks in robotics has also been an important application for image similarity approaches.

In literature, the most effort has been done in obtaining a measure of similarity between images described mathematically. These similarity approaches can be classified taking into account: (i) the transformation done to the original images and (ii) the distance or dissimilarity technique used for comparing the images transformed into the corresponding space. According to this, works in literature can be classified as:

- approaches that transform original images into binary images by segmentation processes and compare the output using intensity-based indices that apply Boolean operations to the corresponding pixel intensities. For an overview of those early approaches see Sampat et al. [2009].

- approaches that transform original images into colour histograms and compare them by calculating: the Euclidean distance between them [Liu et al., 2004], fuzzy correlations of them [Zhai et al., 2005], a similarity distance from the Dirichlet distribution [Missaoui et al., 2004], etc. A comparison by Rubner et al. [2001] of nine dissimilarity measures applied to colour and

texture histograms of images (e.g. Minkowski-form distance, Weighted-Mean-Variance, Kolmogorov-Smirnov distance, etc.) concluded that there is no measure with best overall performance, all of them have a high computational cost to calculate and are effective depending on the task.

- transform original images into mixtures of gaussians (MoG) and compare them using: (i) the Kullback-Liebler divergence, (ii) Monte Carlo simulations, (iii) Mahalanobis match, etc. See a comparison of methods in the work by [Goldberger et al., 2003]. The drawback of all these approaches is its high computational complexity.

- transform original images into energy spectra and interpret them by generating a multidimensional space in which scenes sharing membership in semantic categories (i.e. streets, highways, coasts) are projected closed together [Oliva and Torralba, 2001]. Quattoni and Torralba [2009] have extended this approach including learning techniques such as support vector machines (SVMs) and methods of segmentation of regions of interest (ROI) for indoor scene recognition.

- transform original images into census histograms and apply the principal component analysis operation to compare them [Wu and Rehg, 2008; Wu et al., 2009; Wu and Rehg, 2010]. The advantage of this approach is that evaluates very fast. However, it needs training using SVMs.

- transform original images into SIFT [Lowe, 2004] flow images and compare them by determining the dense correspondence between the compared images. As analysing optical flow or dense sampling of the time domain is assumed to enable tracking, dense sampling in the space of world images is assumed to enable scene alignment [Liu et al., 2011].

- transform panoramic images into signatures consisting of a constellation of descriptors computed over the different types of local affine covariant region descriptors (Harris-Affine, Hessian-Affine, MSER) and SIFT and GLOH descriptors (see the work by Mikolajczyk et al. [2005] for an overview of all these methods) and compare two panoramic images by establishing matches as nearest neighbours between the feature descriptors of both panoramas using the Euclidean distance as similarity measure and then applying complementary methods for rejecting potentially false positives [Ramisa, 2009; Ramisa et al., 2009].

All these approaches obtain a similarity index between images and use it to classify them successfully into categories. However, the problem of understanding the content of the images and interpret it, is only partially solved by these approaches, because:

(i) they do not identify objects in the processed images. For example, the approaches by Quattoni and Torralba [2009] and Wu and Rehg [2010] can classify images of streets, living rooms, bedrooms, etc. but they cannot identify if there is a car, a sofa or a bed in the image or not, and not even where in the image these objects are located.

(ii) they can only identify in the images a set of previously determined objects. For example, all the approaches based on affine covariant region descriptors

or SIFT descriptors, as for example the works by Liu et al. [2011] and Ramisa [2009], need a database of images of objects or scenes to compare with. And, an 'unknown' object or an object that does not appear in this database, it cannot be detected.

However, when we try to enhance the perception of an intelligent agent, the content of the image and how it is arranged is very important. For example, a robotic system can use a generic scene similarity approach [Oliva and Torralba, 2001; Quattoni and Torralba, 2009; Ramisa, 2009; Liu et al., 2011] in order to determine in which room of a house is located. However, if it needs to manipulate the environment where it navigates, it needs to identify objects and where are they located in order to guide itself to them. Another example is given by the image retrieval systems that try to interpret a query given in human language. Human descriptions of images usually are based on object identification and location rather than using only a word to describe the whole image.

Moreover, as Vernon [2008] points out: *'Cognition implies an ability to understand how the world around us might possibly be (...) and being able to interpret a visual scene without having complete data'*. Therefore, a cognitive system should be able to describe and identify an image without having the complete information about it, that is, it should be able to describe objects that have not seen before (or 'unknown' objects) by describing its shape and colour using high-level or qualitative features. Therefore, for increasing the robot cognition and knowledge about the world, there are works that have changed the low-level features used in the previously reviewed similarity calculus approaches for *high-level* or qualitative features that provide a more abstract representation of the image.

As approaches that describe images using *high-level* or qualitative features, it is worth mention the works by: Stefanidis et al. [2002], Deruyver and Hod [2009], Sousa and Fonseca [2009] and our approach. Stefanidis et al. [2002] define a similarity measure between images of airports (in which the user defines the scene and the object configurations) based on the shape, topology, orientation and distance of the objects in the images. Deruyver and Hod [2009] interpret segmentation results of images by establishing cardinal direction relations and topological relations between object subparts. Conceptual graphs are used to describe the general spatial distribution of the subparts of faces, cars, flowers and anatomical cerebral images. Using a constraint satisfaction algorithm, the conceptual graphs are used to retrieve images containing similar objects. Sousa and Fonseca [2009] present an approach to index and retrieve complex vector drawings by content, using topological and geometric information automatically extracted from figures where a graph-based technique is used to describe the spatial arrangement of drawing components, coding the topological relationships of inclusion and adjacency through the specification of links between nodes of a graph.

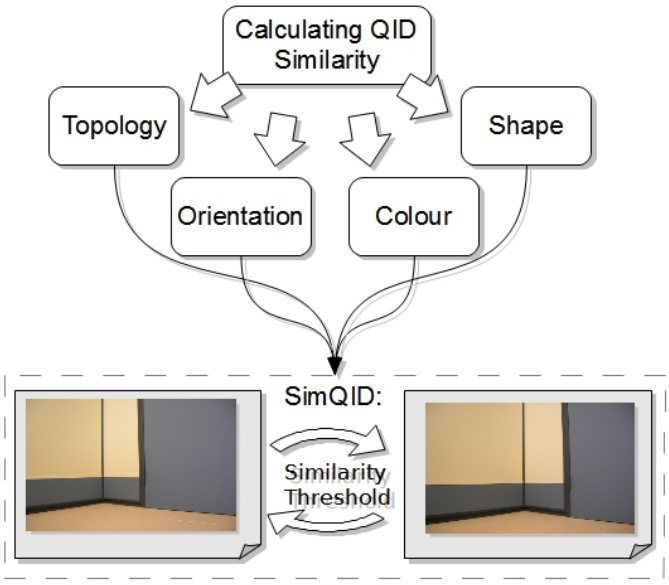
Finally, our Computational Model for Qualitative Image Description (see Chapter 3) describes the qualitative shape, colour, topology and orientation of all the objects (known or unknown) within any digital image. Therefore, as Figure 5.1 shows, in order to find a similarity degree between two image descriptions, it is necessary to define a similarity measure between:

1. Qualitative Shape Descriptions ($QSD$) (Section 3.3),

2. Qualitative Colour Descriptions ($QCD$) (Section 3.4),
3. Topology Descriptions (Section 3.5),
4. Orientation Descriptions(Section 3.6),

And then it is necessary to determine in which degree these properties (shape, colour, topology and orientation) are taken into account for obtaining similarities between objects within an image and between images themselves.

Figure 5.1: Outlining our approach for calculating the similarity between Qualitative Image Descriptions (QIDs).




This Chapter organizes all these work parts as follows. Section 5.1 studies shape similarity and defines two similarity measures for qualitative shape descriptions based on conceptual neighbourhood diagrams and interval distances ($SimQSD$). Section 5.2 studies colour similarity and determines a similarity measure for comparing qualitative colour descriptions ($SimQCD$). Section 5.3 studies the literature on spatial similarity and defines a similarity measure between topology descriptions ($SimTop$) and a similarity measure between fixed orientation descriptions with respect to the centre of the image based on a conceptual neighbourhood diagram ($SimFO$). Finally, Section 5.4 defines how to obtain a similarity measure between two Qualitative Image Descriptions ($QID$) using all the similarities defined before ($SimQID$) and describes the applications of $SimQID$ in three robotic scenarios. Finally, conclusions and future work are presented in Section 5.5.

## 5.1 Shape Similarity

From a cognitive point of view, the problem of shape equivalence involves understanding the conditions under which people perceive two distinct objects as having the same shape: two objects will continue to have the same shape even after they have undergone spatial transformations such as translations (changing position), rotations (changing orientation), dilations (expanding and contracting in size), mirror-image reflections (changing direction) and combinations of these transformations. However, if spatial transformations such as squashing, stretching or deforming in any way are needed to bring two objects into exact correspondence, then they are not considered equivalent, although they can be perceived as very similar [Palmer, 1989]. Hence, an important problem is to obtain a similarity measure between shapes which quantifies the *resemblance* or *closeness* between them.

In order to tackle the problem of shape similarity, Section 5.1.1 reviews related work and Section 5.1.2 presents two similarity measures between Qualitative Shape Descriptions (QSDs) based on conceptual neighbourhood diagrams and on interval distances, and finally Section 5.1.3 presents tests of both measures on objects from the MPEG7 CE Shape-1 library and discusses the results obtained.

### 5.1.1 Related Work on Shape Similarity

Shape similarity has been widely studied. In the literature works can be found that define similarity measures between shapes that are represented by: (i) quantitative information [Super, 2004; Ling and Jacobs, 2007; Attalla and Siy, 2005; Latecki and Lakamper, 2000; Bai et al., 2008; Gdalyahu and Weinshall, 1998; Mori et al., 2001]; (ii) mixed quantitative and qualitative information [Shokoufandeh et al., 2002; Berretti et al., 2000; Siddiqi et al., 1998; Macrini et al., 2008; Sebastian et al., 2001, 2002] and (iii) qualitative information [Gottfried, 2008; Kuijpers et al., 2006; Schuldt et al., 2006].

Approaches to shape similarity calculus based on quantitative representations can be classified into:

- approaches that match points of the shape boundary: Super [2004] defines critical points of high curvature on boundaries and normalises the shape to a reference frame for rotation and scaling before calculating a distance measure used in the matching process; whereas Ling and Jacobs [2007] consider the inner-distance, or the length between landmark points within the shape silhouette to define shape descriptors invariant to articulation, which improved the classification of articulated shapes of 2D objects.

- approaches that match segments of the shape boundary: (i) shapes are segmented at multiple resolutions and a similarity is defined by elastic matching of shape segments in the work by Attalla and Siy [2005]; (ii) a similarity measure between shapes based on the correspondence of visual parts where partial matching can be performed when the scale is known is presented by Latecki and Lakamper [2000] and then it is used for detection and recognition of contour parts in digital images by Bai et al. [2008]; and (iii) a local curve matching algorithm is described by Gdalyahu and

Weinshall [1998] that extracts points of high curvature and calculates a distance between them using efficient alignment.

- approaches that match the context of the shapes: Mori et al. [2001] define a shape feature descriptor vector that is used to represent general shape contour.

Approaches to shape similarity that mix quantitative and qualitative representations are those based on graphs/trees that usually describe the spatial arrangement of the shape parts between them but also contain some measurable properties of each shape part in their edges/nodes. For example: (i) Shokoufandeh et al. [2002] divide the coarse shape of an object into blobs and geometric relationships between them are organised into a graph, which is used for shape comparing; (ii) a shape is divided into tokens, according to its protrusions, and arranged into an M-tree, which is used to calculate distances between tokens and to obtain a dissimilarity measure between the M-trees of two shapes by Berretti et al. [2000]; (iii) *shocks* (singularities) of a curve on bounding contours are organised into a graph for shape matching by Siddiqi et al. [1998] and evolve to skeletons[1] and bone graphs for object recognition in the work by Macrini et al. [2008]; finally, (iv) a distance between shock graphs is defined and used for recognition of shapes in the work by Sebastian et al. [2001] and for retrieval of similar shapes in large databases in the work by Sebastian et al. [2002].

In terms of qualitative approaches to shape similarity, the most representative can be generally classified as:

- based on qualitative shape descriptors: (i) *bipartite arrangements* defined by Gottfried [2008] that relate line segments of a contour of an object to other parts of that same contour and then a similarity measure between these qualitative descriptions of shape; (ii) matrices of qualitative concepts developed by Kuijpers et al. [2006] using the double-cross orientation model by Freksa [1992] to describe polylines and to find a similarity measure between polygons; and finally, (iii) polygons described qualitatively by their scope (calculated as their relative position with respect to one of their line segments where the double-cross grid described by Freksa [1992] is located) and scope histograms generated and used for shape comparing by Schuldt et al. [2006].

- theoretical approaches: the recognition-by-components theory by Biederman [1987] in which any object can be generated from a set of generalized-cone components, called geons; the relational modelling technique by Shapiro et al. [1980] which decomposes objects into sticks, plates and blobs; and finally, the codons by Richards and Hoffman [1985] that are simple primitives for describing closed 2D shapes.

To the best of our knowledge, there is no approach which describes shapes qualitatively using the QSD model (detailed in Section 3.3.2) and obtains a similarity measure between qualitative shape descriptions using conceptual neighbourhood diagrams and interval distances.

[1] A skeleton or axis is a two dimensional arc reflecting some global or local symmetry or regularity within a shape.

### 5.1.2 Similarity Between Qualitative Shape Descriptions ($SimQSD$)

An approach to shape comparison based on the Qualitative Shape Description (QSD) presented in Section 3.3.2 is presented here. Remember that, in our QSD, the relevant points of the shape of a 2D object are described by a set of four elements:

$$<\mathrm{KEC}_p,\ \mathrm{A}_p \mid \mathrm{TC}_p,\ \mathrm{L}_p,\ \mathrm{C}_p>$$

where,

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

$\mathrm{A}_p \in$ {*very_acute, acute, right, obtuse, very_obtuse* / j is a *line_line, line_curve, curve_line, curve_curve*};

$\mathrm{TC}_p \in$ {*very_acute, acute, semicircular, plane, very_plane* / j is a *curvature_point*};

$\mathrm{L}_p \in$ {*much_shorter (msh), half_length (hl), a_bit_shorter (absh), similar_length (sl), a_bit_longer (abl), double_length (dl), much_longer (ml)*}

$\mathrm{C}_p \in$ {*convex, concave*};

Therefore, the complete description of the shape of a 2D object is given from a set of qualitative tags as follows[2]:

$$[\mathrm{KEC}_0,\ \mathrm{A}_0 \mid \mathrm{TC}_0,\ \mathrm{L}_0,\ \mathrm{C}_0],\ [\mathrm{KEC}_1,\ \mathrm{A}_1 \mid \mathrm{TC}_1,\ \mathrm{L}_1,\ \mathrm{C}_1],\ldots,\ [\mathrm{KEC}_{n-1},\ \mathrm{A}_{n-1} \mid \mathrm{TC}_{n-1},\ \mathrm{L}_{n-1},\ \mathrm{C}_{n-1}]$$

where n is the total number of relevant points of the object, 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 other relevant points are described cyclically in a clockwise direction.

From a cognitive point of view, shape is defined by Wilson and Keil [1999] (see Shape Perception entry) as: *'An aspect of a stimulus that remains invariant despite changes in size, position and orientation.'* Therefore, it is important to note that the QSD presented here is:

- invariant to scaling (expansions and contractions in size). If a shape is scaled, then all the edges are expanded or contracted in the same proportion, and therefore the features of shape obtained in both situations are the same. If a shape is expanded or contracted until an edge disappears, then our approach considers that the original shape is transformed into another different one because they have different quantity of edges.

- invariant to translations (changes in position), because an object description is always started at the point closest to the upper-left corner of the image, and therefore does not depend on where the object is located in the image.

Clearly, the QSD is not invariant to rotations (changing orientation), however the comparison of two shapes would be invariant to rotation if both shape descriptions were compared considering each point as the possible starting point

[2] $\mathrm{A}_i \mid \mathrm{TC}_i$ denotes that the angle or the type of curvature that occurs at the point $P_i$.

of the (cyclic) description. This is one constraint that must be considered when defining a cognitive similarity measure between shapes described by our QSD.

With the aim of defining a similarity measure between two QSDs corresponding to two objects, it is necessary to work through three stages:

- defining a similarity measure between the qualitative tags (related to the shape features Edge Connection (EC), Angle (A), Type of Curvature (TC), compared Length (L) and Convexity (C)) that describe the relevant points of each QSD of the objects compared (Section 5.1.2.1);
- obtaining a similarity measure between a pair of relevant points: each one corresponding to the QSD of each compared object (Section 5.1.2.2);
- defining a similarity measure between the QSDs of both objects by establishing a correspondence of pairs of relevant points (Section 5.1.2.3).

#### 5.1.2.1 Similarity of Qualitative Features of Our QSD

In cognitive psychology, human assessments of similarity refer to the closeness of the mental representations made about both of the compared concepts.

In order to compare qualitative features, a model in which dissimilarity is calculated as the number of transformations (or distance) between qualitative concepts is used. Furthermore, as some of these qualitative concepts are defined using a reference system based on interval values (A, TC and L, then a dissimilarity between them can be obtained by interval distances, from a mathematical point of view.

Therefore, the approach presented here provides a similarity measure between qualitative shape descriptions (QSDs) based on the number of transformations required to sequentially pass from one qualitative concept (tag, label) to another. Hence, the fewer transformations, the more similar two shapes are. A cost for transformations is assigned from dissimilarity matrices based on:

- conceptual neighbourhood diagrams (Section 5.1.2.1.1); and
- interval distances (Section 5.1.2.1.2).

##### 5.1.2.1.1 Building Dissimilarity Matrices Using Conceptual Neighbourhood Diagrams (CNDs)

The possible transformations between two labels that describe a feature can be defined from its corresponding CND. The term *conceptual neighbourhood* was first considered by Freksa [1991] in his analysis of the 13 interval relations defined in the temporal logic by Allen [1981]: *'Two relations between pairs of events are conceptual neighbours if they can be directly transformed one into another by continuous deformation (i.e., shortening or lengthening) of the events'.*

Conceptual neighbourhood relations can be found between the qualitative tags defined for each feature of shape in the QSD model. For example, when dealing with angles, the angles *acute* and *right* can be considered conceptual neighbours since a quantitative extension of the angle *acute* leads to a direct transition to the angle *right*. However, angles *acute* and *obtuse* are not conceptual neighbours, since a transition between them must go through the angle *right* first.

In general, *CNDs* can be described as diagrams or graphs containing: (i) nodes that map to a set of individual relations defined on regions or intervals and (ii) paths or edges connecting pairs of adjacent nodes that map to continuous transformations between them.

The CND proposed for each feature of shape can be seen in Figure 5.2, 5.3, 5.4, 5.5, and 5.6. Hence, the dissimilarity between two tags in each CND is calculated in Table 5.1, 5.2, 5.4, 5.3 based on the minimal path between them. Note that, although the features Type of Curvature (TC) and Angle (A) contain different qualitative concepts, their CNDs and dissimilarity matrices have been defined in exactly the same way because the units and intervals are the same.

It is important to note that these dissimilarities are coherent with human common sense because, for example, the dissimilarity between a point that connects two straight lines (*line-line*) and a point that connects a straight line and a curve (*line-curve*) is 1; and that the dissimilarity between a *line-line* point and a *curvature-point* is 2. Therefore, a *line-line* point is more similar to a *line-curve* point than to a *curvature-point*, which is a logical conclusion.

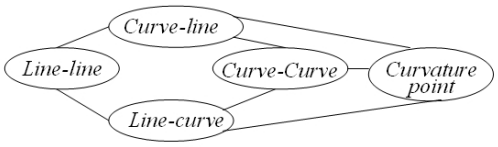



Figure 5.2: CND for feature Kind of Edges Connected.


Figure 5.3: CND for feature Angle.


Figure 5.4: CND for feature Type of Curvature.


Figure 5.5: CND for feature Length.

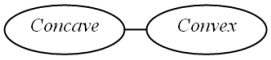



Figure 5.6: CND for feature Convexity.

Table 5.1: Dissimilarity matrix for KEC using a CND.

| ***KEC*** | *line line* | *line curve* | *curve line* | *curve curve* | *curvature point* |
|---|---|---|---|---|---|
| *line line* | 0 | 1 | 1 | 2 | 2 |
| *line curve* | 1 | 0 | 2 | 1 | 1 |
| *curve line* | 1 | 2 | 0 | 1 | 1 |
| *curve curve* | 2 | 1 | 1 | 0 | 1 |
| *curvature point* | 2 | 1 | 1 | 1 | 0 |

Table 5.2: Dissimilarity matrix for TC and A using a CND.

| ***TC or A*** | *Very acute* | *Acute* | *Semicircular or Right* | *Plane or Obtuse* | *Very plane or Very obtuse* |
|---|---|---|---|---|---|
| *Very acute* | 0 | 1 | 2 | 3 | 4 |
| *Acute* | 1 | 0 | 1 | 2 | 3 |
| *Semicircular or Right* | 2 | 1 | 0 | 1 | 2 |
| *Plane or Obtuse* | 3 | 2 | 1 | 0 | 1 |
| *Very plane or Very obtuse* | 4 | 3 | 2 | 1 | 0 |

Table 5.3: Dissimilarity matrix for C using a CND.

| ***Convexity*** | *Concave* | *Convex* |
|---|---|---|
| *Concave* | 0 | 1 |
| *Convex* | 1 | 0 |

Table 5.4: Dissimilarity matrix for L using a CND.

| ***Length*** | msh | hl | qsh | sl | ql | dl | ml |
|---|---|---|---|---|---|---|---|
| *msh* | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| *hl* | 1 | 0 | 1 | 2 | 3 | 4 | 5 |
| *qsh* | 2 | 1 | 0 | 1 | 2 | 3 | 4 |
| *sl* | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
| *ql* | 4 | 3 | 2 | 1 | 0 | 1 | 2 |
| *dl* | 5 | 4 | 3 | 2 | 1 | 0 | 1 |
| *ml* | 6 | 5 | 4 | 3 | 2 | 1 | 0 |

#### 5.1.2.1.2 Building Dissimilarity Matrices Using interval Distances

To define the dissimilarity matrices for the features Angle, Type of Curvature and compared Length, an ordinal scale has been used in Section 5.1.2.1.1. However, these features are defined from intervals of values in their Reference Systems. Therefore, we can take advantage of this by not considering dissimilarity matrices but instead, distance matrices, which are richer mathematically, since the concept of distance is stricter than the concept of dissimilarity.

Now we will briefly introduce the concept of interval distance. Given an open interval (analogously for another kind of interval) of finite dimension, there are two main ways to represent it: from the extreme points as *(a,b)* (classical notation) or as an open ball $B_r(c)$ (Borelian notation) where $c = (a + b)/2$ (centre) and $r = (b - a)/2$ (radius). Thus, for example, the interval *(10,20)* can be written as $B_5(15)$ in Borelian notation.

Given two intervals, $I_1 = (a_1, b_1) = B_{r1}(c_1)$ and $I_2 = (a_2, b_2) = B_{r2}(c_2)$, a family of distances between intervals was defined in Gonzalez-Abril et al. [2009], which depends on three parameters as follows:

$$d^2(I_1, I_2) = (\ \Delta c\ \Delta r\ ) A \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix}$$

where $\Delta c = c_2 - c_1$, $\Delta r = r_2 - r_1$ and $A$ is a symmetrical 2×2 matrix of weights, which must be a positive definite matrix. From the $A$ matrix, the weights given to the position of the intervals and to the radius can be controlled.

Here, the most natural choice for the $A$ matrix is used, which is the identity matrix that provides the next distance (written using the two notations explained above):

$$d(I_1, I_2) = \sqrt{\Delta^2 c + \Delta^2 r} = \sqrt{(c_2 - c_1)^2 + (r_2 - r_1)^2} \tag{5.1}$$

Hence, new dissimilarity matrices have been considered for the features: Angle, Type of Curvature and Length. Let us consider these one by one.

For the feature Angle, the following intervals in Borelian notation are:

$A_{Br(c)} = \{B_{20}(20), B_{22.5}(62.5), B_5(90), B_{22.5}(117.5), B_{20}(160)\}$

and their corresponding distance (dissimilarity) matrix is given in Table 5.5.

Table 5.5: Distance matrix for TC and A using interval distances.

| ***TC or A*** | *Very acute* | *Acute* | *Semicircular or Right* | *Plane or Obtuse* | *Very plane or Very obtuse* |
|---|---|---|---|---|---|
| *Very acute* | 0.0 | 42.6 | 71.6 | 97.5 | 140.0 |
| *Acute* | 42.6 | 0.0 | 32.6 | 55.0 | 97.5 |
| *Semicircular or Right* | 71.6 | 32.6 | 0.0 | 32.6 | 71.6 |
| *Plane or Obtuse* | 97.5 | 55.0 | 32.6 | 0.0 | 42.6 |
| *Very plane or Very obtuse* | 140.0 | 97.5 | 71.6 | 42.6 | 0.0 |

Note that, considering the dissimilarity matrix built from the CND for the feature Angle (Table 5.2), the dissimilarity between the interval [0,40] (corresponding to *very-acute*) and the interval (40,85] (corresponding to *acute*) is 1; and the dissimilarity between the interval [0,40] and the interval (85, 95] (corresponding to *right*) is 2. Mathematically speaking, it is not accurate to say that the second dissimilarity is double the first dissimilarity since proportional values cannot be calculated on an ordinal scale. Nevertheless, considering the dissimilarity matrix built from interval distances for the feature Angle (Table 5.5), the distance between the interval [0,40] and the interval (40,85] is 42.6; and the distance between the interval [0,40] and the interval (85,95] is 71.6. Hence, by using the proportional scale, it is absolutely accurate to compare both distances from their ratio (71.6/42.6), obtaining that one is 1.68 times the other.

For the feature Type of Curvature, the intervals in Borelian notation are:

$\mathrm{TC}_{Br(c)}$= {$\mathrm{B}_{20}(20)$, $\mathrm{B}_{22.5}(62.5)$, $\mathrm{B}_{5}(90)$, $\mathrm{B}_{22.5}(117.5)$, $\mathrm{B}_{20}(160)$}

Note that the distance matrix must be obtained from $\mathrm{TC}_{INT2}$ instead of $\mathrm{TC}_{INT1}$ because if $\mathrm{TC}_{INT1}$ is used, then the distance between the interval [2.75, $\infty$) and the other interval is infinite. The distance matrix obtained for the Type of Curvature is the same as the one obtained for the Angle, which is given in Table 5.5.

For the feature Length, the following intervals in Borelian notation are:

$\mathrm{L}_{Br(c)}$ = {$\mathrm{B}_{0.2}(0.2)$, $\mathrm{B}_{0.1}(0.5)$, $\mathrm{B}_{0.15}(0.75)$, $\mathrm{B}_{0.1}(1.0)$, $\mathrm{B}_{0.4}(1.5)$,$\mathrm{B}_{0.1}(2.0)$, $\mathrm{B}_{3.95}(6.05)$}

and their corresponding distance (dissimilarity) matrix is given in Table 5.6.

The last interval in $\mathrm{L}_{INT}$ is $(2.1, \infty)$ which is infinite, so a new interval $(2.1, 10)$ has been considered in order to obtain finite distances between the defined features of compared distance. A maximum threshold for compared distances has been established at 10, which means that our system considers as a maximum that one edge of a shape is 10 times longer than another edge.

Table 5.6: Distance matrix for L using interval distances.

| ***Length*** | msh | hl | qsh | sl | ql | dl | ml |
|---|---|---|---|---|---|---|---|
| *msh* | 0.00 | 0.32 | 0.55 | 0.81 | 1.32 | 1.80 | 6.95 |
| *hl* | 0.32 | 0.00 | 0.25 | 0.50 | 1.04 | 1.50 | 6.75 |
| *qsh* | 0.55 | 0.25 | 0.00 | 0.25 | 0.79 | 1.25 | 6.52 |
| *sl* | 0.81 | 0.50 | 0.25 | 0.00 | 0.58 | 1.00 | 6.35 |
| *ql* | 1.32 | 1.04 | 0.79 | 0.58 | 0.00 | 0.58 | 5.77 |
| *dl* | 1.80 | 1.50 | 1.25 | 1.00 | 0.58 | 0.00 | 5.59 |
| ***ml*** | 6.95 | 6.75 | 6.52 | 6.35 | 5.77 | 5.59 | 0.00 |

#### 5.1.2.2 Calculating a Similarity Degree Between Relevant Points

As mentioned above, the qualitative shape of an object is described by means of all its relevant points. Therefore, in order to define a similarity measure between shapes, first a similarity between relevant points must be obtained. Hence, given two relevant points, denoted by $RP_A$ and $RP_B$, belonging to the shapes of the objects A and B respectively, a similarity between them, denoted by $Sim(RP_A, RP_B)$, is defined as:

$$Sim(RP_A, RP_B) = 1 - \sum_{i \in \{KEC, A \vee TC, C, L\}} w_i \, \frac{ds(i)}{Ds(i)}, \tag{5.2}$$

where $ds(feature)$ denotes the dissimilarity between relevant points with respect to the *feature* obtained from the dissimilarity matrix previously defined. $Ds(feature)$ denotes the maximum dissimilarity in the dissimilarity matrix related to the *feature* considered at the relevant point. Hence, by dividing $ds(feature)$ and $Ds(feature)$ the proportion of dissimilarity related to a *feature* of $\mathrm{RP}_A$ and $\mathrm{RP}_B$ is obtained, which is between 0 and 1. Moreover, the parameter $w_{feature}$ is the weight assigned to this *feature*, and $w_{KEC}+w_A+w_L+w_C = 1$, $w_A = w_{TC}$ and $w_{feature} \geq 0$ holds for each $feature$.

In this study, with the aim of giving the same importance to all features in (5.2), all the weights have the same value: $\frac{1}{4}$. Clearly, these weights can be tuned if a researcher needs to give more importance to one feature over the others. Furthermore, in (5.2) the weight is subtracted from 1 with the aim of giving a similarity instead of a dissimilarity.

For each $RP_A$ and $RP_B$, it is straightforward to prove that $0 \leq Sim(RP_A, RP_B) \leq 1$ and that this is a symmetrical relation. Furthermore, this similarity is very intuitive since $Sim(RP_A, RP_B) = 0$ means that $ds(feature) = Ds(feature)$, that is, both relevant points have the maximum dissimilarity for all the features and hence, both relevant points are as different as possible.

On the other hand, if $Sim(RP_A, RP_B) = 1$, then this means that $ds(feature) = 0$ for all the features of the relevant points, and hence, these two relevant points have the same QSD. In this case, both relevant points are considered equivalent (a relation of equivalence is established between them).

If one relevant point is a *curvature-point* (feature KEC) and the other compared relevant point is not, the type of curvature (feature TC) of the first

relevant point will be compared to the angle (feature A) of the second relevant point. For instance, in the qualitative description of shape given in Table 3.1, if the relevant point A of the shape is compared to the relevant point B2, the corresponding Angle at A (*right*) will be compared with the corresponding Type of Curvature at B2 (*acute*). However, this is not a problem for our approach because it can compare angles with types of curvature because both features correspond to the same concept, that is, the angular amplitude at the relevant point, and both can be defined by the same values in degrees.

Furthermore, the value $\frac{ds(i)}{Ds(i)}$ in (5.2) can be seen as the importance of changes in each feature of shape. Hence, from the dissimilarity matrices obtained from CNDs, the following maximums ($Ds(i)$) are obtained: for Convexity, 1; for Kind of Edges Connected, 2; for Angle and Type of Curvature, 4; and for Length, 6. As the value assigned to each change is 1, this means that each change in each feature has a different importance in (5.2) and the following priorities among features are given:

$$I(C) = 1 > I(KEC) = \frac{1}{2} > I(A) = I(TC) = \frac{1}{4} > I(L) = \frac{1}{6}.$$

For the interval distance matrices, the maximums ($Ds(i)$) for each feature are: for Angle and Type of Curvature, 140; and for Length, 6.95. The mean value of change between the qualitative concepts of the distance matrix for Angle and Type of Curvature is approximately 35 for each line and row (as the dissimilarity matrix is symmetric), which gives us an importance of change of 35/140 or 1/4. Moreover, the value of change between the tags of the distance matrix for Length is approximately 1.18 for each line and row, which gives us an importance of change in the feature of 1.18/6.95 or 17/100. Hence, it is obtained that

$$I(C) = 1 > I(KEC) = \frac{1}{2} > I(A) = I(TC) = \frac{1}{4} > I(L) = \frac{17}{100}.$$

Therefore, the priorities given when considering dissimilarities from matrices built from CNDs or from interval distances have the same order and approximately equal values of importance.

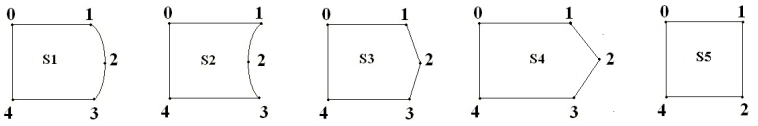



Figure 5.7: Examples of shapes for explaining the intuitive priorities obtained for C, KEC, A, TC and L.

These priorities can be justified as being suitable for comparing shapes intuitively. In Figure 5.7 five shapes are shown ($S1$, $S2$, $S3$, $S4$ and $S5$) that exemplify these priorities. The Convexity (C) is the feature that has the greatest priority because, when it changes, not only the boundary of the object changes, but also its interior (i.e. compare shapes $S1$ to $S2$ in which only the convexity of relevant point 2 changes). The Kind of Edges Connected (KEC) is

the second most important feature because it differentiates between curves and straight lines, which is also an important difference. For example, if we compare shapes $S1$ to $S3$ in which only the KEC of relevant point 2 changes, we will see that they are more similar between them than $S1$ and $S2$ and than $S2$ and $S3$ in which both the KEC and the C of 2 changes. The next most important feature is the Angle or Type of Curvature because it characterises the shape of an object in a more significant way than the lengths of the edges, which usually depend on the angle they define. If we compare $S3$ and $S4$, the most perceptual difference is that the Angle of 2 changes, but also the compared length between relevant points 3-4 and 4-0 changes in both shapes and it is less perceptible. Finally, note that it is also true that the most similar the number of relevant points between shapes, the highest the similarity, since $S1$-$S4$ are more similar between them than any of them to $S5$, which has a relevant point less than the others.

#### 5.1.2.3 Similarity Between QSDs

In order to compare two shapes A and B whose QSDs have the same number of relevant points (denoted by $n$), the similarity between A and B is calculated from (5.2) as an arithmetic mean of the similarity between relevant points of both shapes cyclically in a clockwise direction. Thus, the calculation of the similarity can start each time at a different relevant point of any of the shapes. When all the possible similarities between relevant points are obtained, the similarity between A and B is the highest value.

Let us clarify this similarity calculus with an example. Let T1 and T2 be two triangles, with QSDs given by $\{RP_{T1}(0), RP_{T1}(1), RP_{T1}(2)\}$ and $\{RP_{T2}(0), RP_{T2}(1), RP_{T2}(2)\}$ respectively. In this case, three similarities can be considered (to simplify, we denote $Sim(RP_{T1}(i), RP_{T2}(j))$ as $Sim(i, j)$):

$$\begin{aligned} Sim_1(T1, T2) &= \tfrac{1}{3}\left(Sim(0,0) + Sim(1,1) + Sim(2,2)\right) \\ Sim_2(T1, T2) &= \tfrac{1}{3}\left(Sim(1,0) + Sim(2,1) + Sim(0,2)\right) \\ Sim_3(T1, T2) &= \tfrac{1}{3}\left(Sim(2,0) + Sim(0,1) + Sim(1,2)\right) \end{aligned}$$

and the final similarity between both triangles will be the maximum of these three.

It is important to note that this final similarity provides us with a correspondence between relevant points of two shapes that will be useful later on. Thus, for instance, if the final similarity between the triangle T1 and T2 is given from the $\text{Sim}_2$ (T1, T2), then the correspondence obtained is:

$$RP_{T1}(1) \rightarrow RP_{T2}(0),\ RP_{T1}(2) \rightarrow RP_{T2}(1),\ RP_{T1}(0) \rightarrow RP_{T2}(2)$$

On the other hand, if two shapes A and B whose QSDs have a different number of relevant points are compared, then there are some relevant points of one shape with no corresponding points in the other shape. In this case, the points with no corresponding pairs in the other shape are compared with a new relevant point, the *void* point, and the similarity between both points is zero.

Let us suppose that the number of relevant points of the shapes A and B are $n$ and $m$ respectively, and without loss of generality that $n \geq m$. In this case, $n$-$m$ relevant points of A are compared with the *void* point, and the rest are compared with the relevant points of B in the same way as in the previous case.

Figure 5.8 shows two objects, A and B, with 4 and 5 relevant points respectively. When comparing A and B all the possible correspondences between the relevant points of these two objects are as follows:

{(0,void),(1,0),(2,1),(3,2),(4,3)} , {(0,void),(1,1),(2,2),(3,3),(4,0)}
{(0,void),(1,2),(2,3),(3,0),(4,1)} , {(0,void),(1,3),(2,0),(3,1),(4,2)}
{(0,0),(1,void),(2,1),(3,2),(4,3)} , {(0,1),(1,void),(2,2),(3,3),(4,0)}
{(0,2),(1,void),(2,3),(3,0),(4,1)} , {(0,3),(1,void),(2,0),(3,1),(4,2)}
**{(0,0),(1,1),(2,void),(3,2),(4,3)}** , {(0,1),(1,2),(2,void),(3,3),(4,0)}
{(0,2),(1,3),(2,void),(3,0),(4,1)} , {(0,3),(1,0),(2,void),(3,1),(4,2)}
{(0,0),(1,1),(2,2),(3,void),(4,3)} , {(0,1),(1,2),(2,3),(3,void),(4,0)}
{(0,2),(1,3),(2,0),(3,void),(4,1)} , {(0,3),(1,0),(2,1),(3,void),(4,2)}
{(0,0),(1,1),(2,2),(3,3),(4,void)} , {(0,1),(1,2),(2,3),(3,0),(4,void)}
{(0,2),(1,3),(2,0),(3,1),(4,void)} , {(0,3),(1,0),(2,1),(3,2),(4,void)}

For the objects in Figure 5.8, $Sim(A, B)$ is given from the correspondence {*(0,0),(1,1),(2,void),(3,2),(4,3)*}. Therefore, our approach provides additional information about the shape: the RP 2 in the object B has no correspondent RP in the object A.

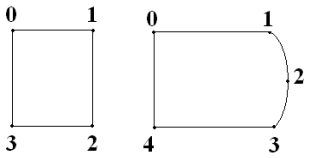



Figure 5.8: Examples of polygons with a different number of relevant points.

Thus, the similarity for each one of all possible correspondences between the relevant points of A and B by considering the *void* point is obtained as:

$$Sim_{\sigma}(A, B) = \frac{1}{n} \sum_{i=1}^{m} Sim(RP_A\sigma(i), RP_B(i)) \tag{5.3}$$

where $\sigma$ denotes a cyclic correspondence of the relevant point of object A and the relevant point of object B.

Note that only $m$ similarities between relevant points must be considered because the similarity between a relevant point of A and the *void* point is always zero. From here, the final similarity between the shapes A and B, called $SimQSD(A, B)$, is the maximum value of these similarities, that is,

$$SimQSD(A, B) = max_{\sigma \in C}(Sim(A, B)) \tag{5.4}$$

where $C$ denotes the set of all possible correspondences between relevant points of A and B.

The main properties of this final similarity are:

- Symmetry: $SimQSD(A, B) = SimQSD(B, A)$.
- Invariance to rotation, translation and scale transformations;

- Upper and lower bounds, that is,

$$0 \leq SimQSD(A, B) \leq 1$$

  for any shapes A and B. Moreover,

$$0 \leq SimQSD(A, B) \leq 1 - \frac{n-m}{n} = \frac{m}{n}$$

  because the difference between the number of relevant points of shapes penalizes the final similarity. This result is desirable because the higher the difference in the number of relevant points, the lower the similarity that should be obtained.

At this point, note that, according to Wilson and Keil [1999], our approach describes and compares shapes in a cognitive way as it takes into account most of the conditions under which people perceive two distinct objects as having the same shape, that is, invariance to changes in size, position and orientation.

In terms of the computational cost (CC), two situations (the best and worst case) can be distinguished:

- if both shapes have the same number of relevant points ($n$), the cost of the algorithm is $O(n^2)$, because the starting point of the comparison can be any point of the second shape.

- if the difference in the number of relevant points between both shapes is $n$-$m$, the number of possibilities for choosing $n$-$m$ points to be compared with the *void* point is a simple combinatory number $\begin{pmatrix} n \\ n-m \end{pmatrix} = \frac{n!}{(n-m)!m!}$ and considering that the starting point of the comparison can be any point of the shape with the highest number of relevant points, the possible costs are:

$$\begin{pmatrix} n \\ n-m \end{pmatrix} m = \frac{n!}{(n-m)!m!} m = \frac{n(n-1)...(m+1)}{(n-m)!} m = O(n^{n-m+1})$$
$$\begin{pmatrix} n \\ n-m \end{pmatrix} m = \frac{n(n-1)...(n-m+1)}{m!} m = O(n^{m+1})$$

and the final cost is the minimum of the above costs:

$$C = min\left(O(n^{m+1}), O(n^{n-m+1})\right) \leq O(n^{r+1})$$

where

$$r = \left\{ \begin{array}{cc} n/2 & if\ n\ is\ odd \\ (n+1)/2 & otherwise \end{array} \right\}$$

Clearly, the computational cost peaks when two shapes with a high number of relevant points are compared and one of them has twice the number of relevant points of the other. However, as the difference in the number of relevant points penalizes the similarity, if two QSDs are compared and one of them has twice as many relevant points as the other, the similarity between them will not be higher than 0.5. Moreover, in this approach, we are interested in finding a similarity relation between two shapes, not a dissimilarity one, so comparisons between shapes with a high difference in relevant points can be assigned a low similarity a priori and the calculation of the exact similarity value can be avoided.

### 5.1.3 Application of $SimQSD$: Comparing images from the MPEG-7 CE Shape-1 library

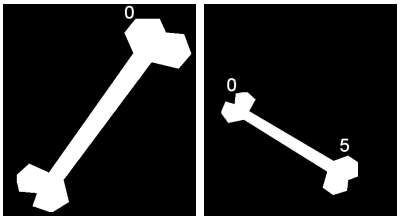
In our experimentation, first the $SimQSD$ is used on some images of the Bone category extracted from the MPEG7 CE Shape-1 library[3] [Latecki et al., 2000] and the similarity values assigned from CNDs or interval distances and the correspondences of relevant points obtained are presented in Section 5.1.3.1. Then, the $SimQSD$ is used to compare the categories Bone, Brick, Glass-Cup, Hammer, Heart and Apple of the MPEG7 CE Shape-1 library (Section 5.1.3.2). Finally, an analysis of the general results of our experimentation is given in Section 5.1.3.3.

#### 5.1.3.1 Similarity Values and Correspondences of Points Between Shapes of Bones

In this section, the $SimQSD$ is tested on some of the images of the Bone category extracted from the MPEG-7 CE Shape-1 library and the results are analysed. These images have been selected because they present interesting aspects to study, such as: deformations, incompleteness and also shapes with high difference in the quantity of relevant points that enable us to test the suitability of the obtained correspondence of relevant points between shapes.

First, our tests show that our approach is invariant to rotations, translations, scaling and combinations of these. The images of bones in Figure 5.9 exemplify this: the similarity calculus started at (0, 5), that is, at relevant point 0 of Bone-1 and at relevant point 5 of Bone-1b. All the relevant points or vertices have a correspondence, because the number of relevant points of both shapes is the same. The $SimQSD = 1.0$, therefore, the shapes of both bones are qualitatively equivalent.

Figure 5.9: Objects with qualitatively equivalent shapes: (a) Bone-1 from the MPEG7 CE Shape-1 library; (b) Bone-1b: Bone-1 translated, rotated and scaled.




(a) Bone-1 (b) Bone-1b

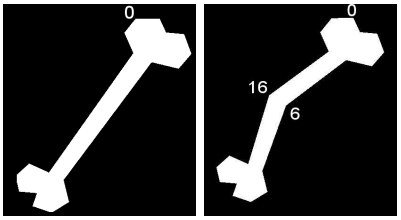
Moreover, it is shown with an example that our approach detects the 'extra' relevant points of a shape intuitively. Given the shapes Bone-1 and Bone-7 (see Figure 5.10), which have a similar shape, the calculation of the $SimQSD$ provides the following results:

- The $SimQSD$ is started at relevant point 1 of Bone-1 and at relevant point 0 of Bone-7, which are the same vertex;

[3] http://www.imageprocessingplace.com/root_files_V3/ image_databases.htm

- The relevant points of Bone-7 with no correspondence in Bone-1 are relevant point 6 and relevant point 16; and,
- The $SimQSD$ between shapes is 0.88 using CNDs and 0.9 using interval distances. A high similarity is obtained by both methods since, Bone-7 is exactly the same as Bone-1 with a bend in it.

Figure 5.10: Two objects from the Bone category in the MPEG7 CE Shape-1 library with a different number of relevant points.




(a) Bone-1 (b) Bone-7

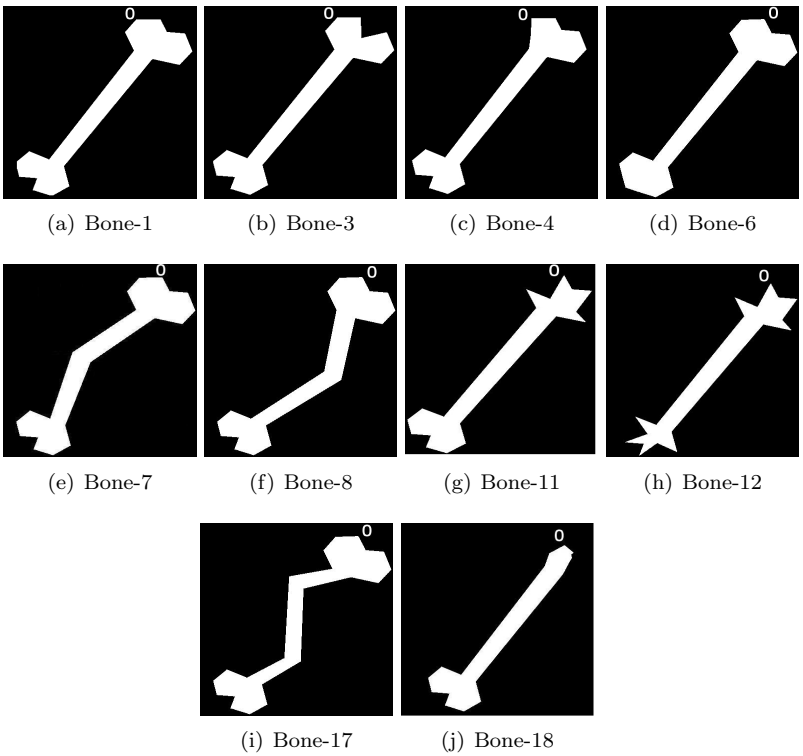
Some other Bone images extracted from the MPEG7 CE Shape-1 library (see Figure 5.11) have been used to calculate the $SimQSD$ between all shapes and to study the obtained correspondence of relevant points. First, using dissimilarity matrices built from CNDs (see Table 5.7) and secondly, using dissimilarity matrices built from CNDs for the features edge connection (EC) and Convexity and using dissimilarity matrices built from interval distances for the features of angle (A), type of curvature (TC) and length (L) (see Table 5.8).

In Table 5.7 and Table 5.8, each cell indicates the $SimQSD$ between the shapes, the starting points of the similarity calculus and the relevant points without correspondence. For example, in Table 5.7 the $SimQSD$ between "Bone-1 and Bone-6" is 0.88, starting the comparison by point 0 of Bone-1 and point 15 of Bone-6 (note that both points have the same location in the images) obtaining that relevant points 10 and 11 of Bone-1 (the shape with the greatest number of relevant points) are compared to the *void* point, that is they have no correspondence with relevant points in Bone-6. In Table 5.8 the $SimQSD$ between "Bone-1 and Bone-6" is 0.89, starting the comparison by point 0 of Bone-1 and point 8 of Bone-6 (another possible alternative) obtaining that relevant points 1 and 2 of Bone-1 are compared to the *void* point.

It can be seen in Table 5.7 and Table 5.8 that the similarities obtained using CNDs or interval distances are very similar. Only a few of the similarity values calculated (marked in bold) are different. Therefore, from now on, only the results of our approach using interval distances will be shown as the weights assigned to the CNDs can be considered a particular case of the interval distances obtained.

Finally, it is important to note that our approach obtains a high similarity value between nearly symmetrical shapes (such as Bone-7 and Bone-8) and it also tackles the problem of deformations and incomplete shapes implicitly. The Bone category of the MPEG-7 library has been chosen because it has some examples of such aspects. Bone-18 can be considered as Bone-1, but incomplete,

Figure 5.11: Images of the Bone category in the MPEG7 CE Shape-1 library used for testing. The starting point of the QSD (0) is also shown.




(a) Bone-1 (b) Bone-3 (c) Bone-4 (d) Bone-6

(e) Bone-7 (f) Bone-8 (g) Bone-11 (h) Bone-12

(i) Bone-17 (j) Bone-18

because the top of the bone that appears in all the other images does not appear in Bone-18. Furthermore, Bone-7, Bone-8 and Bone-17 can be considered as Bone-1, but broken or deformed in the middle. Moreover, the relevant points of one shape that do not have a correspondence in the other shape show where the deformation or the incomplete part of the other shape is. Therefore, our approach can obtain suitable similarities between instances of the same shape that are deformed or incomplete and give an approximate location of such deformation or incompleteness.

#### 5.1.3.2 Similarity Values Between Different Shape Categories of MPEG-7 Library

In this section, an experiment has been performed using *SimQSD* on the images of the categories Bone, Brick, Glass-Cup, Hammer, Heart and Apple from MPEG7 CE Shape-1 library.

In Table 5.9, the similarity values obtained for these shapes within each category is shown, taking Bone-1, Bone-12, Brick-1, Brick-15, Glass-1, Glass-20, Hammer-1, Hammer-12, Heart-1, Heart-20, Apple-1 and Apple-9 as key shapes. Two objects of the same category with different shapes have been chosen to evaluate the performance of our approach in a more accurate way. The table

Table 5.7: *SimQSD* tested on some Bone shapes using dissimilarity matrices built from CNDs.

| **Bone** | **3** | **4** | **6** | **7** | **8** | **11** | **12** | **17** | **18** |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.98 (0,0) ∅ | 0.93 (0,1) {0} | 0.88 (0,15) {10,11} | 0.88 (1,0) {6,16} | 0.88 (1,0) {6,16} | **0.93** (0,17) ∅ | **0.86** (0,17) ∅ | 0.79 (1,0) {5,7, 16,18} | 0.64 (0,0) {0,1,2, 3,6,16} |
| **3** | 1 | 0.92 (0,1) {1} | 0.88 (0,8) {1,2} | 0.88 (1,0) {6,16} | 0.87 (1,0) {6,16} | **0.92** (0,17) ∅ | **0.85** (0,17) ∅ | 0.79 (1,0) {5,7, 16,18} | 0.64 (0,0) {0,2,3, 4,6,16} |
| **4** | | 1 | 0.91 (0,7) {2} | 0.82 (2,0) {0,6,16} | 0.82 (2,0) {0,6,16} | 0.89 (1,0) {17} | **0.82** (1,0) {17} | 0.74 (2,0) {0,5,7, 16,18} | 0.68 (0,11) {0,2,3, 6,16} |
| **6** | | | 1 | 0.79 (0,0) {6,10, 11,16} | 0.78 (9,0) {1,2, 6,16} | **0.81** (8,0) {1,17} | **0.76** (0,0) {10,12} | 0.70 (0,0) {5,7, 12,15, 16,18} | 0.71 (0,5) {6,8, 10,12} |
| **7** | | | | 1 | 0.97 (0,10) ∅ | **0.83** (0,0) {6,16} | **0.77** (0,0) {6,16} | 0.86 (0,0) {7,16} | 0.58 (0,1) {0,1,2, 3,5,6, 17,18} |
| **8** | | | | | 1 | **0.83** (0,0) {6,16} | **0.77** (0,0) {6,16} | 0.87 (0,0) {5,18} | 0.58 (0,0) {0,1,2, 3,5,6, 16,18} |
| **11** | | | | | | 1 | **0.93** (0,0) ∅ | **0.74** (0,1) {0,5,7, 21} | 0.64 (0,1) {0,1,2, 3,4,17} |
| **12** | | | | | | | 1 | **0.69** (1,0) {0,5,7, 21} | **0.57** (0,1) {1,3,5, 12,14, 15} |
| **17** | | | | | | | | 1 | 0.52 (0,10) {0,1,2,3, 5,6,7,16, 19,20} |

showing our results has the following structure: (i) image of the key shape; (ii) images of the same category arranged according to the obtained similarity value.

From Table 5.9, it can be deduced that the *SimQSD* approach is:

i. invariant to scaling, for example, the similarity value obtained when comparing Bone-1 and Bone-13 (the third image in the first row) or between Hammer-1 and the rest of the images of this category is high because, in both situations, both images contain the same shape but the second one has been narrowed;

ii. invariant to translations, proved by the high similarity values obtained for the Brick category, in which nearly all the objects are translated with respect to the key object shape;

iii. invariant to rotations, proved by the high similarity values obtained for the Hammer category, in which the most of the objects are rotated with respect to the key object shape;

iv. invariant to mirror-image reflections because the similarity values obtained between symmetrical shapes (i.e. Hammer-12 and Hammer-8 which is the eighth shape of the eighth row) is high;

v. influenced by the quantity of relevant points contained by the compared shapes: the more difference in relevant points (or the more relevant points compared to the *void* point) the lower the similarity.

Moreover, in Table 5.10, the key shapes used in the previous table are compared and they are arranged according to the highest similarity values obtained. This table has the following structure: (i) image of the key shape; (ii) image of the same category and its corresponding similarity value; (iii) images of the other different categories arranged according to the obtained similarity value.

From Table 5.10, it can be noticed that:

i. the similarity value between shapes of different categories is determined by the quantity of relevant points of each compared shape because the less difference in the number of relevant points of the compared shapes, the higher the similarity (i.e. the shapes of the Bone and Brick category obtain a quite high similarity value because they have nearly the same number of relevant points).

ii. if both shapes have an approximate number of relevant points, the features of shape of each relevant point (i.e. the edge connection, the angle, the convexity, the type of curvature, etc.) are the ones that influences the similarity more, according to the priorities exemplified in Figure 5.7 (i.e. the shapes of the categories Heart and Apple are more similar to each other because they all have curves).

iii. the similarity values obtained are usually higher between shapes within the same category (i.e. shapes of the objects in the first and second column in Table 5.10).

#### 5.1.3.3 Analysis of our Results

After the previous experimentation, the advantages of our approach are summarized here:

- it provides a similarity value between two shapes but also a set of points with no correspondence in the compared shapes;
- it obtains nearly the same similarity values either using CNDs or interval distances;
- it obtains a similarity value between instances of the same shape that are deformed or incomplete and it also gives an approximate location of such deformation or incompleteness because the relevant points of one shape that do not have a correspondence in the other shape show where the deformation or the incomplete part of the other shape is;
- it obtains a high similarity value between translated, rotated, scaled and symmetrical shapes.

From a cognitive point of view, the problem of shape equivalence involves understanding the conditions under which people perceive two distinct objects as having the same shape. In addition to the cognitive definition of shape perception by Wilson and Keil [1999], Palmer [1989] (see Representing Shape and Structure Chapter) considers that two objects have the same *objective shape* even after they have undergone spatial transformations such as translations (changing position), rotations (changing orientation), scaling (expanding and contracting in size), mirror-image reflections (changing direction) and combinations of these transformations. However, if spatial transformations such as squashing, stretching or deforming in any way are needed to bring two objects into exact correspondence, then they have different objective shapes, although they can be perceived as very similar. According to this, our approach has proved to fulfill the requirements of a cognitive perception of shape, because it is invariant to translations, rotations and scaling and also obtains a high similarity value between mirror-image reflections or symmetrical shapes.

Finally, similar arrangements of shapes to those provided by the $SimQSD$ approach in Tables 5.9 and 5.10 are obtained by the approach by Sebastian et al. [2001] when classifying the Bone, Brick, Glass-Cup and Heart category from MPEG7 CE Shape-1 library. Although this approach does not deal with the problem of comparing rotated, translated or scaled shapes, psychological tests would be needed in order to determine which approach assigns similarity values in a more cognitive way. However, from the closeness of the arrangements obtained by $SimQSD$ approach and the arrangements obtained by the shape classifier by Sebastian et al. [2001], it is important to notice that the $SimQSD$ approach could be used as a shape classifier.

#### 5.1.3.4 Discussion

An approach built upon a qualitative model for shape description (QSD) to obtain a similarity measure between shapes was tested using images of the MPEG-7 CE-Shape-1 library. This approach has three steps: (1) comparing qualitative tags related to the same feature of shape by building dissimilarity

matrices using: (a) conceptual neighbourhood diagrams (CNDs); and (b) interval distances; (2) calculating a similarity measure between relevant points; and finally, (3) obtaining a similarity measure between the QSD of the objects by cyclically comparing their relevant points.

Both methods for obtaining dissimilarity matrices for qualitative features of shape (CNDs and interval distances) provide similarity measures that are suitable for our case of study, because there is only a very small difference between them. Furthermore, it is clear that obtaining dissimilarity matrices between qualitative concepts built from CNDs is more intuitive and simpler to calculate. In contrast, dissimilarity matrices between qualitative concepts built from interval distances are more accurate from a mathematical point of view.

The *SimQSD* approach has been tested using images of different categories of the MPEG-7 CE-Shape-1 library and the results obtained show that: (1) the similarity values obtained are invariant to rotations, translations, scaling and mirror-image changes of shapes and also combinations of these; (2) a similarity value can be obtained between deformed or incomplete shapes and the approximate location of the deformation or cut is determined by locating the relevant points with *void* correspondence; and (3) the similarity values obtained by our approach are coherent and cognitive because the lower the difference in shape to the human vision, the higher the similarity.

Table 5.8: *SimQSD* tested on some Bone shapes using dissimilarity matrices built from interval distances for the features of A, TC and L.

| **Bone** | **3** | **4** | **6** | **7** | **8** | **11** | **12** | **17** | **18** |
|---|---|---|---|---|---|---|---|---|---|
| **1** | **1.00** (0,0) ∅ | 0.94 (0,1) {0} | 0.89 (0,8) {1, 2} | 0.90 (1,0) {6,16} | 0.90 (1, 0) {6,16} | **0.97** (0,17) ∅ | **0.94** (0,17) ∅ | 0.80 (1, 0) {5,7, 16,18} | 0.65 (0,0) {1,2,4, 5,6,16} |
| **3** | 1 | 0.94 (0,1) {1} | 0.89 (0,8) {0,2} | 0.90 (1,0) {6,16} | 0.90 (1,0) {6,16} | **0.96** (0,17) ∅ | **0.94** (0,17) ∅ | 0.80 (1,0) {5,7, 17,18} | 0.65 (0,0) {0,1,2, 4,6,16} |
| **4** | | 1 | 0.93 (0,7) {2} | 0.85 (10,0) {6,9,16} | 0.85 (2,0) {0,6,16} | 0.92 (1,0) {17} | **0.90** (1,0) {17} | 0.75 (1,0) {5,7,10, 16,18} | 0.69 (0,11) {0,2,3, 6,16} |
| **6** | | | 1 | 0.80 (9,0) {1,2, 6,16} | 0.80 (9,0) {1,2, 6,16} | **0.87** (9,0) {1,3} | **0.84** (9,0) {1,3} | 0.70 (9,0) {1,2, 5,7, 16,18} | 0.73 (0,5) {6,8, 10,12} |
| **7** | | | | 1 | **0.99** (0,10) ∅ | **0.87** (0,0) {6,16} | **0.84** (0,0) {6,16} | 0.88 (10,0) {5,18} | 0.59 (0,1) {0,1,2, 3,5,6, 16,17} |
| **8** | | | | | 1 | **0.87** (0,0) {6,16} | **0.84** (0,0) {6,16} | 0.89 (0,0) {5,18} | 0.59 (0,0) {0,1,2, 5,16,17, 18,19} |
| **11** | | | | | | 1 | **0.97** (0,0) ∅ | 0.78 (0,9) {9,10, 16,18} | 0.65 (0,11) {1,3,5, 15,16, 17} |
| **12** | | | | | | | 1 | **0.76** (0,0) {9,10, 16,18} | **0.64** (0,11) {1,3,5, 12,14, 15} |
| **17** | | | | | | | | 1 | 0.53 (0,1) {0,1,2,3, 5,6,7,16, 19,20} |

Table 5.9: Results of testing *SimQSD* on some categories from MPEG-7 Shape Library using interval distances.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
|  Bone-1 | 1.00 | 0.99 | 0.99 | 0.99 | 0.97 | 0.94 | 0.94 | 0.90 |
|  Bone-12 | 0.97 | 0.94 | 0.94 | 0.93 | 0.93 | 0.90 | 0.90 | 0.84 |
|  Brick-1 | 0.93 | 0.92 | 0.88 | 0.87 | 0.87 | 0.86 | 0.86 | 0.86 |
|  Brick-15 | 0.91 | 0.89 | 0.89 | 0.88 | 0.87 | 0.87 | 0.87 | 0.86 |
|  Glas-1 | 1.00 | 1.00 | 1.00 | 0.98 | 0.97 | 0.95 | 0.91 | 0.91 |
|  Glas-20 | 0.96 | 0.90 | 0.84 | 0.80 | 0.79 | 0.78 | 0.78 | 0.76 |
|  Hammer-1 | 0.97 | 0.97 | 0.95 | 0.94 | 0.9 | 0.89 | 0.89 | 0.89 |
|  Hammer-12 | 0.94 | 0.92 | 0.91 | 0.91 | 0.91 | 0.9 | 0.9 | 0.9 |
|  Heart-1 | 1.00 | 1.00 | 0.95 | 0.92 | 0.9 | 0.88 | 0.88 | 0.85 |
|  Heart-20 | 0.79 | 0.76 | 0.75 | 0.71 | 0.7 | 0.68 | 0.68 | 0.67 |
|  Apple-1 | 0.9 | 0.87 | 0.84 | 0.8 | 0.8 | 0.78 | 0.78 | 0.77 |
|  Apple-9 | 0.82 | 0.82 | 0.82 | 0.8 | 0.8 | 0.79 | 0.78 | 0.77 |

Table 5.10: Results of testing $SimQSD$ on the key images from MPEG-7 Shape Library used in Table 5.10.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
|  Bone-1 | 0.94 | 0.71 | 0.7 | 0.59 | 0.57 | 0.53 | 0.4 | 0.4 |
|  Brick-1 | 0.87 | 0.7 | 0.68 | 0.58 | 0.57 | 0.53 | 0.51 | 0.51 |
|  Glas-20 | 0.71 | 0.68 | 0.61 | 0.61 | 0.6 | 0.58 | 0.49 | 0.49 |
|  Hammer-12 | 0.87 | 0.76 | 0.61 | 0.59 | 0.57 | 0.51 | 0.51 | 0.39 |
|  Heart-1 | 0.66 | 0.65 | 0.63 | 0.47 | 0.43 | 0.39 | 0.32 | 0.31 |
|  Apple-1 | 0.78 | 0.72 | 0.63 | 0.62 | 0.61 | 0.51 | 0.51 | 0.49 |

## 5.2 Colour Similarity

Qualitative colour names allow to compare the colours of an image in an intuitive way, as human beings do, because colours with similar hue, saturation and ilumination values are given the same name. However, the problem appears when comparing two colour names: How can we define how similar are *red* and *blue* colours? And *grey* and *black*?

This section is organized as follows. The related work on colour similarity is reviewed in subsection 5.2.1 and then, in subsection 5.2.2, a similarity measure between the colour names defined by our Qualitative Colour Model (Section 3.4.2) is presented.

### 5.2.1 Related Work on Colour Similarity

In literature, different colour pixel similarity measures are defined, each one related to a different colour space (for an overview of colour spaces see Section 3.4.1).

Some of the colour similarity measures defined on numerical colour spaces are the following: (i) Euclidean distance is frequently used in cubic representation spaces as RGB or CIE L*ab* and occasionally in cylindric spaces like L*C*H [Sarifuddin and Missaoui, 2005; Vik, 2004]; (ii) a cylindric distance is used for cylindric and conic spaces like HSL, HSV and L*C*H by Plataniotis and Venetsanopoulos [2000]; (iii) the Fuzzy C-Means is used by Seaborn et al. [2005] to define similarity and dissimilarity measures for comparing fuzzy colour categories based on Musell colour space; and (iv) other formulae for computing colour difference in colour spaces as L*C*H and CIECAM02 are proposed by Sarifuddin and Missaoui [2005] and also by Luo et al. [2001].

There are less studies in literature that calculate a similarity measure between colour names. To the best of our knowledge, only psychological studies by Griffin [2001] and Griffin [2006] try to obtain a similarity relation between colour names based on surveys made to people about *which is the most similar pair: A and B or C and D* and then obtain diagrams of the psychological structure of the Basic Colour Terms (BCT).

After learning from all these previous works, it was decided to define an approach for measuring the similarity betweem the colour names of our Qualitative Colour Model (Section 3.4.2) using a conceptual neighbourhood diagram (CND) because: (i) the effectiveness of CNDs was proved in our previous works on shape similarity and (ii) the results of the psicological studies on colour similarity by Griffin [2001] were represented in a diagram of colour connections, which could be used for analising how cognitive is our approach in the future.

### 5.2.2 Similarity between Qualitative Colour Descriptions ($SimQCD$)

Our model for Qualitative Colour Description (QCD) describes colours by dividing the HSL colour coordinates in intervals of Hue, Saturation and Lightness (Section 3.4) and assigning to the corresponding intervals, the following colour names:

$\mathrm{QC}_{LAB_1} = \{$*black, dark_grey, grey, light_grey, white*$\}$

$\mathrm{QC}_{LAB_2} = \{$*red, yellow, green, turquoise, blue, purple, pink*$\}$
$\mathrm{QC}_{LAB_3} = \{$*pale_* + $\mathrm{QC}_{LAB_2}\}$
$\mathrm{QC}_{LAB_4} = \{$*light_* + $\mathrm{QC}_{LAB_2}\}$
$\mathrm{QC}_{LAB_5} = \{$*dark_* + $\mathrm{QC}_{LAB_2}\}$

From the qualitative colours defined by our approach, the Conceptual Neighbourhood Diagram or CND shown in Figure 5.12 can be built. This CND is tridimensional and it has the shape of a double cone, as HSL colour space (see Figure 3.6). The rainbow colours described in this CND (*red, yellow, green, turquoise, blue, purple, pink*) 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* 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 (*black, dark_grey, grey, light_grey, white*).

In this CND, the paths connecting pairs of adjacent nodes that map to continuous transformations have weights assigned in order to establish priorities:

- 1 is the weight assigned between a colour name and the same colour name with a semantic prefix (*pale_, light_, dark_*). For example: $dsColour(red, light\_red) = 1$ and $dsColour(grey, dark\_grey) = 1$.
- 3 is the weight assigned between different colour names in the rainbow scale without or with prefix (*pale_, light_, dark_*). For example: $dsColour(pink, red) = 3$ and $dsColour(pale\_pink, pale\_red) = 3$.
- 5 is the weight assigned in the transition between the rainbow colour scale and the grey scale. For example: $dsColour(pale\_red, grey) = 5$, $dsColour(light\_yellow, light\_grey) = 5$ and $dsColour(dark\_blue, dark\_grey) = 5$.
- 5 is the weight assigned to the different colour names in the grey scale. For example: $dsColour(black, dark\_grey) = 5$.

Finally, according to all these weights, dissimilarity matrices which map the pairs of nodes in the CND to the minimal path distance between them are shown in Table 5.11 - 5.15.

Table 5.11: Dissimilarity matrix for qualitative colours in the grey scale.

| **Grey Scale** | *black* | *dark_grey* | *grey* | *light_grey* | *white* |
|---|---|---|---|---|---|
| *black* | 0 | 5 | 6 | 7 | 12 |
| *dark_grey* | 5 | 0 | 1 | 2 | 7 |
| *grey* | 6 | 1 | 0 | 1 | 6 |
| *light_grey* | 7 | 2 | 1 | 0 | 5 |
| *white* | 12 | 7 | 6 | 5 | 0 |

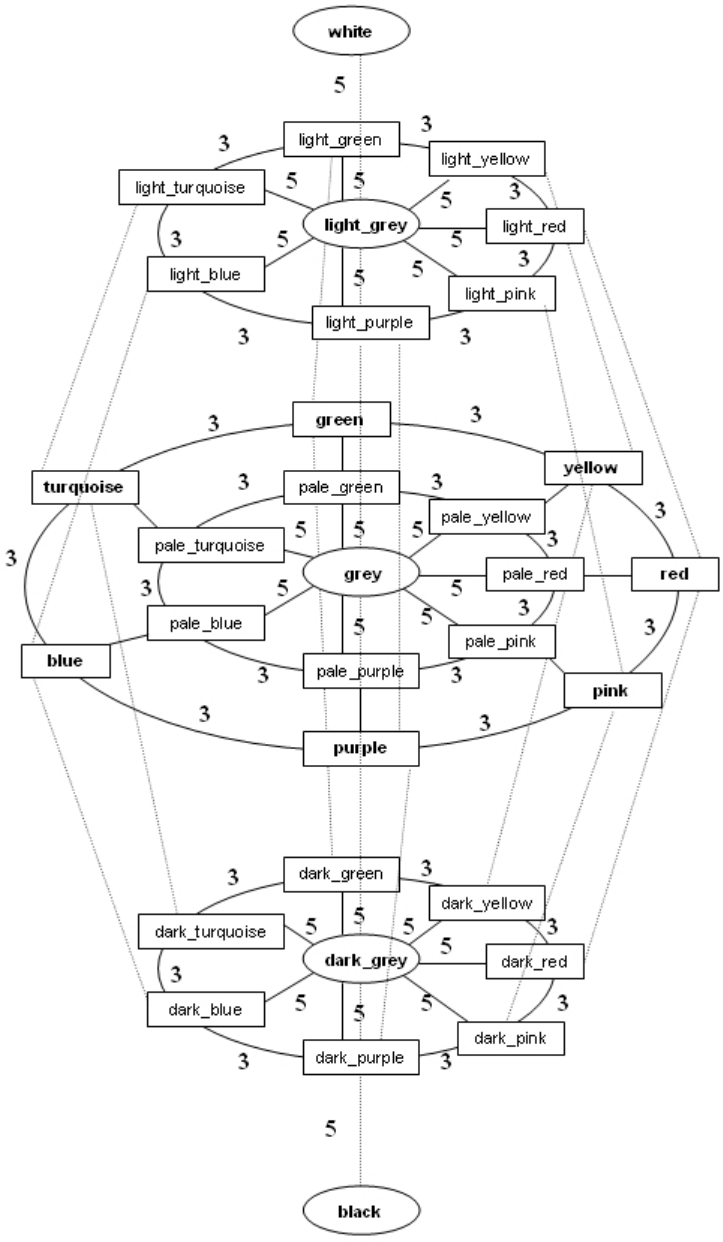



Figure 5.12: CND for our model for Qualitative Colour Description (QCD). Paths with no weight drawn are 1 by default.

Table 5.11 shows the dissimilarity matrix for the transformations for the qualitative colours in the grey scale (*black, dark_grey, grey, light_grey, white*) which correspond to the vertical central nodes of the CND in Figure 5.12.

Table 5.12: Dissimilarity matrix for qualitative colours in the rainbow scale, with or without a prefix.

| **Rainbow Scale (d)** | *(p-) red* | *(p-) yellow* | *(p-) green* | *(p-) turquoise* | *(p-) blue* | *(p-) purple* | *(p-) pink* |
|---|---|---|---|---|---|---|---|
| *(p-) red* | 0 | 3 | 6 | 9 | 9 | 6 | 3 |
| *(p-) yellow* | 3 | 0 | 3 | 6 | 9 | 9 | 6 |
| *(p-) green* | 6 | 3 | 0 | 3 | 6 | 9 | 9 |
| *(p-) turquoise* | 9 | 6 | 3 | 0 | 3 | 6 | 9 |
| *(p-) blue* | 9 | 9 | 6 | 3 | 0 | 3 | 6 |
| *(p-) purple* | 6 | 9 | 9 | 6 | 3 | 0 | 3 |
| *(p-) pink* | 3 | 6 | 9 | 9 | 6 | 3 | 0 |

Table 5.12 shows the matrix containing the dissimilarities between the qualitative colours in the rainbow scale (*red, yellow, green, turquoise, blue, purple, pink*) without prefix or with the same prefix (denoted as *p-* which refers to *pale_*, *light_* or *dark_*). Note that the resulting dissimilarity of this matrix is denoted as $d$ and it will be used later on.

Table 5.13 shows the matrix for obtaining the dissimilarities between qualitative colours in rainbow scale (located in the external central circle) and the *light_/pale_/dark_* qualitative colours in the rainbow scale (located in the three central circles located above/in the middle/below in the CND, respectively). The parameter denoted as $rc$ corresponds to the rainbow colour labels (*red, yellow, green, turquoise, blue, purple, pink*). The dissimilarity denoted as $d$ is the result of the dissimilarity matrix shown in the Table 5.12.

Table 5.14 shows the dissimilarities between the qualitative colours in the rainbow scale (nodes connected to the external central circle) and the qualitative colours in the grey scale (nodes connected to the vertical central line).

Finally, Table 5.15 shows the dissimilarities between the qualitative colours in the grey scale (nodes connected to the vertical central line) and the qualitative *light_/pale_/dark_* colours in the rainbow scale (nodes connected to the three central circles located above/in the middle/below in the CND, respectively).

Therefore, given two qualitative colours, denoted by $QC_A$ and $QC_B$, referring to the colours of the objects A and B respectively, a similarity between them, denoted by $SimQCD(QC_A, QC_B)$, is defined as:

Table 5.13: Dissimilarity matrix for qualitative colours in rainbow scale with different prefixes.

| **Rainbow Scale any Prefix Scale** | *rc* | *pale + rc* | *light + rc* | *dark + rc* |
|---|---|---|---|---|
| *rc* | $d$ | $d+1$ | $d+1$ | $d+1$ |
| *pale + rc* | $d+1$ | $d$ | $d+2$ | $d+2$ |
| *light + rc* | $d+1$ | $d+2$ | $d$ | $d+2$ |
| *dark + rc* | $d+1$ | $d+2$ | $d+2$ | $d$ |

Table 5.14: Dissimilarity matrix for qualitative colours in rainbow scale and grey scale.

| **Rainbow Scale to Grey Scale** | *black* | *dark_grey* | *grey* | *light_grey* | *white* |
|---|---|---|---|---|---|
| *red* | 11 | 6 | 6 | 6 | 11 |
| *yellow* | 11 | 6 | 6 | 6 | 11 |
| *green* | 11 | 6 | 6 | 6 | 11 |
| *turquoise* | 11 | 6 | 6 | 6 | 11 |
| *blue* | 11 | 6 | 6 | 6 | 11 |
| *purple* | 11 | 6 | 6 | 6 | 11 |
| *pink* | 11 | 6 | 6 | 6 | 11 |

Table 5.15: Dissimilarity matrix for the qualitative colours in the grey scale and the qualitative *light_/pale_/dark_* colours in the rainbow scale.

| **Prefix Scale to Grey Scale** | *black* | *dark_grey* | *grey* | *light_grey* | *white* |
|---|---|---|---|---|---|
| *pale + rc* | 11 | 6 | 5 | 6 | 11 |
| *light + rc* | 12 | 7 | 6 | 5 | 10 |
| *dark + rc* | 10 | 5 | 6 | 7 | 12 |

$$SimQCD(QC_A, QC_B) = 1 - \frac{dsColour(QC_A, QC_B)}{MaxDsColour}, \tag{5.5}$$

where $dsColour(QC_A, QC_B)$ denotes the dissimilarity between the qualitative colour names obtained from the dissimilarity matrices previously defined. *MaxDsColour* denotes the maximum dissimilarity for all colour names, which is 12 for

our case of study. Hence, by dividing $dsColour(QC_A, QC_B)$ and $MaxDsColour$ the proportion of dissimilarity related to qualitative colours $\mathrm{QC}_A$ and $\mathrm{QC}_B$ is obtained, which has values between 0 and 1. Finally, this value is substracted from 1 with the aim of provinding a similarity instead of a dissimilarity.

The computational cost (CC) of calculating the similarity of two qualitative colours is $O(1)$.

The main properties of this final similarity are:

- Symmetry: $SimQCD(QC_A, QC_B) = SimQCD(QC_B, QC_A)$.
- Upper and lower bounds: $0 \leq SimQCD(QC_A, QC_B) \leq 1$.
- Intuitive: $SimQCD(QC_A, QC_B) = 0$ means that $dsColour(QC_A, QC_B) = MaxDsColour$, that is, both colours are as different as possible.

The following examples of $SimQCD$ shown some intuitive properties (from the point of view of human thinking) of our colour similarity approach:

- the null similarity is given between *black* and *white* and also between any *light* rainbow colour ($rc$) and *black* and any *dark* $rc$ and *white*:

  $SimQCD(white, black) = 1 - 12/12 = 0$

  $SimQCD(light + rc, black) = SimQCD(dark + rc, white) = 1 - 12/12 = 0$

- the same similarity is given between any $rc$ and *black/white* or any *pale* $rc$ and *black/white*:

  $SimQCD(rc, black) = SimQCD(rc, white) = 1 - 11/12 = 0.08$

  $SimQCD(pale + rc, black) = SimQCD(pale\_rc, white) = 1 - 11/12 = 0.08$

- the same similarity is given between any *light* $rc$ and *white* and any *dark* $rc$ and *black*:

  $SimQCD(light + rc, white) = SimQCD(dark + rc, black) = 1 - 10/12 = 0.16$

- the same similarity is given between any $rc$ and the same *dark*, *pale* or *light* $rc$:

  $SimQCD(pale + rc, rc) = SimQCD(light + rc, rc) = SimQCD(dark + rc, rc) = 1 - 1/12 = 0.92$

- the same similarity is given between any *prefix* (*pale*, *dark* or *light*) of the same $rc$:

  $SimQCD(pale + rc, dark + rc) = SimQCD(pale + rc, light + rc) = SimQCD(dark + rc, light + rc) = 1 - 2/12 = 0.83$

- the same similarity is given between any *pale* $rc$ and *grey*, and between any *light* $rc$ and *light_grey*, and between any *dark* $rc$ and *dark_grey*:

  $SimQCD(pale + rc, grey) = SimQCD(light + rc, light\_grey) = SimQCD(dark + rc, dark\_grey) = 1 - 5/12 = 0.58$

- any *light* $rc$ is more similar to *white* than any *pale* $rc$ to *white* and, in the same way, any *dark* $rc$ is more similar to *black* than any *pale* $rc$ to *black*:

  $SimQCD(light + rc, white) > SimQCD(pale + rc, white)$

  $SimQCD(dark + rc, black) > SimQCD(pale + rc, black)$

### 5.2.3 Application of $SimQSD$ and $SimQCD$: A Pragmatic Approach for Assembling Tile Mosaics by Qualitative Shape and Colour Similarity Matching

A mosaic is a decorative art form in which small ceramic tiles are assembled to form a predefined image. An approach for describing qualitatively the shape of any ceramic tile piece and its RGB colour was provided by Museros and Escrig [2004] and a prototype for assemblying ceramic mosaics automatically by exact matching of qualitative shape descriptions and RGB colour values was developed by Museros and Escrig [2007].

However, as some of the qualitative terms used in a description of shape are related to an interval of values (for example, length, angular amplitude, etc.), sometimes two very close numerical values to be qualified can be: one close to the ending point of an interval and the other one close to the starting point of another interval, and therefore each one can belong to different intervals of values and different qualitative terms can be obtained for those very close numerical values. For avoiding this problem an approach for measuring the similarity between qualitative terms is needed.

In this section, the application of assembling tile mosaics is revisited in order to test the suitability of: (i) the new model for qualitative shape description presented in Section 3.3.2, which is an improvement of the previous model by Museros and Escrig [2004]; and (ii) the approach for measuring the similarity between qualitative descriptions of shape presented in Section 5.1. For completing the qualitative description of a tile piece of a mosaic, the qualitative colour of each tile (instead of its RGB colour) is obtained using the model in Section 3.4.2, and hence, the colour similarity measure presented in Section 5.2.2 is applied and tested.

#### 5.2.3.1 Scenario

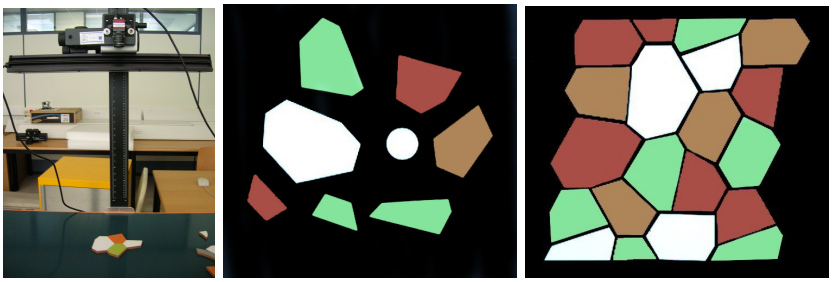
In our scenario, some tiles are placed on a working table above which an industrial camera is located as Figure 5.13 (a) shows. A digital image of the table is taken by the camera (Figure 5.13 (b) shows an example) and the shape and the colour of each tile piece located there is described qualitatively by our algorithms. Then these qualitative descriptions are compared to the description of the pieces of the mosaic to assemble (Figure 5.13 (c) showns an example) and a matching degree is obtained.

Specifically, this industrial application uses:

- the model for Qualitative Shape Description (QSD) (Section 3.3.2) for describing the shape of the tile pieces;
- the approach for obtaining a similarity measure between $QSD$s ($SimQSD$s) based on matrices built from conceptual neighbourhood diagrams (CND) (Section 5.1.2), since it obtained successful results after testing on MPEG-7 library (Section 5.1.3);
- the model for Qualitative Colour Description (QCD) (Section 3.4.2) for the description of the colour of tile pieces. As tiles are closed objects and have uniform colours, the HSL colour coordinates of the centroid of the object are representative enough to give a name to the perceptual colour of the object;

- the approach for obtaining a similarity measure between $QCD$s ($SimQCD$s) based on matrices built from conceptual neighbourhood diagrams (CND) (Section 5.2.2).

Figure 5.13: Our scenario for assembling tile mosaics: (a) working table and industrial camera; (b) image containing real tiles used to assemble the mosaic design; and (c) mosaic designed using a graphics editing program.



(a) Working table and industrial camera

(b) Image captured by the camera

(c) Mosaic design

#### 5.2.3.2 Matching Algorithm using Qualitative Shape and Qualitative Colour Similarity

Our approach can calculate the matches of the tile pieces captured by the camera with its corresponding place in the mosaic by taking into account:

- shape and colour similarity: the QSD and QCD of each tile is compared to all the pieces composing the mosaic design. If the similarity between qualitative colour descriptions ($SimQCD$) is higher than a threshold determined by experimentation, then the similarity between qualitative shape descriptions ($SimQSD$) is calculated, otherwise, the matching is discarded. If the $SimQSD$ is higher than a threshold also determined by experimentation, then the matching is done, otherwise it is discarded.

- only shape similarity: the QSD of each tile is compared to all the pieces composing the mosaic design. If the $SimQSD$ is higher than a threshold defined by experimentation, then the matching is done, otherwise it is discarded.

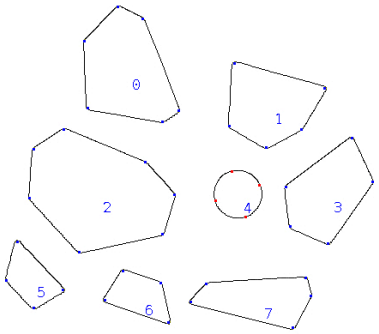
#### 5.2.3.3 Results

After testing our approach, promising results are obtained.

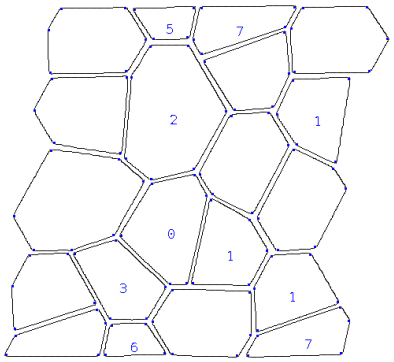
Figure 5.14 (b) shows an example of the matches obtained with a $SimQCD > 0.75$ and $SimQSD > 0.98$. All the matches done in this situation correspond to correct locations in the mosaic design. Therefore, the results of our matching algorithm are successful.

Figure 5.14 (c) shows the matching results only considering shape similarity with a $SimQSD > 0.80$. In this situation, two aspects are important:

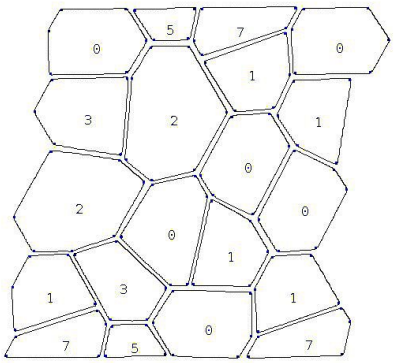
Figure 5.14: Results on our scenario for assembling tile mosaics: (a) description obtained from the image captured by the industrial camera; (b) shape and colour matching results; (c) only shape matching results.




(a) Image processed by our approach




(b) Results if $SimQCD > 0.75$ and $SimQSD > 0.98$




(c) Results if $SimQSD > 0.80$

- The first one is that tile 1 is match to a place in the mosaic design corresponding to a tile with the same shape but with different colour (instead of *red*, *white*), which is an interesting result which demonstrates that our approach can generate perceptually visual different mosaics only by changing the colour of the tiles which compose it.
- The second important aspect is that false positive matches are obtained for tile 3 and tile 0 because of the lower value of the $SimQSD$ threshold and the high perceptual similarity of the shape of the mistaken tiles. However these false positive matches are controlled by a threshold and they are not a problem for our approach.

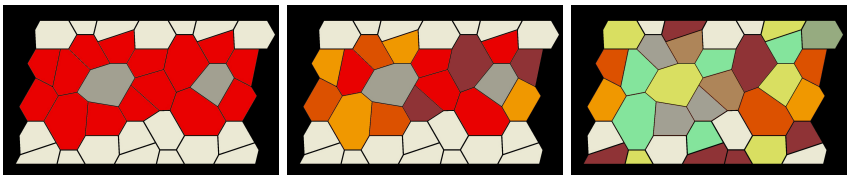
The approach presented can accelerate the process of mosaic assembling of Museros and Escrig [2007] because the time spent in waiting for a tile piece that provides an exact matching of qualitative descriptions is not needed, instead

approximate matchings with high similarity values (for avoiding false positives) can be used successfully.

Figure 5.15 shows an interesting result of the presented approach: by relaxing the colour similarity threshold ($ColourThreshold$), perceptually visual different mosaics can be generated. Figure 5.15 shows: (a) the initial mosaic design, (b) the mosaic produced by the approximate matching algorithm presented with a $SimQCD \geq 0.75$, and (c) the mosaic produced considering only $SimQSD \geq 0.98$. Note that, in the last mosaic produced (Figure 5.15 (c)), each tile has been assembled in a free place of the mosaic design according to its shape, and the order of appearance on the working table.

Finally, note that, in our current approach, we are not managing the size of the tiles because we know a priori that all the tiles correspond to the mosaic design and that there are not tiles with the same shape and different size.

Figure 5.15: Results of the application relaxing the colour similarity measure.



(a) (b) (c)

#### 5.2.3.4 Discussion

A first approximation to an approach for assembling tile mosaics by shape and colour similarity matching based on conceptual neighbourhood diagrams has been presented and tested in a real application and the results obtained are promising. A remarkable result of the current approach is that it can produce perceptually visual different mosaics from a given design by just relaxing the colour similarity threshold.

As future work, we intend to: (1) test our matching algorithm using the shape similarity and colour similarity calculated from matrices of intervalar distances and compare the results obtained in this case with the current results; and (2) use the false positive matches obtained when decreasing the shape similarity threshold for assembling trencadís mosaics. In this application, we will manage the size of the tiles in the working table with respect to their possible location in the trencadís mosaic design.

## 5.3 Spatial Similarity

The spatial features considered in our approach for Qualitative Image Description (QID) for describing any region within a digital image are: (1) its topology relations (Section 3.5) and its fixed and relative orientation (Section 3.6).

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.

Although our approach for QID provides two kinds of orientation for each region or object within the image (fixed and relative), to start to tackle the problem of orientation similarity we deal only with the fixed orientation relations of objects with respect to the centre of the image. The rest of the orientation descriptions are left for future work.

Therefore, this section is organized as follows. Related work on spatial similarity is summarized in Section 5.3.1. A similarity measure between topological descriptions is presented in Section 5.3.2 and a similarity measure between fixed orientation descriptions with respect to the image is given in Section 5.3.3.

### 5.3.1 Related Work on Spatial Similarity

In literature, different approaches that obtain a degree of similarity between two spatial descriptions can be found.

Dealing with topological similarity, Egenhofer and Al-Taha [1992] reason about gradual changes to binary topological relationships caused by deformations as translations, rotations, reductions and/or expansions of an object. A conceptual neighbourhood diagram (CND) is defined according to the deformation applied and a table containing the topology distance between the eight topological relationships considered is also defined.

Dealing with orientation or directional similarity, the approach by Goyal and Egenhofer [2001] determines the directional similarity between extended spatial objects, based on the direction-relation matrix by Goyal and Egenhofer [2000] and combines the qualitative model based on projections by Mukerjee and Joe [1990b] and the cardinal orientation approach by Frank [1991].

Moreover, in literature there are also approaches that calculate the spatial similarity of two scenes or images by studying the relationships of topology, direction/orientation and distance of the objects contained in them [Bruns and Egenhofer, 1996; Papadias and Delis, 1997; Randell and Witkowski, 2004; Li and Fonseca, 2006]. Bruns and Egenhofer [1996] present a method to assess the spatial similarity of two scenes based on the number of changes required to transform one scene into another. They combine qualitative models of topology, distance and direction/orientation into a single model and compute similarity between scenes containing the same number of objects. The approach by Papadias and Delis [1997] calculates the spatial similarity between a query and the images contained in a data base. This approach is based on relations of topology, direction/orientation and distance and also uses the notion of conceptual neighbourhood in N-dimensional spaces. Randell and Witkowski [2004] present an approach for tracking regions over time in indexed images that is based on similarity measures defined according to the conceptual neighbourhood diagrams (CND) of (1) the topological relations defined by the Region

Connection Calculus (RCC-8) [Randell and Cohn, 1989; Randell et al., 1992b], and (2) the direction relations left/right and above/below. By implementing and testing this approach in a simulator, they prove that, by composing the CND of the three models, the ambiguity in identity mappings of the regions in the images is reduced. The model by Li and Fonseca [2006] calculates a spatial similarity degree based on findings of psychological similarity research by considering: (i) commonalities between the stimulus pairs, (ii) structural alignments, (iii) inter/intra-group transformation costs, and (iv) an order of priority between the spatial features (topology < direction < distance).

All these works have inspired our studies and have been a reference for defining similarity measures between the qualitative concepts of topology and fixed orientation described by our approach for QID. However, those measures have been defined for and influenced by our application.

### 5.3.2 Similarity between Topological Descriptions ($SimTop$)

Our approach for qualitative image description (QID) gives the topological relationship of an object A with respect to (wrt) the location in space of another object B (A wrt B) and distinguishes four relationships between the objects:

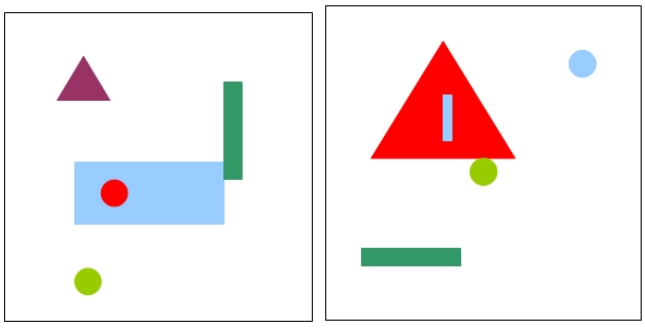
$$T_{LAB} = \{disjoint,\ touching,\ completedly\_inside,\ container\}$$

When referring to the topological situation of two objects, A and B, located respectively in two different images, $\mathrm{Im}_A$ and $\mathrm{Im}_B$, what is described is:

- the *containers* that objects A and B have,
- the components that A and B have (the other objects located *completedly_inside* A and B),
- the objects that are *touching* A and B, and
- the objects that are *disjoint* A and B.

In order to compare the topological situations of objects A and B, the four relationships described should be considered.

Figure 5.16: Drawings containg objects with different topological situations.



(a) Drawing 1 (b) Drawing 2

Let us consider the two drawings in Figure 5.16. If we compare the topological situations of the rectangle in Drawing 1 and the triangle in Drawing 2, we can observe that they have both: (i) a *container* (the image itself); (ii) an object located *completedly inside*; (iii) a neighbour *touching*, and (iv) two neighbours *disjoint*. Therefore, the rectangle in Drawing 1 and the triangle in Drawing 2 have the same topological situation ($SimTop = 1$).

According to the aforementioned, the important information provided by our QID approach that has to be considered to calculate the topological similarity of two objects is: (i) the quantity ($Q$) of *containers*; (ii) the quantity of objects *completedly inside* or components; (iii) the quantity of neighbours *touching*; and (iv) the quantity of neighbours *disjoint* that each object has. By considering the quantity of objects rather than which objects are exactly, a degree of similarity depending only in the topological situation in the space of a specific object A wrt another specific object B is calculated.

Therefore, the topological similarity of two objects A and B, with respect to its *containers*, objects *completedly inside* and neighbours *touching* or *disjoint* is defined as:

$$TopRel(x, A, B) = \left\{ \begin{array}{cc} 1 & if\ Q(x, A)\ =\ Q(x, B) \\ \frac{MinQ(x,A,B)}{MaxQ(x,A,B)} & otherwise \end{array} \right\}$$

where $x$=\{*containers*, *completedly_inside*, *disjoint*, *touching*\}, $Q$ is the quantity of $x$, $MinQ$ is the minimun quantity of $x$ of A and B, and finally $MaxQ$ is the maximum quantity of $x$ of A and B that is always different from zero. Note that it is hold that $0 \leq TopRel(x, A, B) \leq 1$ because $MinQ(x, A, B) \leq MaxQ(x, A, B)$.

Therefore, the similarity of two objects A and B, with respect to all its topological relations is:

$$\begin{array}{rcc} SimTop(A, B) & = & w_C \cdot TopRel(containers, A, B) \\ & + & w_{CI} \cdot TopRel(completedly\_inside, A, B) \\ & + & w_D \cdot TopRel(disjoint, A, B) \\ & + & w_T \cdot TopRel(touching, A, B) \end{array} \tag{5.6}$$

where $w_C + w_{CI} + w_D + w_T = 1$ and, therefore, it is straightforward to prove that $0 \leq SimTop(A, B) \leq 1$. For our application $w_C = w_{CI} = w_D = w_T = 0.25$.

As an example, the similarity between the triangle in Drawing 1 ($T_{D1}$) and the triangle in Drawing 2 ($T_{D2}$) is calculated as:

$TopRel(container, T_{D1}, T_{D2}) = 1$
$TopRel(completedly\_inside, T_{D1}, T_{D2}) = 0/1 = 0$
$TopRel(touching, T_{D1}, T_{D2}) = 0/1 = 0$
$TopRel(disjoint, T_{D1}, T_{D2}) = 2/3$
$SimTop(T_{D1}, T_{D2}) \approx 0.42$

Finally, the computational cost of calculating the similarity of the topology situation of two objects ($CC(SimTop)$) is the cost of doing four comparisons (wrt containers, components, neighbours disjoint or touching) that is $O(4)$.

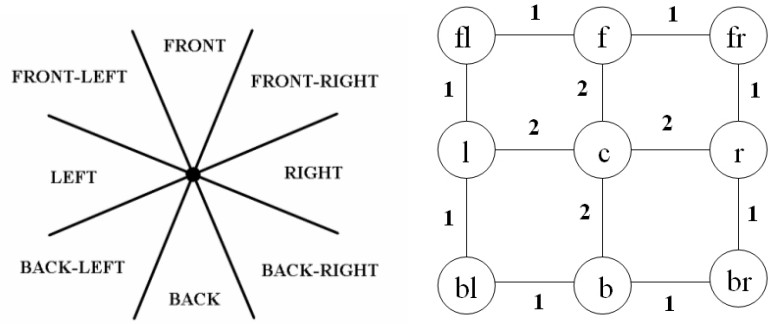
### 5.3.3 Similarity between Orientations ($SimFO$)

Our approach for qualitative image description (QID) provides the orientation of an object A wrt its container by distinguishing eight regions labelled as:

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

From the orientation model by Hernández [1991] (shown in Figure 5.17 (a)), the Conceptual Neighbourhood Diagram or CND shown in Figure 5.17 (b) can be built. The weights in our CND have been established in order to get the maximum dissimilarity (zero) between opposite orientation relationships, that is: *front* vs. *back*, *right* vs. *left*, *front-left* vs. *back-right* etc. Therefore, the dissimilarity matrix obtained from this CND is that shown in Table 5.16.

Figure 5.17: (a) Hernandez's orientation model; (b) CND for fixed orientation.




(a) (b)

Table 5.16: Dissimilarity matrix for fixed orientation using a CND.

| ***Orientation*** | *f* | *fr* | *r* | *br* | *b* | *bl* | *l* | *fl* | *c* |
|---|---|---|---|---|---|---|---|---|---|
| *f* | 0 | 1 | 2 | 3 | 4 | 3 | 2 | 1 | 2 |
| *fr* | 1 | 0 | 1 | 2 | 3 | 4 | 3 | 2 | 3 |
| *r* | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 3 | 2 |
| *br* | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 3 |
| *b* | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 2 |
| *bl* | 3 | 4 | 3 | 2 | 1 | 0 | 1 | 2 | 3 |
| *l* | 2 | 3 | 4 | 3 | 2 | 1 | 0 | 1 | 2 |
| *fl* | 1 | 2 | 3 | 4 | 3 | 2 | 1 | 0 | 3 |
| *c* | 2 | 3 | 2 | 3 | 2 | 3 | 2 | 3 | 0 |

Therefore, given two fixed orientations, denoted by $FO_A$ and $FO_B$, referring to the orientations of the objects A and B respectively, a similarity between them, denoted by $SimFORel(FO_A, FO_B)$, is defined as:

$$SimFORel(FO_A, FO_B) = 1 - \frac{dsFORel(FO_A, FO_B)}{MaxDSimFO} \tag{5.7}$$

where $dsFORel(FO_A, FO_B)$ denotes the dissimilarity between the fixed orientations obtained from the dissimilarity matrix previously defined. $MaxDSimFO$ denotes the maximum dissimilarity for all fixed orientations, which is 4 for our case of study. Hence, by dividing $dsFORel(FO_A, FO_B)$ and $MaxDSimFO$ the proportion of dissimilarity related to fixed orientations $FO_A$ and $FO_B$ is obtained, which has values between 0 and 1. Finally, this value is substracted from 1 with the aim of provinding a similarity instead of a dissimilarity. Therefore $0 \leq SimFORel(FO_A, FO_B) \leq 1$.

As an example, let us consider the FO of the Triangle in Drawing 1 ($T_{D1}$) and the FO of the blue Circle in Drawing 2 ($C_{D2}$):

$T_{D1} = \{front\_left(fl)\}$
$C_{D2} = \{front\_right(fr)\}$
$SimFORel(fl, fr) = 1 - \frac{2}{4} = 0.5$

In this case, as the fixed orientation of each object is determined by only one label, the orientation similarity between both objects $T_{D1}$ and $C_{D2}$ is the similarity between the labels, that is $SimFORel(fl, fr) = 0.5$.

However, if the quantity of fixed orientations of the objects A and B are $n$ and $m$ respectively, and $n \geq m$, the $n$-$m$ fixed orientations of A are compared with the *void* orientation and their similarity is 0, whereas the rest are compared with the fixed orientations of B as shown in (5.7).

Finally, the similarity of two objects A and B, with respect to their fixed orientations is:

$$SimFO(A, B) = \sum_{FO_A \in A, FO_B \in B}^{n} \frac{SimFORel(FO_A, FO_B)}{n} \tag{5.8}$$

As an example, let us consider the FO of the Triangle in Drawing 1 ($T_{D1}$) and the FO of the green Rectangle in Drawing 2 ($R_{D2}$):

$T_{D1} = \{front\_left(fl)\}$
$R_{D2} = \{back\_left(bl), back(b)\}$
$SimFORel(fl, bl) = 1 - \frac{2}{4} = 0.5$
$SimFORel(void, b) = 0$
$SimFO(T_{D1}, R_{D2}) = (0.5 + 0)/2 = 0.25$

In this case, the fixed orientation similarity between both objects $SimFO$ ($T_{D1}$, $R_{D2}$) is 0.25.

Finally, the computational cost (CC) of calculating the similarity of the fixed orientation situation of two objects with respect to the image is $O(n)$ that is the greater number of relevant points of the objects compared.

## 5.4 Image Similarity

From a cognitive point of view, when human beings compare two images, they can determine that both images are: (i) equal, if all the objects and properties of the objects within both images are the same; (ii) similar, if both images share objects and properties; and (iii) different, if they have nothing in common. Determining that two images are equal or completely different can be relatively easy. However, determining how similar two images are, it is a more complex task, because, as Goodman suggested *'X is similar to Y' is totally unconstrained until it is completed by an explanatory clause as 'with respect to property Z'* (*Similarity* entry in *The MIT Encyclopedia of the Cognitive Science* [Wilson and Keil, 1999]). For example: (i) a banana and an apple are similar because they are both fruits; (ii) a banana and a tennis ball are similar because they are both yellow; and (iii) an apple and a tennis ball are similar because they are both round. Therefore, it is important to define the property used to compare objects in order to obtain a useful degree of similarity between them. In the same way, when comparing two Qualitative Image Descriptions ($QID$s) is important to determine with respect to which properties the comparison is done and which are the most suitable ones for obtaining an appropriate explanation of similarity degrees for each scenario.

The Qualitative Image Description approach (see Chapter 3) describes the qualitative shape, colour, topology and orientation of the objects within any digital image. In previous sections, measures of similarity between qualitative shape ($SimQSD$), colour ($SimQCD$), topology ($SimTop$) and orientation ($Sim\ FO$) descriptions have been defined. Now it is necessary to study in which degree these properties are taken into account for obtaining similarities between objects within an image and between images themselves. This section tackles with this problem by defining a similarity measure between qualitative image descriptions ($QID$) (Section 5.4.1) and applying it to three robotic scenarios (Section 5.4.2).

### 5.4.1 Similarity between Qualitative Image Descriptions ($SimQID$s)

In order to define a similarity measure between images described by the QID approach, first a similarity measure between objects described by its shape, colour, topology and orientation must be obtained.

Hence, given two objects, denoted by $A$ and $B$, a similarity between them, denoted by $SimObj(A, B)$, is defined as:

$$\begin{aligned} SimObj(A,B) &= w_{QSD} \cdot SimQSD(A,B) \\ &+ w_{QCD} \cdot SimQCD(A,B) \\ &+ w_{Top} \cdot SimTop(A,B) \\ &+ w_{FO} \cdot SimFO(A,B) \end{aligned} \tag{5.9}$$

where the parameters $w_{QSD}$, $w_{QCD}$, $w_{Top}$ and $w_{FO}$ are the weights assigned to the shape simarity ($SimQSD$), the colour similarity ($SimQCD$), the topology similarity ($SimTop$) and the orientation similarity ($SimFO$), respectively. Moreover, it is hold that $w_{QSD} + w_{QCD} + w_{Top} + w_{FO} = 1$. Clearly, these weights can be tuned in order to give more importance to one similarity (shape, colour, topology or orientation) over the others. Finally, for each $A$ and $B$,

it is straightforward to prove that $0 \leq SimObj(A,B) \leq 1$ and that this is a symmetrical relation, that is, $SimObj(A,B) = SimObj(B,A)$.

The computational cost (CC) of calculating the similarity of the qualitative shape and colour and the topological and fixed orientation situation of two objects is the following:

$$CC(SimObj(A,B)) = CC(SimQSD(A,B)) + CC(SimQCD(A,B)) +$$

$$CC(SimTop(A,B)) + CC(SimFO(A,B)) =$$

$$O(n^2) \text{ or } O(n^{r+1}) + O(1) + O(4) + O(n) = O(n^2) \text{ or } O(n^{r+1})$$

Therefore, the $CC(SimObj(A,B))$ is determined by the CC of the shape similarity calculus, that is, in the best case $O(n^2)$ and in the worst case $O(n^{r+1})$ where $n$ is the number of relevant points of object A and $r = n/2$ if $n$ is odd and $r = (n+1)/2$ otherwise.

Furthermore, in order to compare two images $Im_A$ and $Im_B$ whose QIDs have the same number of objects (denoted by $N$), the similarity between $Im_A$ and $Im_B$ is calculated from (5.9) as an arithmetic mean of the similarity between objects:

$$SimQID_\sigma(Im_A, Im_B) = \frac{1}{N}\sum_{i=1}^{N} SimObj(A_i, B_{\sigma(i)}) \tag{5.10}$$

where $\sigma$ is all the possible correspondences of objects (A,B) that can be obtained from images $Im_A$ and $Im_B$, that is, all the permutations of N.

Note that, depending on which objects of $Im_A$ are compared with which objects of $Im_B$, different values of $SimQID$ are obtained. As in the case of shape similarity, the final correspondences between objects obtained by our approach will be those which maximize the similarity between images:

$$SimQIDFinal(Im_A, Im_B) = max_{\sigma \in C}(SimQID(Im_A, Im_B)) \tag{5.11}$$

where $C$ denotes the set of all possible correspondences between objects, that is, the set of all possible permutations of N ($N!$).

If the two images compared, $Im_A$ and $Im_B$, have a different number of objects, then there are some objects of one image with no corresponding objects in the other image. In this case, the objects with no corresponding pairs in the other image are compared with the *void* object, and the similarity between both objects is zero.

Let us suppose that the number of objects of the images $Im_A$ and $Im_B$ are $N$ and $M$ respectively, and that $N \geq M$. In this case, $N$-$M$ objects of $Im_A$ are compared with the *void* object, and the rest are compared with the objects of $Im_B$ in the same way as in the previous case. Taking into account this situation, the similarity between images is obtained from 5.10 and 5.11 in the same way as before.

Note that only $M$ similarities between objects must be calculated since the similarity between an object of $Im_A$ and the *void* object is always zero.

Finally, the main properties of $SimQIDFinal(Im_A, Im_B)$ are:

- Symmetry: $SimQIDFinal(Im_A, Im_B)) = SimQIDFinal(Im_B, Im_A)$.

- Upper and lower bounds, that is,

$$0 \leq SimQIDFinal(Im_A, Im_B)) \leq 1$$

for any images $Im_A$ and $Im_B$.

In terms of the computational cost ($CC$), as in the shape similarity approach, two situations (the bests and worsts cases) can be distinguished:

1. if both images have the same number of objects ($N$), the cost of the algorithm is $O(N!) \cdot CC(SimObj)$, because the starting object of the comparison can be any object of the second image. If all the objects put into correspondence have the same number of relevant points ($n$), the cost of the algorithm is the best case for this situation $O(N! \cdot n^2)$. However, if all the objects put into correspondence have different number of relevant points, as the starting point for the shape comparison can be anyone of them, the computational cost increases to $O(N! \cdot n^{r+1})$.

2. if the difference in the number of objects between both images is $N - M$, the number of possibilities for choosing $N - M$ objects to be compared with the *void* object is a simple combinatory number $\begin{pmatrix} N \\ N-M \end{pmatrix}$ and considering that the starting object of the comparison can be any object of the image with the highest number of objects, the CC in the worst case is:

$$\begin{pmatrix} N \\ N-M \end{pmatrix} M!$$

Clearly, the computational cost peaks when two images with a high number of objects ($N$) are compared and one of them has twice the number of objects of the other. And also, the objects compared have a high difference in the number of relevant points ($n$) that define their shape. In order to compensate, the *Branch and Bound* technique has been used for speeding the process of finding the correspondences of objects and compensate the $CC$.

### 5.4.2 Applications of $SimQID$: Obtaining the Similarity between Images obtained in Robotic Scenarios

Our approach for comparing Qualitative Image Descriptions (QIDs) has been tested in three scenarios that involve comparisons of:

- images of tile compositions (Section 5.4.2.1),
- images of indoor scenes taken by a robot camera (Section 5.4.2.2), and
- images of landmarks (corners) detected by the laser sensor and previously seen in the robot world (Section 5.4.2.3).

#### 5.4.2.1 Scenario I: Comparing Images of Tile Compositions

Scenario I consists on images of tile compositions that are located on a platform from which a robot arm picks and places tile pieces for assembling tile mosaics as shown in Figure 5.13 (a). The images used in the testing were captured by an industrial camera AVT-Guppy F033C at different times of the day showing different illumination conditions.

In order to analyse the similarity values obtained, the images captured in this scenario have been classified into two groups:

- images containing the same number of objects, shown in Figure 5.18; and
- images containing different number of objects, shown in Figure 5.19.

For extracting the objects of the images, the segmentation method by Canny [1986] is used. The QID approach (Chapter 3) describes all the compared images qualitatively and then similarity values between qualitative image descriptions are obtained applying the formula in (5.11), as explained in Section 5.4.1.

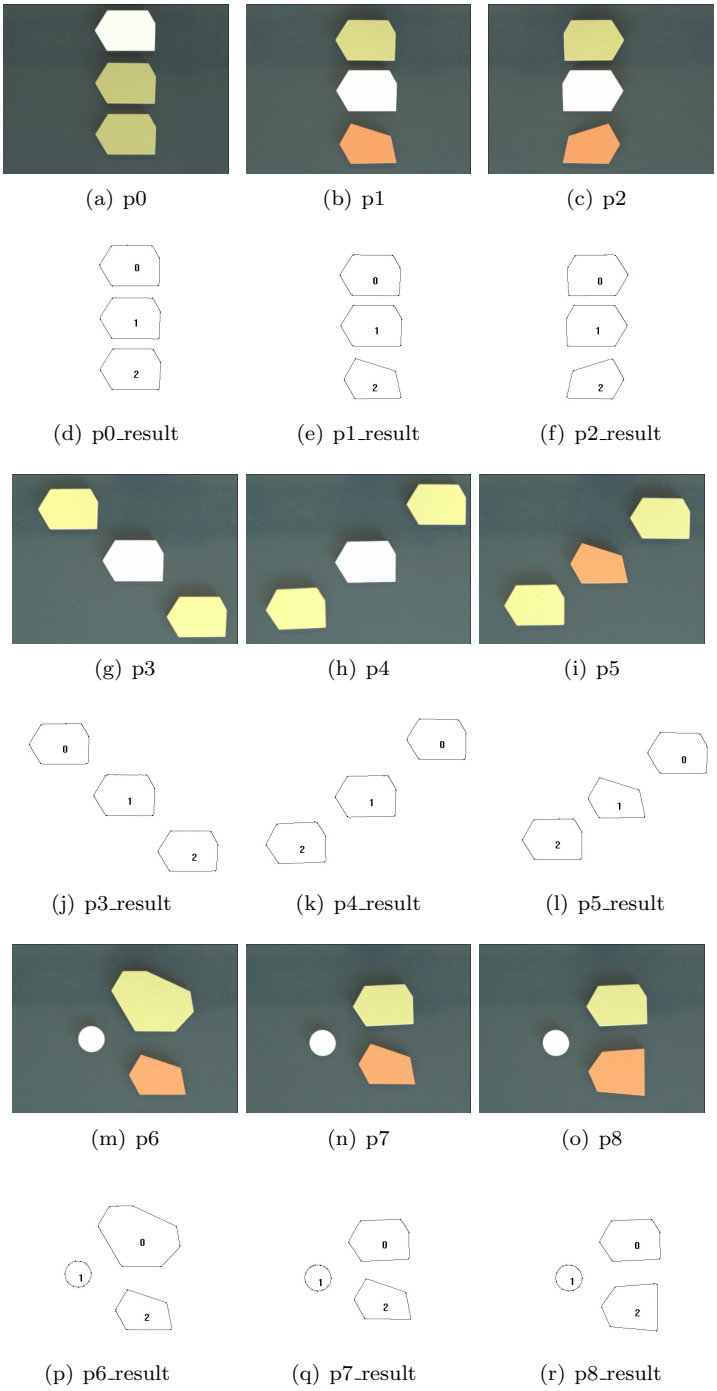
The similarity values obtained after comparing the images containing the same number of objects, shown in Figure 5.18, are given in Table 5.17 where each cell indicates: (i) the $SimQID$ value between the compared images and (ii) the correspondences of objects found. For example, the $SimQID$ between images $p0$ and $p1$ is 0.83 ($SimQID(p0, p1) = 0.83$) and the object correspondences are {1,0,2} that mean the following: object 0 in $p0$ is compared to object 1 in $p1$ (both are the same object), object 1 in $p0$ is compared to object 0 in $p1$ (both are the same object), and object 2 in $p0$ is compared to object 2 in $p1$. The weights used in these comparisons are $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$ in order to give the same importance to all the features described.

The total time of execution of all the similarity calculus in Table 5.17 is 0.91 seconds using a computer with an Intel Core 2 processor at 2.66 GHz and 4 GB of RAM, running under an Ubuntu 10.04 (lucid) with a Linux kernel 2.6.32-21-generic.

After analysing the images presented in Figure 5.18 and the results of similarity presented in Table 5.17, it is important to note that:

- images containing the same objects with mirror changing orientation obtain high similarity values, i.e. $SimQID(p1, p2) = 0.93$, $SimQID(p3, p4) = 0.88$;

Figure 5.18: Comparing images of different tile compositions with the same number of objects.



(a) p0 (b) p1 (c) p2

(d) p0_result (e) p1_result (f) p2_result

(g) p3 (h) p4 (i) p5

(j) p3_result (k) p4_result (l) p5_result

(m) p6 (n) p7 (o) p8

(p) p6_result (q) p7_result (r) p8_result

Table 5.17: Results of $SimQID$ values and correspondences of objects obtained from the images in Figure 5.18 with the $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$.

| | **p1** | **p2** | **p3** | **p4** | **p5** | **p6** | **p7** | **p8** |
|---|---|---|---|---|---|---|---|---|
| **p0** | 0.83 {1,0,2} | 0.78 {1,0,2} | 0.70 {1,0,2} | 0.73 {1,2,0} | 0.80 {2,0,1} | 0.75 {1,0,2} | 0.76 {1,0,2} | 0.77 {1,0,2} |
| **p1** | **1** | 0.93 {0,1,2} | 0.78 {0,2,1} | 0.82 {0,2,1} | 0.87 {0,2,1} | 0.90 {0,1,2} | 0.92 {0,1,2} | 0.91 {0,1,2} |
| **p2** | | **1** | 0.80 {0,2,1} | 0.80 {1,2,0} | 0.83 {0,2,1} | 0.88 {0,1,2} | 0.89 {0,1,2} | 0.88 {0,1,2} |
| **p3** | | | **1** | 0.88 {0,2,1} | 0.74 {2,0,1} | 0.72 {2,1,0} | 0.73 {2,1,0} | 0.74 {2,1,0} |
| **p4** | | | | **1** | 0.78 {1,0,2} | 0.80 {2,0,1} | 0.79 {0,2,1} | 0.80 {2,0,1} |
| **p5** | | | | | **1** | 0.85 {0,2,1} | 0.88 {0,2,1} | 0.87 {0,2,1} |
| **p6** | | | | | | **1** | 0.95 {0,1,2} | 0.97 {0,1,2} |
| **p7** | | | | | | | **1** | 0.97 {0,1,2} |

- images containing similar objects with the same spatial orientation obtain high similarity values, i.e. $SimQID(p6, p7) = 0.95$, $SimQID(p7, p8) = 0.97$;
- the similarity values obtained between these images are all quite high ($SimQID(p_i, p_j) \geq 0.70$) because all the images have the same number of objects (3) and some of these objects have the same/similar colour and/or shape and/or the same/similar spatial distribution.

The similarity values obtained after comparing the other group of images in Figure 5.19 are given in Table 5.18, where as previously explained, each cell indicates the $SimQID$ between the images and the correspondences of objects found. However, in this table, the comparisons with the *void* object are also implicitly given. For example, the $SimQID$ between images $j2$ and $j4$ is 0.59 ($SimQID(j2, j4) = 0.59$) and the object correspondences are {0,1,2} that mean the following: object 0 in $j2$ is compared to object 0 in $j4$, object 1 in $j2$ is compared to object 2 in $j4$ (both objects are white and round and have similar orientations), object 2 in $j2$ is compared to object 1 in $j4$ (both objects are red and have similar shapes) and object 3 in image $j4$ is compared to the *void* object.

The weights used in these comparisons are also $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$ in order to give the same importance to all the features described. And the total time of execution of the similarity calculus in Table 5.18 is 12.3 seconds using a computer with an Intel Core 2 processor at 2.66 GHz and 4 GB of RAM, running under an Ubuntu 10.04 (lucid) with a Linux kernel 2.6.32-21-generic. Note that this execution time is higher than the one obtained before because the images compared have different number of objects and these objects

have different number of relevant points, and therefore, the computational cost is increased, as explained in Section 5.4.1.

Table 5.18: Results of $SimQID$ values and correspondences of objects obtained from the images in Figure 5.19 with the $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$.

| | **j1** | **j2** | **j3** | **j4** | **j5** | **j6** | **j7** | **j8** |
|---|---|---|---|---|---|---|---|---|
| **j0** | 0.78 {0,2,1} | 0.94 {0,1,2} | 0.56 {3,1,0} | 0.58 {3,1,0} | 0.57 {3,2,1} | 0.34 {1,5,2} | 0.30 {2,7,3} | 0.26 {3,8,4} |
| **j1** | **1** | 0.73 {2,1,0} | 0.57 {0,2,1} | 0.60 {1,0,2} | 0.58 {0,2,3} | 0.37 {0,4,5} | 0.32 {3,6,7} | 0.28 {4,7,8} |
| **j2** | | **1** | 0.56 {0,2,3} | 0.59 {0,2,1} | 0.60 {1,2,3} | 0.35 {1,3,2} | 0.30 {2,6,7} | 0.27 {3,7,8} |
| **j3** | | | **1** | 0.92 {1,0,2,3} | 0.84 {1,0,2,3} | 0.42 {2,3,5,4} | 0.38 {0,1,5,3} | 0.34 {0,1,2,4} |
| **j4** | | | | **1** | 0.90 {1,0,2,3} | 0.46 {5,0,3,4} | 0.40 {7,3,5,2} | 0.35 {8,4,6,3} |
| **j5** | | | | | **1** | 0.46 {0,2,3,4} | 0.40 {3,0,5,7} | 0.36 {4,0,6,8} |
| **j6** | | | | | | **1** | 0.81 {3,2,0, 1,6,5,7} | 0.72 {4,3,0, 2,7,6,8} |
| **j7** | | | | | | | **1** | 0.88 {0,1,3,4, 5,6,7,8} |

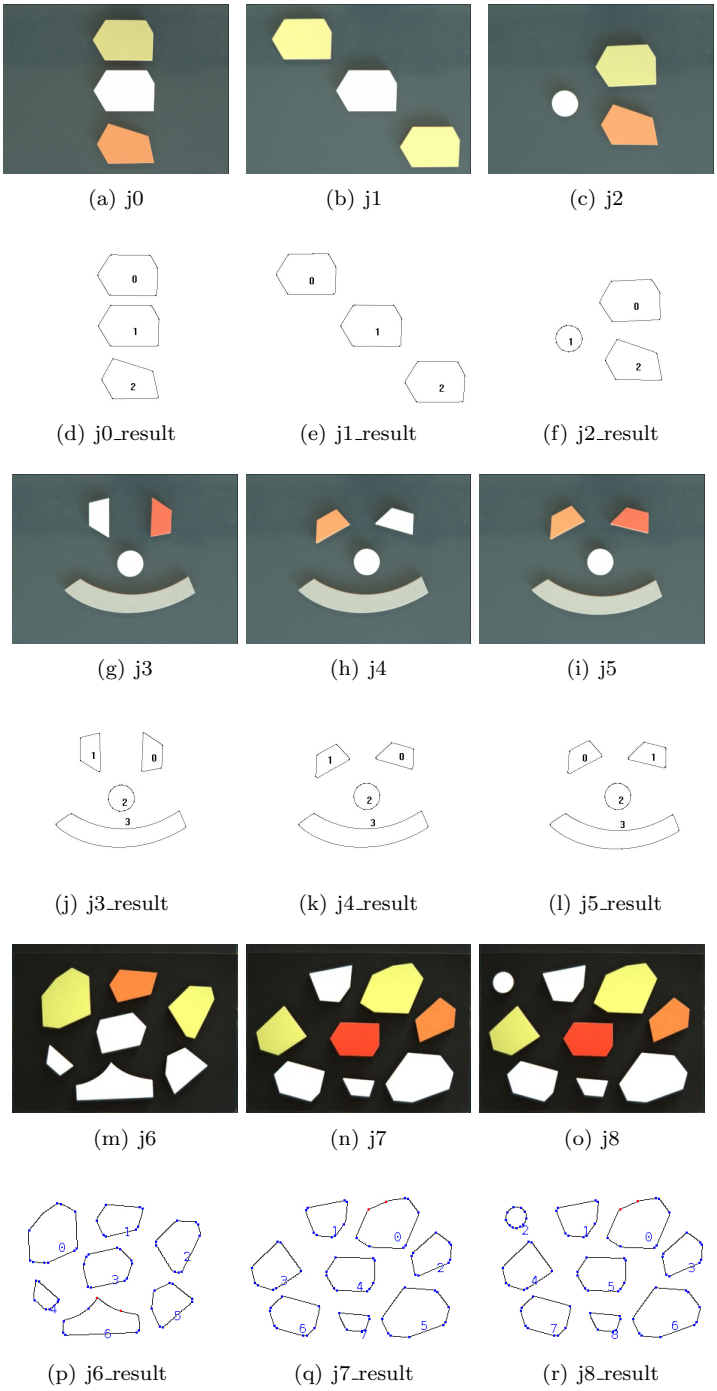
After analysing the images presented in Figure 5.19 and the results of similarity presented in Table 5.18, it is important to note that:

- comparisons of images with the same number of objects obtain high similarity values: $SimQID(j_u, j_v) \geq 0.72$;
- comparisons of images with different number of objects obtain low similarity values: $SimQID(j_u, j_v) \leq 0.60$;
- considering only the images in Figure 5.19 a similarity threshold ($SimTh$) of 0.70 can classify similar images and separate them from the rest;
- the higher the difference of object number between the images, the lower the similarity between the images. For example, the difference in the number of objects between images $j0$ and $j3$ is 1 object, whereas between $j3$ and $j6$ is 3 objects and between $j0$ and $j6$ is 4 objects and the similarity values obtained hold that:

$$SimQID(j0, j3) > SimQID(j3, j6) > SimQID(j0, j6)$$

Finally, $SimQID$ has been tested for image retrieval considering all the images and a similarity threshold (SimTh) of 0.74 and the results obtained are shown in Table 5.19. After analysing the classification obtained, the general conclusions for this scenario are that:

Figure 5.19: Comparing images of different tile compositions with different number of objects.



(a) j0 (b) j1 (c) j2

(d) j0_result (e) j1_result (f) j2_result

(g) j3 (h) j4 (i) j5

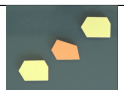
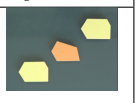
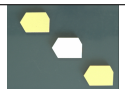
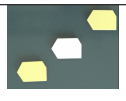
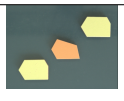
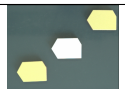
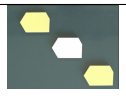
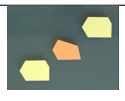
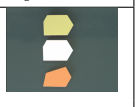
(j) j3_result (k) j4_result (l) j5_result

(m) j6 (n) j7 (o) j8

(p) j6_result (q) j7_result (r) j8_result

- images containing the same objects (same shape and colour) have very high similarity values;
- images with the same number of objects and similar spatial distributions have high similarity values;
- the higher the difference of object number between the images, the lower the similarity between the images;
- the topology feature is the same in all the images (*disjoint*) because the tile pieces usually does not appear touching in the conveyor belt. As future work, more tests can be done in order to study the similarity between images containing separate tile pieces and between images containing tile pieces touching each other building a mosaic;
- as the spatial distribution is not very important in this scenario, then the correspondences of objects obtained can be improved as future work using a heuristic built only using shape and colour similarity instead of considering all the features described by the QID approach.

Table 5.19: Results of testing *SimQID* on tile compositions as a classifier with the $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$ and $SimTh = 0.74$.

| | | | | | |
|---|---|---|---|---|---|
|  p0 |  p1: 0.83 |  p5: 0.80 |  p2: 0.78 |  p8: 0.77 |  p7: 0.76 |
|  p1, j0 |  p2: 0.93 |  p7: 0.92 |  p8: 0.91 |  p6: 0.90 |  p5: 0.87 |
|  p2 |  p1: 0.92 |  p7: 0.89 |  p8: 0.88 |  p2: 0.88 |  p6: 0.83 |
|  p3, j1 |  p4: 0.88 |  p2: 0.80 |  p1: 0.78 |  p5: 0.74 |  p8: 0.74 |
|  p4 |  p3: 0.88 |  p1: 0.82 |  p2: 0.80 |  p8: 0.80 |  p6: 0.80 |
|  p5 |  p7: 0.88 |  p8: 0.87 |  p1: 0.87 |  p6: 0.85 |  p2: 0.83 |
|  p6 |  p8: 0.97 |  p7: 0.95 |  p1: 0.90 |  p2: 0.88 |  p5: 0.85 |
|  p7, j2 |  p8: 0.97 |  p6: 0.95 |  p1: 0.94 |  p2: 0.89 |  p5: 0.88 |
|  p8 |  p7: 0.97 |  p6: 0.97 |  p1: 0.93 |  p2: 0.88 |  p5: 0.87 |
|  j3 |  j4: 0.92 |  j5: 0.84 | | | |
|  j7 |  j8: 0.88 |  j6: 0.81 | | | |

#### 5.4.2.2 Scenario II: Identifying Indoor Scenes in the Robot World

Scenario II consists on the comparison of images captured by a webcam Logitech Quickcam Pro 9000[4] with a Carl Zeiss optic lense and 2 Megapixel resolution located on the top of ActiveMedia Pioneer 2 dx mobile robot (see Figure 5.20).

Images of indoor scenes captured in this scenario are shown in Figure 5.21. For extracting the objects of the images, the segmentation method by Felzenszwalb and Huttenlocher [2004] is used. The QID approach (Chapter 3) describes all the compared images qualitatively and then similarity values between qualitative image descriptions are obtained applying the formula in (5.11), as explained in Section 5.4.1.

Figure 5.20: Robotic platform used to test our approach.

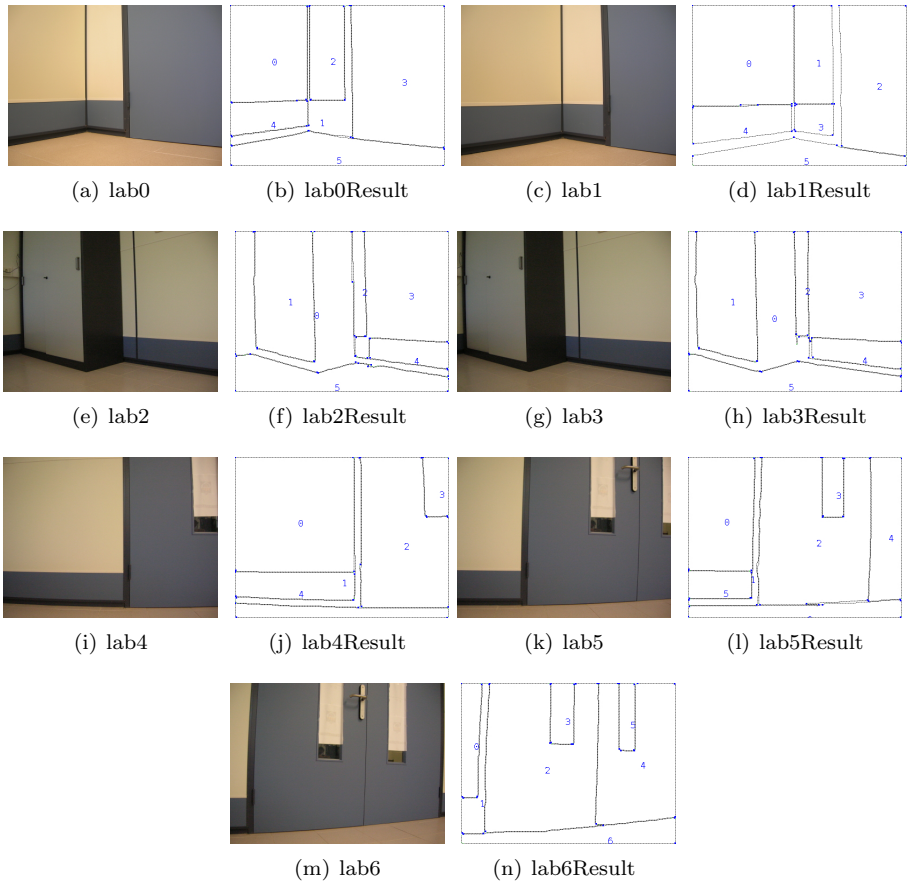
The similarity values obtained after comparing the group of images in Figure 5.21 are given in Table 5.20 where, as previously explained, each cell indicates the $SimQID$ between the images and the correspondences of objects found. For example, the similarity value obtained between images $lab0$ and $lab1$ is 0.84 ($SimQID(lab0, lab1) = 0.84$) and the object correspondences obtained are {0, 3, 1, 2, 4, 5} that mean the following:

- object 0 in $lab0$ is compared to object 0 in $lab1$ (both objects have the same colour, shape and spatial location);
- object 1 in $lab0$ is compared to object 3 in $lab1$ (both objects have the same spatial location);
- object 2 in $lab0$ is compared to object 1 in $lab1$ (both objects have the same colour, shape and spatial location);
- object 3 in $lab0$ is compared to object 2 in $lab1$ (both objects have the same colour, shape and spatial location);
- object 4 in $lab0$ is compared to object 4 in $lab1$ (both objects have the same colour, shape and spatial location);
- object 5 in $lab0$ is compared to object 5 in $lab1$ (both objects have the same colour, shape and spatial location).

Note that there are not correspondences with the *void* object because both images have the same number of objects. Moreover, the weights used in these comparisons are $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$ in order to give the same importance to all the features described. And the total time of execution of the similarity calculus in Table 5.20 is 8.75 seconds using a computer with an Intel Core 2 processor at 2.66 GHz and 4 GB of RAM, running under an Ubuntu 10.04 (lucid) with a Linux kernel 2.6.32-21-generic.

[4]http://www.logitech.com

Figure 5.21: Comparing images of visual landmarks inside our laboratory.



(a) lab0 (b) lab0Result (c) lab1 (d) lab1Result

(e) lab2 (f) lab2Result (g) lab3 (h) lab3Result

(i) lab4 (j) lab4Result (k) lab5 (l) lab5Result

(m) lab6 (n) lab6Result

By considering a Similarity Threshold of 0.84 ($SimTh = 0.84$), the results obtained (corresponding to the $SimQID$ values in bold in Table 5.20) are the following:

- Images similar to *lab0*: *lab1* and *lab4*;
- Images similar to *lab1*: *lab0* and *lab4*;
- Images similar to *lab2*: *lab3*;
- Images similar to *lab3*: *lab2*;
- Images similar to *lab4*: *lab0* and *lab1*;
- Images similar to *lab5*: *lab6*;
- Images similar to *lab6*: *lab5*.

After analysing the images presented in Figure 5.21 and the results of similarity presented in Table 5.20, it is important to note that:

- the similarity values obtained for images *lab0*, *lab1* and *lab4* are so high ($SimQID(lab0, lab4) = 0.84$ and $SimQID(lab1, lab4) = 0.86$) that produce false positives because the QID approach is two-dimensional, that is, it does not consider the depth in the images as human vision does. Therefore, the QID approach cannot differenciate if there is a corner in images

Table 5.20: Results of $SimQID$ values and correspondences of objects obtained from the images in Figure 5.21 with the $w_{QSD} = w_{QCD} = w_{Top} = w_{FO} = 0.25$.

| | **lab1** | **lab2** | **lab3** | **lab4** | **lab5** | **lab6** |
|---|---|---|---|---|---|---|
| **lab0** | **0.84** {0,3,1, 2,4,5} | 0.78 {3,0,4, 2,1,5} | 0.79 {3,0,4, 2,1,5} | **0.84** {0,1,2, 3,4,5} | 0.72 {1,2,4, 3,5,6} | 0.70 {3,2,4, 5,1,6} |
| **lab1** | **1** | 0.83 {3,2,1, 4,0,5} | 0.82 {3,2,1, 4,0,5} | **0.86** {0,4,3, 2,1,5} | 0.75 {0,1,3, 2,5,6} | 0.76 {1,0,5, 4,2,6} |
| **lab2** | | **1** | **0.86** {0,1,2, 3,4,5} | 0.82 {1,0,2, 3,5,4} | 0.72 {0,1,6, 3,4,5} | 0.71 {0,3,6, 5,4,2} |
| **lab3** | | | **1** | 0.80 {1,0,2, 3,5,4} | 0.72 {2,0,1, 3,4,5} | 0.68 {2,0,3, 5,6,1} |
| **lab4** | | | | **1** | 0.77 {0,1,4, 3,5,6} | 0.74 {3,2,4, 5,0,6} |
| **lab5** | | | | | **1** | **0.88** {0,3,2, 5,4,1,6} |

*lab0* and *lab1* and not in image *lab4*. Considering only two-dimensions, image *lab4* can be considered quite similar to images *lab0* and *lab1*.

- the relative size of the objects in the image could be described and, using it, some correspondences of objects (i.e. object 2 in *lab1* to object 3 in *lab4* or object 3 in *lab1* to object 2 in *lab4*) could be improved. The absolute size of the objects in the image (already calculated) cannot be considered if a similarity value independent to the distance to the target is needed.

- the similarity values obtained for images *lab1*, *lab2* and *lab3* are quite high ($SimQID(lab1, lab2) = 0.83$ and $SimQID(lab1, lab3) = 0.82$) but they do not produce false positives because are lower than the similarity threshold defined ($SimTh = 0.84$). These images can be considered symmetric, as the object of interest is on the right in *lab1* while it is on the left in *lab2* and *lab3*: Object 0 of image *lab1* (the wall) is compared to object 3 of image *lab2* (also the wall) and object 2 of image *lab1* (the door) is compared to object 1 of image *lab2* (the door of the cupboard). The same correspondences are obtained for image *lab1* and *lab3*. Therefore, as these images can be considered cognitively symmetric, they are similar according to its qualitative description. As different correspondences of objects are obtained between image *lab0* and images *lab2* and *lab3*, the $SimQID$ values obtained for them are not so high.

Comparing the results obtained in the previous scenario (images of tile compositions) with the results obtained in the current scenario (images of indoor

scenes), it is important to note that the *Similarity Threshold* ($SimTh$) in this scenario must be defined more precisely to avoid false positives. However, sometimes some of them cannot be avoided because our approach is affected by the quantity of relevant objects detected in both compared images which depend on the segmentation process and on the illumination conditions.

In the previous scenario, the relevant objects in the image are differenciated more easily and the segmentation approach by Canny [1986] can obtain the boundary of all the objects without problems because they are contained completely inside the image and the illumination is more controlled. In our current scenario, parts of doors, parts of the floor, etc. appear and their boundary is more difficult to extract, although the results of the approach by Felzenszwalb and Huttenlocher [2004] are quite acceptable.

#### 5.4.2.3 Scenario III: Recognizing Landmarks in the Robot World

For discretizing our problem of identifying scenes of indoor environments, our $SimQID$ approach is tested here for calculating the similarity between previously detected landmarks of the robot world. A landmark is a reference of place used for robot localization. In the approach by Peris and Escrig [2005] (previously explained in Chapter 1 and 2), corners are detected by a laser sensor and considered the landmarks of indoor environments. However, in their approach corners are only described qualitatively as *convex* or *concave* and after its detection and description, more information is needed for its recognition. Figure 5.22 outlines our problem of landmark recognition: a robot navigates through a building (the dotted line represents the route done by the robot), then the robot enters a room and scans it by detecting the corners with its laser sensor and taking photos of them, finally it recognizes a corner and determines inside which room it is at the moment. In this Section, the problem tackled is recognizing a corner by calculating a similarity value between images of corners described qualitatively.

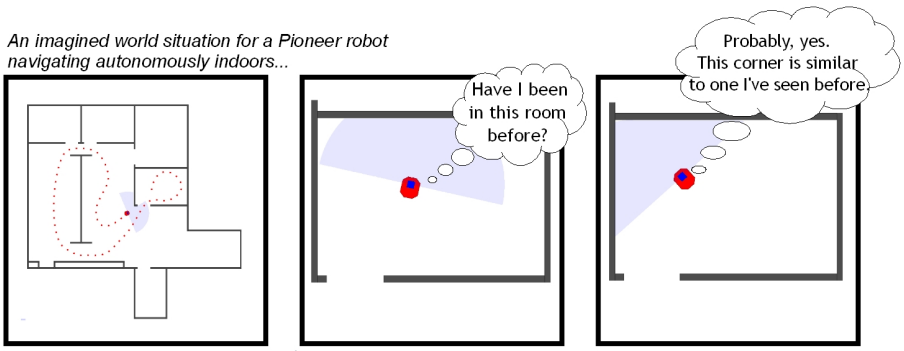



Figure 5.22: Outlining the problem of landmark recognition.

The technical details of the testing done in this scenario are the following. The physical robotic platform used is an ActiveMedia Pioneer 2 dx mobile

robot incorporating eight sonar sensors, a Leuze Rotoscan RS4 laser sensor [5] and a webcam Logitech Quickcam Pro 9000 with a Carl Zeiss optic lense and 2 Megapixel resolution. As Figure 5.20 shows, the laser sensor is mounted on the top of the robot and the webcam is located on it. To carry out the experimentation of our approach, the Player/Stage [6] has been used as the network server for robot control, which provides a simple interface to the robot's sensors and actuators.

Figure 5.24 outlines the steps followed by our robot controller. Once the robot is located inside a room, it: (i) detects the corners of the room with its laser sensor and takes photos of them using its webcam; (ii) describes qualitatively the captured images; and (iii) calculates a similarity value between the qualitative description of the current image and the qualitative description of the images in memory corresponding to previously seen landmarks.

Figure 5.23: Empty office room used as the scenario to test our approach.

The testing scenario consists of an empty office room of our University building shown in Figure 5.23. In a first execution of our controller, the four images of the corners of the room shown in Figure 5.25 are obtained and stored in memory. In the second execution of our controller, the four images obtained of the same corners are obtained and shown in Figure 5.26.

As in Scenario II, the segmentation method by Felzenszwalb and Huttenlocher [2004] is used for extracting the objects of the images (adjusting the segmentation parameters to: $\sigma = 0.6$, $k = 700$ and $min = 1000$). Note that, although perspectives of the corners are quite similar, illumination reflexes produce different segmentation results as shown in Figures 5.25 and 5.26. The pictures obtained are dimensioned to $400 \times 300$ in order to accelerate the segmentation and the process of extracting the boundary of the detected objects, the features extraction time (*FE time*) is given. The QID approach (Chapter 3) describes all the images qualitatively, the QID time of each description is also given.

The similarity values between the QID of the four images obtained in the first

[5] http://www.leuze.com
[6] http://playerstage.sourceforge.net

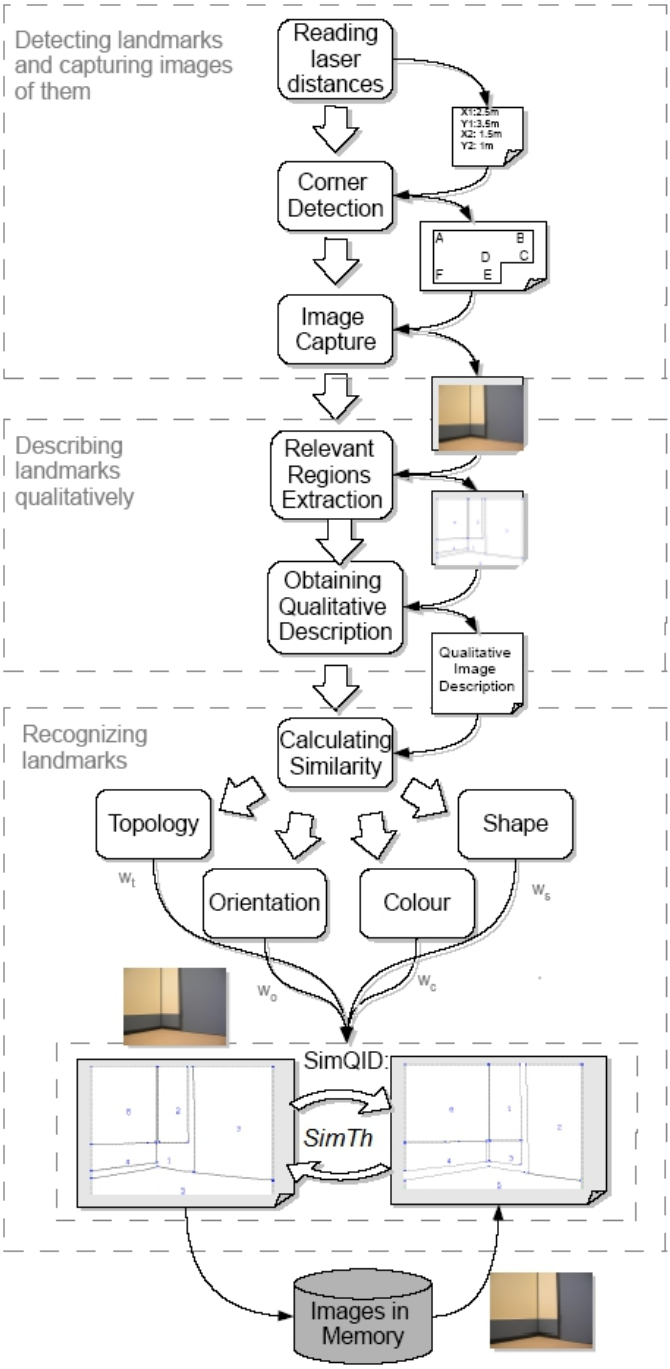
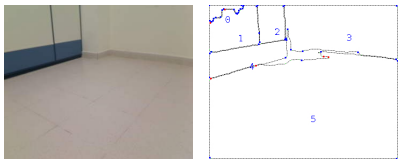



Figure 5.24: Diagram of the controller for recognizing corners as landmarks of indoor environments.

Figure 5.25: Digital images of the four detected corners (*C1, C2, C3, C4*) taken by the robot webcam in its first exploration (*ex1*) of the environment. The result of processing each image by the QID approach and the features extraction time and the QID time are also given.

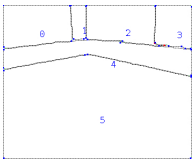


(a) C1_ex1



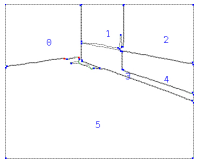
(b) C1_ex1_Result, $FE\ time$ : 0.15, $QID\ time$ : 0.3

(c) C2_ex1



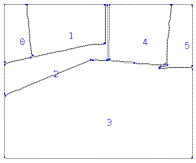
(d) C2_ex1_Result, $FE\ time$ : 0.17, $QID\ time$ : 0.37

(e) C3_ex1



(f) C3_ex1_Result, $FE\ time$ : 0.16, $QID\ time$ : 0.34

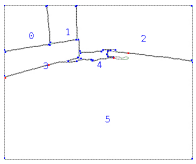
(g) C4_ex1



(h) C4_ex1_Result, $FE\ time$ : 0.16, $QID\ time$ : 0.42

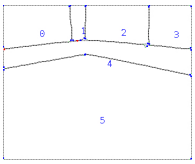
Figure 5.26: Digital images of the four detected corners (*C1, C2, C3, C4*) taken by the robot webcam in its second exploration *ex2* of the environment. The result of processing each image by the QID approach and the features extraction time and the QID time are also given.

(a) C1_ex2



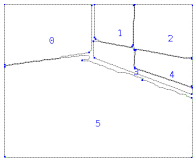
(b) C1_ex2_Result, $FE\ time$ : 0.15, $QID\ time$ : 0.26

(c) C2_ex2



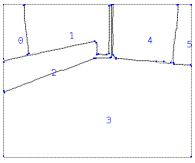
(d) C2_ex2_Result, $FE\ time$ : 0.16, $QID\ time$ : 0.36

(e) C3_ex2



(f) C3_ex2_Result, $FE\ time$ : 0.17, $QID\ time$ : 0.34

(g) C4_ex2



(h) C4_ex2_Result, $FE\ time$ : 0.19, $QID\ time$ : 0.44

execution and each of the images obtained in the second execution is obtained calculating the $SimQIDFinal$ (5.11), as explained in Section 5.4.1. Table 5.21

Table 5.21: Results of comparing the QID of the captured image corresponding to a corner currently detected by the laser sensor with the QID of images of corners previously stored in memory.

| **Current Image** | **Corner 1** | **Corner 2** | **Corner 3** | **Corner 4** |
|---|---|---|---|---|
|  |  |  |  |  |
| **C1_ex2**<br>SimQID time:<br>**49.6** | **C1_ex1**<br>SimQID: **0.95**<br>{1, 2, 3, 4, 0, 5} | C2_ex1<br>SimQID: 0.89<br>{0, 5, 1, 4, 2, 3} | C3_ex1<br>SimQID: 0.85<br>{0, 2, 4, 3, 1, 5} | C4_ex1<br>SimQID: 0.70<br>{0, 1, 6, 2, 3, 5} |
|  |  |  |  |  |
| **C2_ex2**<br>SimQID time:<br>**2.25** | C1_ex1<br>SimQID: 0.84<br>{2, 1, 3, 5, 4, 0} | **C2_ex1**<br>SimQID: **0.95**<br>{0, 1, 2, 3, 4, 5} | C3_ex1<br>SimQID: 0.86<br>{0, 5, 2, 4, 3, 1} | C4_ex1<br>SimQID: 0.70<br>{0, 2, 1, 6, 5, 3} |
|  |  |  |  |  |
| **C3_ex2**<br>SimQID time:<br>**1.96** | C1_ex1<br>SimQID: 0.80<br>{1, 3, 2, 5, 4, 0} | C2_ex1<br>SimQID: 0.86<br>{0, 2, 1, 4, 3, 5} | **C3_ex1**<br>SimQID: **0.93**<br>{5, 1, 2, 3, 4, 0} | C4_ex1<br>SimQID: 0.75<br>{0, 4, 1, 5, 6, 3} |
|  |  |  |  |  |
| **C4_ex2**<br>SimQID time:<br>**4.08** | C1_ex1<br>SimQID: 0.75<br>{0, 2, 1, 5, 3, 4} | C2_ex1<br>SimQID: 0.80<br>{0, 1, 2, 3, 4, 5} | C3_ex1<br>SimQID: 0.84<br>{5, 0, 1, 2, 3, 4} | **C4_ex1**<br>SimQID: **0.90**<br>{0, 3, 2, 1, 4, 5} |

presents the results obtained. Each row shows the results of comparing the QID of the image currently detected by the laser sensor with the QIDs of images previosly stored in memory. The first cell of each row indicates the query image and the time of calculating the similarity between this image and the other four images of corners. The other cells indicate the similarity value obtained and the correspondences of the objects within the compared images, which are numbered in Figures 5.25 and 5.26.

By considering a Similarity Threshold of 0.90 ($SimTh = 0.90$), the results obtained (corresponding to the $SimQID$ values in bold in Table 5.21) show that all the corners are correctly recognized. The similarity approach defined has a high computational cost and it can be very time consuming in the worst case. However, the similarity approach is calculated by a separated thread, and in this way, the robot can continue moving for detecting corners with the laser

sensor and taking photos of them while the recognition is in process. Taking this into account, the execution times corresponding to corners C2, C3 and C4, that are 2-4 seconds approximately are acceptable. However, the execution time obtained for corner C1 (50 seconds approximately) is too high for an efficient robot location and navigation, maybe produced by the irregularity in the shape of region 4 in image *C1_ex2* generated by a light reflex.

We believe that the overall similarity time can be reduced by decreasing the computational cost of the shape similarity calculus. This can be obtained by considering that the robot camera is fixed and that the objects/regions to be compared cannot be rotated and therefore the process of starting the comparison at any of the relevant points of the shape can be avoided.

To further enhance the system precision, the algorithm that assign the correspondence of the objects can be adjusted to the situation by considering only the suitable features (maybe first considering only colour and orientation, etc.) and also the weights applied to each feature can be more affined by experimentation.

#### 5.4.2.4 Discussion: How to improve the SimQID approach

In the QID approach, as the properties of the objects within the images have been abstracted/generalized to intervals of values from the beginning (shape, colour, etc.), a highly precise similarity measure between two images cannot be obtained. However, it is interesting to observe that a high precision is not needed in some situations and that such abstraction has provided the QID approach with a high adaptability to different scenarios.

The *SimQID* approach calculates all the correspondences of the objects within an image description and all the correspondences of the relevant points within a shape description, obtaining an effective rotation invariant image similarity approach, but at a high computational cost that it is very time consuming in the worst case. In order to obtain a more efficient computation, the calculus of the best correspondence of relevant points in the shape similarity could be avoided where the situation does not require similarity rotation invariance, such as Scenarios II and III, in which the camera is fixed on the top of the robot.

Furthermore, the semantic information extracted by the reasoners (as explained in Chapter 4) could also be used for adjusting the correspondence of objects, for example, if an object is categorized as a *door* it could only correspondence to another object categorized also as a *door*.

Finally, other aspects to improve the similarity calculus would be to define and use a similarity measure between the relative orientation (RO) of the objects in the image, to consider the relative size of the objects in the image, to obtain and compare the depth information of the scene from a laser scanner or a Kinnect sensor, etc. However, all these properties have to be selected and organized according to the application and the situation, otherwise using all of them could increase the computational cost without adding more precision.

## 5.5 Conclusions

In this chapter, an approach for obtaining a similarity measure between qualitative image descriptions ($QID$s) is presented. This approach obtains a similarity measure between qualitative shape descriptions ($QSD$), qualitative colour descriptions ($QCD$), and between topology and fixed orientation ($FO$) descriptions of the objects within the images.

A similarity measure between qualitative shape descriptions ($SimQSD$) has been defined using conceptual neighbourhood diagrams and also interval distances and it has been used to compare shapes from the categories Bone, Brick, Glass-Cup, Hammer, Heart and Apple of the MPEG-7 CE-Shape-1 library. The results obtained from the tests are that, according to the definitions by Wilson and Keil [1999] and Palmer [1989], the $SimQSD$ fulfils the requirements of a cognitive perception of shape, because it is invariant to translations, rotations and scaling and also obtains a high similarity value between mirror-image reflections or nearly symmetrical shapes. It also obtains a similarity measure between deformed or incomplete shapes and the approximate location of the deformation or cut is determined by locating the relevant points with *void* correspondence.

A similarity measure between qualitative colour descriptions ($SimQCD$) has been built using conceptual neighbourhood diagrams and has been combined with the $SimQSD$ to define a pragmatic approach for assembling tile mosaics by qualitative shape and colour similarity matching. A remarkable result of this application is that it accelerates the process of mosaic assembling compared to that by Museros and Escrig [2007] and it can produce perceptually visual different mosaics from a given design by just relaxing the colour similarity threshold.

A similarity measure between topology descriptions ($SimTop$) have been defined independently of the specific objects involved in the description and a similarity measure between fixed orientation descriptions ($SimFO$) of the objects within the images have also been defined using conceptual neighbourhood diagrams (CNDs). The $SimTop$ and the $SimFo$ combined with $SimQSD$ and $SimQCD$ have been used to built a similarity measure between qualitative image descriptions ($SimQID$).

For evaluating the $SimQID$ approach, tests has been carried out in three scenarios that involve comparisons of: images of tile compositions (Scenario I), images of indoor scenes taken by a robot camera (Scenario II), and images of landmarks (corners) detected by the laser sensor and previously seen in the robot world (Scenario III).

After analysing the results of image similarity in Scenario I, it is important to note that: (i) images containing the same objects (same shape and colour) have very high similarity values; (ii) images with the same number of objects and similar spatial distributions have high similarity values; (iii) the higher the difference of object number between the images, the lower the similarity between the images. And from the results obtained in this scenario, the main conclusion is that the $SimQID$ approach could be used in applications that involve a human understanding image description, such as image classification and retrieval in databases in general, but specifically it would be really interesting to apply it to the retrieval of vector-drawings, icon or clip-art image search by example or to detect design plagiarisms of tile mosaics.

In Scenario II, the similarity threshold must be defined more precisely than

in Scenario I to avoid false positives. In Scenario I the depth information and the illumination conditions are more controlled. However, the depth information in Scenario II cannot be described by the QID approach and therefore it cannot be taken into account in the similarity process, producing false positives in some situations.

In Scenario III, the depth information for all the images described is the same because they all are photographs of corners. In this scenario, the results show that all the corners are correctly recognized. The landmark recognition process, depending on the $SimQID$ calculus, it can be very time consuming in the worst case. However, the $SimQID$ calculus is computed by a separated thread, and in this way, the robot can continue moving for detecting corners with the laser sensor and taking photos of them while the recognition is in process.

The main conclusion after the tests in Scenario II and III is that, although the $SimQID$ approach is effective for these application, it could be more efficient after accelerating the execution time, tuning the weights by experimentation for giving more importance to the corresponding relevant features of the scenario and using inferred semantic information for obtaining more accurate correspondences of objects.

As future work, we intend to: (i) decrease the computational cost and the execution time of the $SimQID$ in Scenarios II and III; (ii) further test all the proposed similarity measures in a wider dataset; (iii) apply the $SimQID$ to solve other problems such as image retrieval in data bases in general, but specifically retrieval of vector-drawings, icon or clip-art image search by example, and moreover to detect plagiarisms of tile mosaic designs and (iv) compare the results obtained to those of other quantitative or qualitative approaches in the literature in order to identify more advantages or disadvantages of the approaches presented and enhance their performance.

# Chapter 6

# Conclusions

This thesis addresses the issues of qualitative distances extraction and representation and qualitative and semantic description and recognition of images for natural landmark representation and identification in robotics. The contributions presented are described in Chapters 2 to 5 and are summarized in this chapter.

**Chapter 2** presents the definition of a general method for obtaining fuzzy distance patterns from the data obtained by any kind of distance sensors incorporated in a mobile robot and the definition of a dissimilarity factor to compare those fuzzy patterns. Algorithms to integrate distances from sonar and laser distance sensors in order to detect special obstacles such as mirrors and glass windows are also explained. And a method for defuzzifying the fuzzy distances provided for obtaining a smooth robot speed has also been used. Moreover, the fuzzy distance patterns obtained between two corners (considered as indoor landmarks by Peris and Escrig [2005]) are also studied in order to classify pairs of corners into open reference systems (those that determine a space that can include more corners and therefore needs a more accurate exploration) or closed reference systems.

In order to evaluate the general method presented, the specific case of an ActivMedia Pioneer 2 dx mobile robot incorporating eight sonar sensors and a SICK LMS-200 laser range scanner have been considered. The Player/Stage has been used as the robot control interface and the tests have been carried out in two real scenarios that incorporate glass windows at a building of University Jaume I.

The results obtained show that this approach enables the robot to: (1) detect non-working sensors, such as laser sensor disconnection, and avoiding crashing into obstacles when lacking the information provided by them, (2) detect mirrors and glass windows as obstacles; (3) obtain the real distance to corners, since the sonar sound reflections are detected; (4) approach obstacles at a smooth speed and avoid crashing into them; and (5) properly classify reference systems ($RS$s) as *open* or *closed*, including those that incorporate glass surfaces.

To introduce visual information in the description and recognition of natural indoor landmarks, in **Chapter 3** a qualitative image description (QID) approach has been defined taking into account studies on how people describe images. The QID approach can describe any digital image by extracting the

visual and spatial features of all the characteristic regions/objects within an image. New qualitative models for shape description (QSD) and for colour description (QCD) have been defined for representing the visual features of the relevant regions within any image, whereas the qualitative model of topology by Egenhofer and Franzosa [1991] and the qualitative models of orientation by Hernández [1991] and by Freksa [1992] have been used for describing the spatial features of the relevant regions within any image.

Regarding the validation of the QID approach, the results show that it is useful for describing images obtained in two real robot-working scenarios: images of indoor corridors captured by a webcam Logitech Quickcam Pro 9000 located on the top of an ActiveMedia Pioneer 2 dx mobile robot (Scenario I), and the description of mosaic images captured by an industrial camera AVT-Guppy F033C located on a platform from which a robot arm picks and places tile pieces to assemble mosaics automatically (Scenario II).

According to the results obtained, the QID approach is proved to be independent of the image segmentation method used. Therefore, the most convenient approach from the literature can be selected depending on the application. Successful results have been obtained using a boundary-based segmentation method [Canny, 1986] in Scenario I and a region-based segmentation method [Felzenszwalb and Huttenlocher, 2004] in Scenario II. The QID execution times obtained in the tests are acceptable which depend on the number of regions or objects extracted from the image and the biggest number of relevant points that define a region in that image.

In order to provide a formal and explicit meaning to the qualitative description of images generated by the QID approach, a Description Logic (DL) based ontology, QImageOntology, has been designed and presented in **Chapter 4**. The qualitative concepts described by the QID approach and their relations are stored as instances of an ontology and contextualized descriptions that characterize kinds of objects are defined in the ontology schema. Although, a contextualized ontology definition for every possible object detected in an image is not provided, the approach presented has the advantage of automatically processing any random image and obtaining a set of DL-axioms that describe it visually and spatially.

In order to evaluate QImageOntology, tests have been carried out using digital images captured by a webcam Logitech Quickcam Pro 9000 located on the top of an ActiveMedia Pioneer 2 dx mobile robot which correspond to the corridors of a building at University Jaume I. The ontology facts obtained from each image, together with the contextualized descriptions defined, have been automatically classified using the ontology reasoner HermiT, although another reasoner (e.g. FaCT++ or Pellet) could have been used. Moreover, the approach presented proposes a method for restricting/closing the world for each particular image using constructors within the *image facts layer* of QImageOntology. The proper classification of the ontology facts obtained in accordance with QImageOntology has been checked using Protégé as front-end.

Finally, the results obtained show that our approach can characterize regions of any image as *walls*, *floor*, *office doors* and *fire extinguishers*, under different illumination conditions and from different observer viewpoints. This 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.

In order to increase the utility of the qualitative image descriptions (QIDs) obtained, an approach for measuring the similarity between two QIDs ($SimQID$) has been developed and presented in **Chapter 5**. This approach obtains a similarity measure between qualitative shape descriptions ($QSD$), qualitative colour descriptions ($QCD$), and between topology and fixed orientation ($FO$) descriptions of the objects within the images.

A similarity measure between qualitative shape descriptions ($SimQSD$) has been defined using conceptual neighbourhood diagrams (CNDs) and also interval distances. It has been evaluated on images of different categories of the MPEG-7 CE-Shape-1 library. The results obtained from the tests are that, according to the definitions by Wilson and Keil [1999] and Palmer [1989], the $SimQSD$ fulfills the requirements of a cognitive perception of shape, because it is invariant to rotations, translations, scaling and mirror-image changes of shapes and also combinations of these. It also obtains a similarity measure between deformed or incomplete shapes and the approximate location of the deformation or cut is determined by locating the relevant points with *void* correspondence.

A similarity measure between qualitative colour descriptions ($SimQCD$) has been built using conceptual neighbourhood diagrams (CNDs) and has been combined with the $SimQSD$ to define a pragmatic approach for assembling tile mosaics by qualitative shape and colour similarity matching. A remarkable result of this application is that it accelerates the process of mosaic assembling compared to that by Museros and Escrig [2007] and it can produce perceptually visual different mosaics from a given design by just relaxing the colour similarity threshold.

A similarity measure between topology descriptions ($SimTop$) have been defined independently of the specific objects involved in the description and a similarity measure between fixed orientation descriptions ($SimFO$) of the objects within the images have also been defined using conceptual neighbourhood diagrams (CNDs). The $SimTop$ and the $SimFO$ combined with $SimQSD$ and $SimQCD$ have been used to built a similarity measure between qualitative image descriptions ($SimQID$).

For evaluating the $SimQID$ approach, tests has been carried out in three scenarios that involve comparisons of: images of tile compositions (Scenario I), images of indoor scenes taken by a robot camera (Scenario II), and images of landmarks (corners) detected by the laser sensor and previously seen in the robot world (Scenario III).

Scenario I is a real human and robot-working place where images of tile compositions are taken by an industrial camera AVT-Guppy F033C at different times of the day showing different illumination conditions. The tile pieces are located on a conveyor belt from which a robot arm picks and places tile pieces for assembling tile mosaics. After analysing the results of the $SimQID$ approach on images taken in this scenario, it is important to note that: (i) images containing the same objects (same shape and colour) have very high similarity values; (ii) images with the same number of objects and similar spatial distributions have high similarity values; (iii) the highest the difference of object number between the images, the lower the similarity between the images. And from the results obtained in this scenario, the main conclusion is that the $SimQID$

approach could be used in applications that involve a human understanding image description, such as image classification and retrieval in databases in general, but specifically it would be really interesting to apply it to the retrieval of vector-drawings, icon or clip-art image search by example or to detect design plagiarisms of tile mosaics.

Scenario II is determined by the images taken by a webcam Logitech Quickcam Pro 9000 located on the top of an ActiveMedia Pioneer 2 dx mobile robot while the robot is navigating through the corridors of a building at University Jaume I. After the evaluation of $SimQID$ in the current scenario, the results obtained show that, the similarity threshold must be defined more precisely than in Scenario I to avoid false positives. In Scenario I the depth information is always the same for all the images and the illumination conditions are more controlled. However, the depth information in the current scenario (different for each image) cannot be described by the QID approach and therefore it cannot be taken into account in the similarity process, producing false positives in images that could be considered similar if the depth information would not be considered.

Scenario III is determined by the images of natural landmarks, which are corners of indoor rooms for the location and navigation context of our robot, that are detected by a Leuze Rotoscan RS4 laser sensor located on the top of an ActiveMedia Pioneer 2 dx mobile robot incorporating eight sonar sensors and a webcam Logitech Quickcam Pro 9000. A controller has been designed using the Player/Stage as the robot control interface. This controller allows the robot to detect the landmarks/corners of the room, take a photograph of them, obtain its corresponding QID and compare it using the $SimQID$ to the QIDs of the images previously captured and stored in memory. In this scenario, the depth information for all the images described is the same because they are all photographs of corners. After the evaluation of the controller designed, the results show that all the corners are correctly recognized. The landmark recognition process, depending on the $SimQID$ calculus, it can be very time consuming in the worst case. However, the $SimQID$ calculus is computed by a separated thread, and in this way, the robot can continue moving for detecting corners with the laser sensor and taking photos of them while the recognition is in process.

The $SimQID$ approach calculates all the correspondences of the objects within an image description and all the correspondences of the relevant points within a shape description obtaining an effective rotation invariant image similarity approach, but at a high computational cost when the difference in the number of objects or in the number of relevant points is high. In order to accelerate the $SimQID$ calculus in the future, the process of obtaining the best correspondence of relevant points of a shape could be avoided where the situation does not require similarity rotation invariance. In Scenarios II and III the robot camera is fixed and the objects/regions to be compared usually will not have rotated shapes, therefore the process of starting the comparison at any of the relevant points of the shape can be avoided.

After the tests in Scenario II and III, the $SimQID$ approach shows to be effective for these applications. However, it could be more efficient after accelerating the execution time, tuning the weights by experimentation for giving more importance to the corresponding relevant features of the scenario and using inferred semantic information for obtaining more accurate correspondences of

objects.

Finally, the main conclusion after the evaluation of the *SimQID* approach is that, as the properties of the objects within the images have been abstracted or generalized to intervals of values from the beginning (shape, colour, etc.), a highly precise similarity measure between two images cannot be obtained. However, it is interesting to observe that a high precision is not needed in some situations and that such abstraction has provided the QID approach with a high adaptability to different scenarios.

## Future Work

There are many lines in which the work done in this thesis can be continued, specially the part regarding the qualitative and semantic interpretation and recognition of images.

In general, all the qualitative and semantic information regarding distances and visual information that are extracted by the approaches described in this thesis should be stored in a knowledge data base that the robot could use in localization and navigation processes and other tasks that could involve interpreting its surroundings in the future.

The Qualitative Image Description approach can describe any image containing objects that are unknown by the robot. It would be interesting combining it with feature invariant detectors, such as SIFT, SURF, Harris-Affine, Hessian-Affine, etc. (see the work by Mikolajczyk et al. [2005] or by Ramisa [2009] for an overview of all these methods) in order to detect known objects, that is, objects segmented by experts and provided to the robot before the object description and identification process. The main disadvantage of these methods is that a repository of all the possible images of objects existing in the world is still not a reality. However, by combining these detectors with the QID-approach, the interpretation of the environment would be enhanced because the known objects would be detected but also all the unknown objects/regions in the images would also be described qualitatively without having previous information of them.

Moreover, the qualitative image descriptions (QIDs) obtained can be easily translated into a natural language form, so that they could be understood and interpreted by human-users. In this way, the user-machine communication in many applications could be enhanced. For example, the QID could be easily post-processed to produce a narrative written description of any image that could be included in a user-interface or read aloud by a speech synthesizer application for blind users to know what the image shows. Once this is obtained, the descriptions provided by the QID-approach could be compared with those produced by human-users in order to study how to improve the cognitive perspective of the QID-approach.

Finally, the QID-approach can be extended to the 3D QID-approach by incorporating the depth information of the images, which is easily obtained now using new devices such as a Kinnect sensor. In this case, the topological model of the scene would change because overlapping objects would be allowed and qualitative models for visibility [Fogliaroni et al., 2009; Tassoni et al., 2011] could be applied for reasoning and extracting more information from the environment.

The QImageOntology can be integrated with other domain ontologies (e.g.,

DOLCE [Gangemi et al., 2002]) and standards such as MPEG-7 [Hunter, 2006] and extended in order to characterize other objects from the robot context. Moreover, other methods for restricting/closing the world for each particular image can be tested in order to accelerate the inference process.

Finally, the reasoner could be integrated into the robot system, so that the new knowledge obtained can be provided to the robot in real time without using the Protégé front-end. And as future applications in robotics, our approach could be usefully applied for *general* robot localization purposes by extending QImageOntology for characterizing objects in different scenarios (e.g. laboratories/classrooms/libraries or outdoor areas) and determining the kind of scenario the robot is navigating through by the classification obtained.

The $SimQID$ approach can be easily applied to image retrieval in data bases in general, but specifically promising results are expected from the retrieval of vector-drawings or icon or clip-art images which requires less effort in image segmentation and the extraction of the relevant regions/objects in the image. And it could be also applied to detect plagiarisms of tile mosaic designs.

It is also important to study how to decrease the computational cost and the execution time of the $SimQID$ in the worst case without losing cognitive properties of the similarity calculus. Moreover, the weights applied in the $SimQID$ approach have to be tuned for giving more importance to the different relevant features of each scenario. And the correspondence of the objects obtained by the $SimQID$ approach can be improved by selecting the more representative features for each scenario and also by using the semantic information extracted from the QImageOntology.

Furthermore, the shape, colour and image similarity approaches proposed in this thesis ($SimQSD$, $SimQCD$, $SimQID$, respectively) need further testing in larger datasets for obtaining a more exhaustive evaluation of its performance so that we can compare the results obtained to those of other quantitative or qualitative approaches in the literature.

Finally, other aspects to improve the $SimQID$ approach would be to define and use a similarity measure between the relative orientation (RO) of the objects in the image, to consider the relative size of the objects in the image, to obtain the depth information of the scene from a laser scanner or a Kinnect sensor and incorporate it in the similarity calculus, etc. However, all these properties have to be selected and organized according to the application and the situation, otherwise using all of them could increase the computational cost without adding more precision.

# Appendix A

# Publications of the Author related to this PhD Thesis

- International Journals with Peer-review:

  1. Falomir Z., Jiménez-Ruiz E., Escrig M. T., Museros L. (2011). Describing Images using Qualitative Models and Description Logics, *Spatial Cognition and Computation*, 11(1):45-74. Submitted: 30/11/2009, accepted: 2/12/2010, appeared: 4/3/2011.
     DOI: 10.1080/13875868.2010.545611.
  2. Falomir Z., Castelló V., Escrig M. T., Peris J. C. (2011). Fuzzy Distance Sensor Data Integration and Interpretation, the *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* (*IJUFKS*), vol. 19 (3). Submitted: 1/6/2010, accepted: 5/12/2010, appeared: 13/05/2011. DOI: 10.1142/S0218488511007106.
  3. Falomir Z., Escrig M. T., Museros L., Gonzalez-Abril L., Ortega J. A., A Model for Qualitative Description of Images based on Visual and Spatial Features, *Journal of Computer Vision and Image Understanding*, submitted: 27/10/10, second revision sent: 05/05/2011.
  4. Falomir Z., Gonzalez-Abril L., Museros L., Ortega J., Measures of Similarity between Objects from a Qualitative Shape Description, *Spatial Cognition and Computation*, submitted: 18/2/2011.

- Book Chapters:

  1. Falomir Z., Martí I., Viana W., Museros L., Escrig M. T. (2010). A Pragmatic Approach for Qualitative Shape and Qualitative Colour Similarity Matching. In René Alquézar, Antonio Moreno and Josep Aguilar (Eds.), *Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*, IOS Press, Vol. 220, pp. 281-290, ISBN 978-1-60750-642-3, ISSN 0922-6389.
  2. Falomir Z., Almazán J., Grande J., Museros L., Escrig M. T. (2009). A Similarity Calculus for Comparing Qualitative Shape Descripitons. In Sandra Sandri, Miquel Sànchez-Marrè and Ulises Cortés (Eds.),

*Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*, IOS Press, Vol. 202, pp. 318-326, ISBN 978-1-60750-061-2, ISSN 0922-6389.

3. Falomir Z., Escrig M. T., Peris J. C., Castelló V. (2007). Distance Sensor Data Integration and Prediction. In Cecilio Angulo and Lluís Godo (Eds.), *Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*, IOS Press. Vol. 163, pp. 339-348, ISBN 978-1-58603-798-7, ISSN 0922-6389.
4. Falomir Z., Peris J. C., Escrig M. T. (2006). Building a Local Hibrid Map from Sensor Data Fusion. In M. Polit, T. Talbert, B. López and J. Meléndez (Eds.), *Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*, IOS Press. Vol. 146, pp. 203-210, ISBN 1-58603-663-7.
5. Falomir Z., Escrig M. T. (2004). Qualitative multi-sensor data fusion. In Jordi Vitriá, Petia Radeva, Isabel Aguiló (Eds.), *Recent Advances in Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*, IOS Press. Vol. 113, pp. 259-266, ISBN 1-58603-466-9.

- International Conferences with Peer-review:

1. Falomir Z., Museros L., Ortega J. A., Velasco F. (2011). A Model for Qualitative Colour Description and Comparison. *25th International Workshop on Qualitative Reasoning (QR'2011),* co-located at the *Joint International Conference on Artificial Intelligence (IJCAI'2011),* Barcelona, Spain.
2. Falomir Z., Jiménez-Ruiz E., Museros L., Escrig M. T. (2009). An Ontology for Qualitative Description of Images, *Workshop on Spatial and Temporal Reasoning for Ambient Intelligence Systems, Conference on Spatial Information Theory (COSIT'09).* http://www.sfbtr8.spatial-cognition.de/papers/ COSIT-ST-AmI-2009-Proc.pdf
3. Falomir Z., Escrig M. T. (2008). Qualitative Models of Shape, Size, Orientation and Distance Applied to the Description of Images Containing 2D Objects, *Proceedings of the Qualitative Reasoning Workshop (QR'08)*, Colorado, USA.
4. Falomir Z., Escrig M. T. (2008). Integration of Qualitative Visual Information and Qualitative Distances applied to Robotics. *VI International Conference on Spatial Cognition, Doctoral Colloquium*, Freiburg, Germany.
5. Falomir Z., Almazán J., Museros L., Escrig M. T. (2008). Describing 2D Objects by using Qualitative Models of Color and Shape at a Fine Level of Granularity, *Proceedings of the Spatial and Temporal Reasoning Workshop at the 23rd AAAI Conference on Artificial Intelligence*, ISBN: 978-1-57735-379-9, Chicago, Illinois, USA.
6. Falomir Z., Escrig M. T. (2004). A hybrid model for multi-sensor data fusion, *Spatial and Temporal Reasoning Workshop, European Conference on Artificial Intelligence (ECAI)*, pp. 3-13.

- National Conferences with Peer-review:

  1. Falomir Z., Viana W., Gonzalez-Abril L., Escrig M. T. (2010). A New Approach for Comparing Qualitative Description of Images. In J. A. Ortega, C. Angulo and L. Gonzalez-Abril (Eds), *XII Jornadas de la Asociación Española de Razonamiento Cualitativo y Apliaciones (JARCA'10). Eficiencia Energética y Sostenibilidad en Inteligencia Ambiental*, pp. 69-82, Universidad de Sevilla, ISBN: 978-84-614-6457-9.
  2. Falomir Z., Martí I., Museros L., Gonzalez-Abril L., Velasco F., Escrig M. T. (2010). A New Approach for Qualitative Colour Naming and Comparing. In J. A. Ortega, C. Angulo and L. Gonzalez-Abril (Eds), *XII Jornadas de la Asociación Española de Razonamiento Cualitativo y Apliaciones (JARCA'10). Eficiencia Energética y Sostenibilidad en Inteligencia Ambiental*, pp. 59-68, Universidad de Sevilla, ISBN: 978-84-614-6457-9, Mallorca, Spain.
  3. Falomir Z., Escrig M. T., Studying how to Compare Qualitative Descriptions of Images (2009). In L. Gonzalez-Abril, F. Velasco y J. A. Ortega (Eds), *XI Jornadas de la Asociación Española de Razonamiento Cualitativo y Apliaciones (JARCA'09). Sistemas Cualitativos, Diagnosis, Robótica, Sistemas Domóticos y Computación Ubicua*, pp. 39-44, Universidad de Sevilla, ISBN: 978-84-613-71-587, Mallorca, Spain.
  4. Falomir Z., Escrig M. T., Museros L., Almazán J. (2008). An Approach for Qualitative Description of Images Containing 2D Objects, *X Jornadas de la Asociación Española de Razonamiento Cualitativo y Apliaciones (JARCA'08). Sistemas Cualitativos y Diagnosis, Robótica, Sistemas Domóticos y Computación Ubicua*, ISBN 978-84-89315-54-9, Tenerife, Spain.
  5. Falomir Z., Escrig M. T., Peris J. C., Castell V. (2007). Sensor Data Integration for a Qualitative and Robust Interpretation of the Robot Environment. In J. Vehí, J. Armengol, J. A. Ortega, *IX Jornadas de la Asociación Española de Razonamiento Cualitativo y Aplicaciones (JARCA'07), Sistemas Cualitativos y Diagnosis*, pp. 33-39, ISBN 978-84-8458-231-1, Girona, Spain.
  6. Falomir Z., Escrig M. T., A Fuzzy Qualitative Approach to Laser and Sonar Data Fusion (2006). In M. T. Escrig, L. Museros, J. A. Ortega, *VIII Jornadas de la Asociación Española de Razonamiento Cualitativo y Aplicaciones (JARCA'06), Razonamiento Cualitativo y Aplicaciones. Robótica, Economía, Diagnosis y Clasificación*, pp. 59-64, ISBN 84-611-1401-9, Castellón, Spain.

# Appendix B

# Other Relevant Publications of the Author

- International Journals with Peer-review:

  1. Museros L., Gonzalez-Abril L., Velasco F., Falomir Z. (2010). A Pragmatic Qualitative Approach for Juxtaposing Shapes, *Journal of Universal Computer Science (J.UCS)*, 16(11): 1410-1424, submitted: 5/1/10, accepted: 28/5/09, appeared: 1/6/10.
  2. Museros L., Falomir Z., Velasco F., Gonzalez-Abril L., Isabel Martí (2011). 2D Qualitative Shape Matching applied to Ceramic Mosaics Assembling, *Journal of Intelligent Manufacturing*, 22: 1-11, DOI 10.1007/s10845-011-0524-6, submitted: 19/7/2010, appeared: 3/3/2011. DOI: 10.1007/s10845-011-0524-6.

- National Journals:

  1. Falomir Z., Neuro-Robótica. In *I+S Informática y Salud*, 77: 59-77, ISSN 1579-8070, December 2009.
  2. Coltell O., Arregui MA., Falomir Z., Arregui M., Puig A., La Bioinformática Clínica: aplicación en el ámbito de la sanidad. In *Todo Hospital*, 218: 426-435, ISSN 0212-19721, July-August 2005.

# Bibliography


Allen, J. (1981). An Interval-Based Representation of Temporal Knowledge. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pages 221–226.

Attalla, E. and Siy, P. (2005). Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recognition*, 38(12):2229–2241.

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.

Bai, X., Yang, X., and Latecki, L. J. (2008). Detection and recognition of contour parts based on shape similarity. *Pattern Recogn.*, 41(7):2189–2199.

Berk, T., Brownston, L., and Kaufman, A. (1982). A new color-naming system for graphics languages. *IEEE Computer Graphics and Applications*, 2:37–44.

Berretti, S., Bimbo, A. D., and Pala, P. (2000). Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transactions on Multimedia*, 2(4):225–239.

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*, 24(3).

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, 32:29–73.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147.

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. et al. (1997). International journal of human-computer studies. *IEEE Intelligent Systems Journal*, 46(2-3).

Brady, M. (1983). Criteria for shape representations. In *Human and Machine Vision*. Academic Press, New York.

Brisson, E. (1993). Representing geometric structures in d dimensions: Topology and order. *Discrete & Computational Geometry*, 9:387–426.

Bruns, H. T. and Egenhofer, M. J. (1996). Similarity of spatial scenes. In *7 th Symposium on Spatial Data Handling*, pages 31–42.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 8:679–697.

Clementini, E. and Di Felice, P. (1993). An object calculus for geographic databases. In *SAC '93: Proceedings of the 1993 ACM/SIGAPP symposium on Applied computing*, pages 302–308, New York, NY, USA. ACM.

Clementini, E. and Di Felice, P. (1997). A global framework for qualitative shape description. *Geoinformatica*, 1(1):11–27.

Clementini, E. and Felice, P. D. (1997). Approximate topological relations. *Int. J. Approx. Reasoning*, 16(2):173–204.

Clementini, E., Felice, P. D., and van Oosterom, P. (1993). A small set of formal topological relationships suitable for end-user interaction. In *Proc. of the 3rd Int'l. Symposium on Advances in Spatial Databases*, pages 277–295.

Cohn, A. (1995). A hierarchical representation of qualitative shape based on connection and convexity. In *Proc COSIT95, LNCS*, pages 311–326. Springer Verlag.

Cohn, A. and Gotts, N. (1995). The 'egg-yolk' representation of regions with indeterminate boundaries. In *in P. Burrough and A. M. Frank (eds), Proceedings, GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries*, pages 171–187. Francis Taylor.

Cohn, A., Randell, D., Cui, Z., Bennett, O., and Gooday, J. (1994). Taxonomies of logically defined qualitative spatial relations. In *in N. Guarino and R. Poli (eds), Formal Ontology in Conceptual Analysis and Knowledge Representation*, pages 831–846. Kluwer.

Cohn, A. G., Bennett, B., Gooday, J. M., and Gotts, N. (1997). Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1:275–316.

Conway, D. (1992). An experimental comparison of three natural language colour naming models. In *Proc. East-West International Conference on Human-Computer Interactions*, pages 328–339.

Corridoni, J. M., Del Bimbo, A., and Vicario, E. (1998). Image retrieval by color semantics with incomplete knowledge. *Journal of the American Society for Information Science*, 49(3):267–282.

Courant, R. and Robbins, H. (1996). *What Is Mathematics? An Elementary Approach to Ideas and Methods*. Oxford University Press.

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.

Damski, J. C. and Gero, J. S. (1996). A logic-based framework for shape representation. *Computer-aided Design*, 28(3):169–181.

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.

Deruyver, A. and Hod, Y. (2009). Qualitative spatial relationships for image interpretation by using a conceptual graph. *Image and Vision Computing*, 27:876–886.

Diosi, A. and Kleeman, L. (2004). Advanced sonar and laser range finder fusion for simultaneous localization and mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, volume 2, pages 1854–1859.

Diosi, A., Taylor, G., and Kleeman, L. (2005). Interactive SLAM using laser and advanced sonar. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1103–1108.

Düntsch, I., Wang, H., and McCloskey, S. (2001). A relation — algebraic approach to the region connection calculus. *Theoretical Computer Science*, 255(1-2):63–83.

Dylla, F., Frommberger, L., Wallgrün, J. O., Wolter, D., Nebel, B., and Wölfl, S. (2007). SailAway: Formalizing navigation rules. In *Proceedings of the AISB'07 Artificial and Ambient Intelligence Symposium on Spatial Reasoning and Communication*.

Egenhofer, M. (1989). A formal definition of binary topological relationships. In *Foundations of Data Organization and Algorithms, 3rd International Conference, FODO 1989*, volume 367 of *Lecture Notes in Computer Science*, pages 457–472. Springer-Verlag.

Egenhofer, M. and Herring, J. (1991). *Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases*. Department of Surveying Engineering, University of Maine, Orono, ME.

Egenhofer, M. J. (2005). Spherical topological relations. *Journal on Data Semantics III*, 3534:25–49.

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., Clementini, E., and Felice, P. D. (1994). Topological relations between regions with holes. *Int. Journal of Geographical Information Systems*, 8:129–142.

Egenhofer, M. J. and Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174.

Egenhofer, M. J. and Vasardani, M. (2007). Spatial reasoning with a hole. In Winter, S., Duckham, M., Kulik, L., and Kuipers, B., editors, *Spatial Information Theory, 8th International Conference, COSIT 2007, Melbourne, Australia, September 19-23, 2007, Proceedings*, volume 4736 of *Lecture Notes in Computer Science*, pages 303–320. Springer.

Escrig, M. T. and Toledo, F. (2000). Autonomous robot navigation using human spatial concepts. *Intl. Journal on Intelligent Systems*, 15(3):165–196.

Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.

Flynn, P. J. and Jain, A. K. (1991). CAD-based computer vision: From CAD models to relational graphs. *IEEE Transaction on P.A.M.I.*, 13(2):114–132.

Fogliaroni, P., Wallgrün, J. O., Clementini, E., Tarquini, F., and Wolter, D. (2009). A qualitative approach to localization and navigation based on visibility information. In Hornsby, K. S., Claramunt, C., Denis, M., and Ligozat, G., editors, *COSIT*, volume 5756 of *Lecture Notes in Computer Science*, pages 312–329. Springer.

Frank, A. U. (1991). Qualitative spatial reasoning with cardinal directions. In Kaindl, H., editor, *7. Österreichische Artificial Intelligence Tagung*, Berlin. Springer.

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. Intl. Conf. GIS—From Space to Territory*, volume 639 of *Lecture Notes in Computer Science*, pages 162–178, Berlin. Springer.

Frommberger, L. and Wolter, D. (2008). Spatial abstraction: Aspectualization, coarsening, and conceptual classification. In Freksa, C., Newcombe, N. S., Gärdenfors, P., and Wölfl, S., editors, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space, International Conference Spatial Cognition 2008, Freiburg, Germany, September 15-19, 2008. Proceedings*, volume 5248 of *Lecture Notes in Computer Science*, pages 311–327. Springer.

Galindo, J. (2007). *Conjuntos y Sistemas Difusos (Lógica Difusa y Aplicaciones)*. Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga.

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.

Gdalyahu, Y. and Weinshall, D. (1998). Flexible syntactic matching of curves. In Burkhardt, H. and Neumann, B., editors, *ECCV (2)*, volume 1407 of *Lecture Notes in Computer Science*, pages 123–139. Springer.

Gero, J. S. (1999). Representation and reasoning about shapes: Cognitive and computational studies in visual reasoning in design. In Freksa, C. and Mark, D. M., editors, *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, International Conference COSIT '99, Stade, Germany, August 25-29, 1999, Proceedings*, volume 1661 of *Lecture Notes in Computer Science*, pages 315–330. Springer.

Goldberger, J., Gordon, S., and Greenspan, H. (2003). An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 487–494, Washington, DC, USA. IEEE Computer Society.

Gonzalez-Abril, L., Velasco, F., Ortega, J. A., and Cuberos, F. J. (2009). A new approach to qualitative learning in time series. *Expert Systems with Applications*, 36:9924–9927.

Gottfried, B. (2005). Global Feature Schemes for Qualitative Shape Descriptions. In Ligozat, G., Guesgen, H. W., and Freksa, C., editors, *IJCAI 05 WS Spatial-Temporal Reasoning*.

Gottfried, B. (2008). Qualitative similarity measures: The case of two-dimensional outlines. *Computer Vision and Image Understanding*, 110(1):117–133.

Gotts, N. M. (1996). Using the RCC formalism to describe the topology of spherical regions. Technical Report, Report 96.24, School of Computer Studies, University of Leeds, 1996.

Goyal, R. and Egenhofer, M. (2000). Consistent queries over cardinal directions across different levels of detail. In *DEXA Workshops*, pages 876–880. IEEE Computer Society.

Goyal, R. and Egenhofer, M. (2001). Similarity of cardinal directions. *Lecture Notes in Computer Science*, 2121:36–55.

Greisdorf, H. and O'Connor, B. (2002). Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation*, 58(1):6–29.

Griffin, L. D. (2001). Similarity of psychological and physical colour space shown by symmetry analysis. *Color Research and Application*, 26(2):151–157.

Griffin, L. D. (2006). Basic colour foci and landmarks of the body colour solid. *Journal of Vision*, 6(13):47–47.

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: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. IOS Press, Amsterdam, The Netherlands, The Netherlands, 1st edition.

Guesgen, H.-W. (1989). Spatial reasoning based on Allen's temporal logic. Technical Report TR-89-049, ICSI, Berkeley, CA.

Hawkins, J. and Blakeslee, S. (2004). *On Intelligence*. Times Books.

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.

Herrero, D., Zamora, M., and Martinez, H. (2002). Fusion de sonar y laser mediante mapas de segmentos difusos. In *Proceedings of the III Workshop en Agentes Fsicos (WAF), Spain*, pages 57–69. Springer-Verlag.

Horrocks, I., Kutz, O., and Sattler, U. (2006). The even more irresistible $\mathcal{SROIQ}$. In *Proceedings of 10th International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, AAAI Press*, pages 57–67.

Horrocks, I., Patel-Schneider, P. F., and van Harmelen, F. (2003). From $\mathcal{SHIQ}$ and RDF to OWL: the making of a web ontology language. *Journal of Web Semantics*, 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, Stanford University, California, USA*, pages 261–283.

Hunter, J. (2006). Adding multimedia to the semantic web: Building 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.

Jörgensen, C. (1998). Attributes of images in describing tasks. *Information Processing Management: an International Journal*, 34(2-3):161–174.

Jungert, E. (1993). Symbolic spatial reasoning on object shapes for qualitative matching. In Frank, A. U. and Campari, I., editors, *Spatial Information Theory. A Theoretical Basis for GIS. European Conference, COSIT'93*, volume 716 of *Lecture Notes in Computer Science*, pages 444–462, Berlin. Springer.

Katz, Y. and Grau, B. C. (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.

Kuijpers, B., Moelans, B., and de Weghe, N. V. (2006). Qualitative polyline similarity testing with applications to query-by-sketch, indexing and classification. In de By, R. A. and Nittel, S., editors, *14th ACM International Symposium on Geographic Information Systems, ACM-GIS 2006, Proceedings*, pages 11–18. ACM.

Kurata, Y. (2008). The $9^+$-intersection: A universal framework for modeling topological relations. In Cova, T. J., Miller, H. J., Beard, K., Frank, A. U., and Goodchild, M. F., editors, *Geographic Information Science, 5th International Conference, GIScience 2008, Park City, UT, USA, September 23-26, 2008. Proceedings*, volume 5266 of *Lecture Notes in Computer Science*, pages 181–198. Springer.

Kurata, Y. and Egenhofer, M. (2007). The $9^+$-intersection for topological relations between a directed line segment and a region. In Gottfried, B., editor, *1st Workshop on Behaviour Monitoring and Interpretation, KI 2007*, pages 62–76.

Lai, X.-C., Kong, C.-Y., Ge, S. S., and Mamun, A. A. (2005). Online map building for autonomous mobile robots by fusing laser and sonar data. In *IEEE International Conference on Mechatronics and Automation*, volume 2, pages 993–998.

Laine-Hernandez, M. and Westman, S. (2006). Image semantics in the description and categorization of journalistic photographs. In Grove, A. and Steffl-Mabry, J., editors, *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology*, volume 43, pages 1–25.

Lammens, J. M. G. (1994). *A Computational Model of Color Perception and Color Naming.* PhD thesis, Faculty of the Graduate School of State University of New York at Buffalo, USA.

Landau, B. and Jackendoff, R. (1993). 'what' and 'where' in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2):217–265.

Latecki, L., Lakamper, R., and Eckhardt, U. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *Proceedings of the Computer Vision on Pattern Recognition (CVPR)*, pages 424–429.

Latecki, L. J. and Lakamper, R. (2000). Shape similarity measure based on correspondence of visual parts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1185–1190.

Leyton, M. (1988). A process-grammar for shape. *Artificial Intelligence*, 34:213–247.

Leyton, M. (2004). Shape as memory storage. In Cai, Y., editor, *Ambient Intelligence for Scientific Discovery*, volume 3345 of *Lecture Notes in Computer Science*, pages 81–103. Springer.

Li, B. and Fonseca, F. (2006). TDD - a comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation*, 6(1):31–62.

Li, S. and Ying, M. (2003). Region connection calculus: Its models and composition table. *Artificial Intelligence*, 145(1-2):121 – 146.

Li, S. and Ying, M. (2004). Generalized region connection calculus. *Artificial Intelligence*, 160(1-2):1 – 34.

Li, X., Huang, X., Dezert, J., Duan, L., and Wang, M. (2005). Robot map building from sonar and laser information using DSmT with discounting theory. *International Journal of Information Technology*, 3(2):78–85.

Ligozat, G. (1993). Qualitative triangulation for spatial reasoning. In Frank, A. U. and Campari, I., editors, *Spatial Information Theory. A Theoretical Basis for GIS. European Conference, COSIT'93*, volume 716 of *Lecture Notes in Computer Science*, pages 54–68, Berlin. Springer.

Ling, H. and Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:286–299.

Liu, C., Yuen, J., and Torralba, A. (2011). SIFT flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). To appear.*

Liu, H. (2008). A fuzzy qualitative framework for connecting robot qualitative and quantitative representations. *IEEE Transactions on Fuzzy Systems*, 16(6):1522–1530.

Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282.

Liu, Y., Zhang, D., Lu, G., and ying Ma, W. (2004). Region-based image retrieval with perceptual colors. In *Proc. of Pacific-Rim Multimedia Conference (PCM2004)*, pages 931–938.

Lovett, A., Dehghani, M., and Forbus, K. (2006). Learning of qualitative descriptions for sketch recognition. In *Proceedings of 20th International Workshop on Qualitative Reasoning (QR), Hanover, New Hampshire, USA, July 10-12.*

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.

Luo, M. R., Cui, G., and Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *COLOR Research and Application*, 26:340–350.

Macrini, D., Siddiqi, K., and Dickinson, S. J. (2008). From skeletons to bone graphs: Medial abstraction for object recognition. In *CVPR*. IEEE Computer Society.

Maillot, N. and Thonnat, M. (2008). Ontology based complex object recognition. *Image Vision Comput.*, 26(1):102–113.

Mark, D. M. and Egenhofer, M. J. (1992). An evaluation of the 9-intersection for region-line relations. In *GIS/LIS '92*, pages 513–521.

Martin, C., Schaffernicht, E., Scheidig, A., and Gross, H.-M. (2006). Multimodal sensor fusion using a probabilistic aggregation scheme for people detection and tracking. *Robotics and Autonomous Systems*, 54(9):721–728.

Menegaz, G., Troter, A. L., Sequeira, J., and Boi, J. M. (2007). A discrete model for color naming. *EURASIP J. Appl. Signal Process, Special Issue on Image Perception*, 2007(1):1–10.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65(2):43–72.

Missaoui, R., Sarifuddin, M., and Vaillancourt, J. (2004). An effective approach towards content-based image retrieval. In *Proc. of the International Conference on Image and Video Retrieval (CIVR)*, pages 335–343.

Mojsilovic, A. (2005). A computational model for color naming and describing color composition of images. *IEEE Transactions on Image Processing*, 14(5):690–699.

Moratz, R., Dylla, F., and Frommberger, L. (2005). A Relative Orientation Algebra with Adjustable Granularity. In *Proceedings of the Workshop on Agents in Real-Time and Dynamic Environments (IJCAI 05)*.

Moratz, R., Renz, J., and Wolter, D. (2000). Qualitative Spatial Reasoning about Line Segments. In Horn, W., editor, *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI00)*, Amsterdam. IOS Press.

Mori, G., Belongie, S., and Malik, J. (2001). Shape contexts enable efficient retrieval of similar shapes. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:723–730.

Moroney, N., Fairchild, M. D., Hunt, R. W. G., Li, C., Luo, M. R., and Newman, T. (2002). The CIECAM02 color appearance model. In *IST/SID 10 th Color Imaging Conference: Color Science, System and Application*, pages 23–27.

Motik, B., Horrocks, I., and Sattler, U. (2007). Bridging the gap between owl and relational databases. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 807–816, New York, NY, USA. ACM.

Motik, B., Sattler, U., and Studer, R. (2005). Query answering for owl-dl with rules. *J. Web Sem.*, 3(1):41–60.

Mukerjee, A. and Joe, G. (1990a). A qualitative model for space. In *Proceedings of the 8th National Conference of the American Association for Artificial Intelligence (AAAI-90)*, pages 721–727, Boston, MA. MIT Press.

Mukerjee, A. and Joe, G. (1990b). A qualitative model for space. In *Proceedings of the 8th National Conference of the American Association for Artificial Intelligence (AAAI-90)*, pages 721–727, Boston, MA. MIT Press.

Museros, L. and Escrig, M. T. (2004). A qualitative theory for shape representation. *Inteligencia Artificial. Revista Iberoamericana de IA*, 23:139–147.

Museros, L. and Escrig, M. T. (2007). Automating assembly of ceramic mosaics using qualitative shape matching. In *The 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, ISBN: 1-4244-0912-8*.

Neumann, B. and Moller, R. (2008). On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82–101.

Nickerson, D. (1976). History of the munsell color system, company, and foundation. *Color Research and Application*, 1(1):7–10.

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.

Oliveira, L., Costa, A., Schnitman, L., and de Souza, J. A. M. F. (2005). An architecture of sensor fusion for spatial location of objects in mobile robotics. In Bento, C., Cardoso, A., and Dias, G., editors, *Progress in Artificial Intelligence, 12th Portuguese Conference on Artificial Intelligence, EPIA 2005, Covilhã, Portugal, December 5-8, 2005, Proceedings*, volume 3808 of *Lecture Notes in Computer Science*, pages 462–473. Springer.

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. MIT Press.

Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In Shepp, B. E. and Ballesteros, S., editors, *Object perception: Structure and process*, pages 121–163. Erlbaum, Hilldale, NJ.

Papadias, D. and Delis, V. (1997). Relation-based similarity. In Laurini, R., Bergougnoux, P., Makki, K., and Pissinou, N., editors, *Proceedings of the 5th International Workshop on Advances in Geographic Information Systems (GIS-97)*, pages 1–4, New York. ACM Press.

Peris, J. C. and Escrig, M. T. (2005). Cognitive maps for mobile robot navigation: A hybrid representation using reference systems. In *Proceedings of the 19th International Workshop on Qualitative Reasoning (QR'05)*, Granz, Austria.

Plataniotis, K. N. and Venetsanopoulos, A. N. (2000). *Color image processing and applications*. Springer-Verlag, NY, USA.

Pobil, A. P. d. and Serna, M. A. (1995). *Spatial Representation and Motion Planning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

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), University of Warwick, UK, September 10-13*, pages 610–619.

Quattoni, A. and Torralba, A. (2009). Recognizing indoor scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 0, pages 413–420, Los Alamitos, CA, USA. IEEE Computer Society.

Ramisa, A. (2009). *Localization and Object Recognition for Mobile Robots*. PhD thesis, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, SPAIN. Supervised by Prof. Ramon López de Mántaras and Dr. Ricardo Toledo.

Ramisa, A., Tapus, A., Aldavert, D., Toledo, R., and de Mántaras, R. L. (2009). Robust vision-based robot localization using combinations of local feature region detectors. *Autonomous Robots Journal*, 27:373–385.

Randell, D.A., Cui, Z., and Cohn, A. G. (1992a). A spatial logic based on regions and connection. In *Proceedings 3rd International Conference on Knowledge Representation and Reasoning*.

Randell, D. and Witkowski, M. (2004). Tracking regions using conceptual neighbourhoods. In *Workshop on Spatial and Temporal Reasoning, ECAI-2004*, pages 63–72.

Randell, D. A. and Cohn, A. G. (1989). Modelling topological and metrical properties in physical processes. In *KR*, pages 357–368.

Randell, D. A., Cui, Z., and Cohn, A. (1992b). A spatial logic based on regions and connection. In Nebel, B., Rich, C., and Swartout, W., editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176. Morgan Kaufmann, San Mateo, California.

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.

Reece, S. (1997a). Data fusion and parameter estimation using qualitative models: The qualitative kalman filter. In *Proceedings of the Eleventh International Qualitative Reasoning Workshop (QR)*, pages 143–153.

Reece, S. (1997b). Qualitative model-based multisensor data fusion and parameter estimation using infinity-norm dempster-shafer evidential reasoning. In *SPIE Eleventh Annual International Symposium on Aerospace/Defence Sensing, Simulation, and Controls*, pages 52–63.

Reece, S. (2000). Self-adaptive multi-sensor systems. In P. Robertson, H. S. and Laddaga, R., editors, *First International Workshop Self-Adaptive Software (IWSAS)*, volume 1936 of *Lecture Notes in Computer Science (LNCS)*, pages 224–241. Springer-Verlag.

Reece, S. and Durrant-Whyte, H. F. (1995). A qualitative approach to sensor data fusion for mobile robot navigation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 36–41.

Requicha, A. A. G. (1980). Representations for rigid solids: Theory, methods, and systems. *Computing Surveys*, 12(4):437–464.

Richards, W. and Hoffman, D. D. (1985). Codon constraints on closed 2d shapes. In *Computer Vision, Graphics, and Image Processing*, pages 265–281.

Rubner, Y., Puzicha, J., Tomasi, C., and Buhmann, J. M. (2001). Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding*, 84:25–43.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., and Markey, M. K. (2009). Complex wavelet structural similarity: A new image similarity index. *IEEE Transactions on Image Processing*, 18(11):2385–2401.

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.

Schlieder, C. (1994). Qualitative shape representation. In *In*, pages 123–140. Taylor and Francis.

Schlieder, C. (1995). Reasoning about ordering. In Frank, A. and Kuhn, W., editors, *Proceedings of the 3rd Conference on Spatial Information Theory (COSIT)*, pages 341–349. Springer, Heidelberg, Berlin.

Schockaert, S., De Cock, M., Cornelis, C., and Kerre, E. E. (2008). Fuzzy region connection calculus: An interpretation based on closeness. *International Journal of Approximate Reasoning*, 48(1):332–347.

Schockaert, S., De Cock, M., and Kerre, E. E. (2009). Spatial reasoning in a fuzzy region connection calculus. *Artificial Intelligence*, 173(2):258–298.

Schuldt, A., Gottfried, B., and Herzog, O. (2006). Retrieving shapes efficiently by a qualitative shape descriptor: The scope histogram. In *Image and Video Retrieval, 5th International Conference, CIVR 2006, Proceedings*, pages 261–270.

Scivos, A. and Nebel, B. (2001). Double-crossing: Decidability and computational complexity of a qualitative calculus for navigation. In Montello, D. R., editor, *Spatial Information Theory: Foundations of Geographic Information Science, International Conference, COSIT 2001, Morro Bay, CA, USA, September 19-23, 2001, Proceedings*, volume 2205 of *Lecture Notes in Computer Science*, pages 431–446. Springer.

Seaborn, M., Hepplewhite, L., and Stonham, T. J. (2005). Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recognition*, 38(2):165–177.

Sebastian, T. B., Klein, P. N., and Kimia, B. B. (2001). Recognition of shapes by editing shock graphs. In *In IEEE International Conference on Computer Vision*, pages 755–762.

Sebastian, T. B., Klein, P. N., and Kimia, B. B. (2002). Shock-based indexing into large shape databases. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part III*, pages 731–746, London, UK. Springer-Verlag.

Shapiro, L., Moriarty, J., Mulgaonkar, P., and Haralick, R. (1980). Sticks, plates, and blobs: Three-dimensional object representation for scene analysis. In *AAAI-80*, pages 28–31.

Shokoufandeh, A., Dickinson, S. J., Jonsson, C., Bretzner, L., and Lindeberg, T. (2002). On the representation and matching of qualitative shape at multiple scales. In *ECCV*, page III: 759 ff.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., and Zucker, S. W. (1998). Shock graphs and shape matching. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 222, Washington, DC, USA. IEEE Computer Society.

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.

Sousa, P. and Fonseca, M. J. (2009). Geometric matching for clip-art drawing retrieval. *J. Vis. Commun. Image R.*, 20:71–83.

Stanchev, P. L., Jr., D. G., and Dimitrov, B. (2003). High level colour similarity retrieval. *International Journal on Information Theories and Applications*, 10(3):363–369.

Stefanidis, A., Agouris, P., Georgiadis, C., Bertolotto, M., and Carswel, J. D. (2002). Scale- and orientation-invariant scene similarity metrics for image queries. *International Journal of Geographical Information Science*, 16(8):749–772.

Stell, J. G. (1999). Boolean connection algebras: A new approach to the region-connection calculus. *Artificial Intelligence*, 122:111–136.

Super, B. J. (2004). Fast correspondence-based system for shape retrieval. *Pattern Recogn. Lett.*, 25(2):217–225.

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.

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

Tassoni, S., Fogliaroni, P., Bhatt, M., and Felice, G. D. (2011). Toward a qualitative 3D visibility model. In *Proceedings of 25th International Workshop on Qualitative Reasoning (QR'2011), co-located at the Joint International Conference on Artificial Intelligence (IJCAI'2011).*

Tobies, S. (2001). Complexity results and practical algorithms for logics in knowledge representation. Technical report, RWTH Aachen, Germany. PhD thesis.

Vamossy, Z., Kladek, D., and Fazekas, L. (2004). Environment mapping with laser-based and other sensors. In *Proceedings of the International Workshop on Robot Sensing (ROSE)*, pages 74–48.

Vasardani, M. and Egenhofer, M. J. (2009). Comparing relations with a multi-holed region. In Hornsby, K. S., Claramunt, C., Denis, M., and Ligozat, G., editors, *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009, Proceedings*, volume 5756 of *Lecture Notes in Computer Science*, pages 159–176. Springer.

Vernon, D. (2008). Image and vision computing special issue on cognitive vision. *Image and Vision Computing*, 26:1–4.

Vik, M. (2004). Industrial colour difference evaluation: Lcam textile data. In *AIC 2004 Color and Paints, Interim Meeting of the International Color Association, Proceedings*, pages 138–142.

Wang, R. and Freeman, H. (1990). Object recognition based on characteristic view classes. In *Pattern Recognition, 1990. Proceedings., 10th International Conference on*, volume i, pages 8–12 vol.1.

Wang, X., Matsakis, P., Trick, L., Nonnecke, B., and Veltman, M. (2008). A study on how humans describe relative positions of image objects. In *Lecture Notes in Geoinformation and Cartography, Headway in Spatial Data Handling, ISBN: 978-3-540-68565-4, ISSN:1863-2246*, pages 1–18. Springer, Berlin Heidelberg.

Weijer, J. V. D. and Schmid, C. (2007). Applying color names to image description. In *ICIP (3)*, pages 493–496. IEEE.

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.

Williams, M.-A. (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.

Wilson, R. A. and Keil, F. C., editors (1999). *The MIT Encyclopedia of the Cognitive Sciences*. the MIT Press, Cambridge, Massachusetts.

Wolter, D., Dylla, F., Frommberger, L., Wallgrün, J. O., Nebel, B., and Wölfl, S. (2007). Qualitative spatial reasoning for rule compliant agent navigation. In Wilson, D. and Sutcliffe, G., editors, *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference, May 7-9, 2007, Key West, Florida, USA*, pages 673–674. AAAI Press.

Wu, J., Christensen, H. I., and Rehg, J. M. (2009). Visual place categorization: problem, dataset, and algorithm. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, IROS'09, pages 4763–4770, Piscataway, NJ, USA. IEEE Press.

Wu, J. and Rehg, J. M. (2008). Where am I: Place instance and category recognition using spatial PACT. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8.

Wu, J. and Rehg, J. M. (2010). CENTRIST: A visual descriptor for scene categorization. *Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on*, pages 1–8.

Xia, L. and Li, S. (2006). On minimal models of the region connection calculus. *Fundamenta Informaticae*, 69(4):427–446.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8:338–353.

Zender, H., Mozos, Ó. M., Jensfelt, P., Kruijff, G.-J. M., and Burgard, W. (2008). Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502.

Zhai, H., Chavel, P., Wang, Y., Zhang, S., and Liang, Y. (2005). Weighted fuzzy correlation for similarity measure of color-histograms. *Optics Communications*, 247:49–55.

Zivkovic, Z., Booij, O., Kröse, B. J. A., Topp, E. A., and Christensen, H. I. (2008). From sensors to human spatial concepts: An annotated data set. *IEEE Transactions on Robotics*, 24(2):501–505.