DEPARTAMENTO DE LENGUAJES, SISTEMAS INFORMÁTICOS E
INGENIERÍA DEL SOFTWARE

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INFORMÁTICOS

TESIS DOCTORAL

Modelo Computacional para la Generación de Indicaciones en la Localización de
Objetos en Entornos Virtuales: Aspectos espaciales y perceptivos

Autora: **Graciela Lara López**
Licenciada en Informática por la
Universidad de Guadalajara

Directoras: **PhD. Angélica de Antonio Jiménez**
Doctora en Informática por la
Universidad Politécnica de Madrid

PhD. Adriana Peña Pérez Negrón
Doctora en Informática por la
Universidad Politécnica de Madrid

Madrid, Junio de 2016

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INFORMÁTICOS

DEPARTAMENTO DE LENGUAJES, SISTEMAS INFORMÁTICOS E
INGENIERÍA DEL SOFTWARE

DOCTORAL THESIS

Computational model for the generation of directions for object location in
virtual environments: spatial and perceptual aspects

Author: Graciela Lara López

Supervisors: PhD. Angélica de Antonio Jiménez
PhD. Adriana Peña Pérez Negrón

Madrid, June of 2016

Título:

Modelo Computacional para la Generación de Indicaciones en la Localización de
Objetos en Entornos Virtuales: Aspectos espaciales y perceptivos

Autor:
Graciela Lara López

Tribunal:

Presidente : D. Xavier Ferré Grau

Vocales : D. José Pascual Molina Massó
D. Pablo Gervas Gómez-Navarro
D. Arcadio Reyes Locuona

Secretario : D. Jaime Ramírez Rodríguez

Suplentes : D. Gonzalo Méndez Pozo
D. Salvador Sánchez Alonso

Acuerdan otorgar la calificación de

Madrid, 20 de Junio de 2016

Agradecimientos / Acknowledgments

*Doy gracias a mi padre **Dios** por ser mi compañero de vida ahora y por siempre, por ser quien me da la fuerza interior, para tener fe y creer en los milagros.*

Este trabajo no hubiera sido posible sin la colaboración de muchas personas que me brindaron su ayuda y sus conocimientos.

Quedo ***ESPECIALMENTE AGRADECIDA*** con mis directoras de tesis la ***Dra. Angélica de Antonio Jiménez*** y la ***Dra. Adriana Peña Pérez Negrón*** por su amistad, dirección, paciencia, disponibilidad, comentarios, revisión y corrección de este trabajo, pero sobre todo por su apoyo y comprensión en algunos momentos difíciles a lo largo de estos últimos años.

De igual manera agradezco:

- A la Dra. María Adriana Corona Nakamura por su ayuda y sus consejos, pero especialmente por su valiosa amistad.
- Al Dr. Cesar Octavio Monzón y al Mtro. Sergio Fernando Limones Pimentel por todo el apoyo que me brindaron, por su amistad y su confianza en mí.
- A la Universidad de Guadalajara, y la Universidad Politécnica de Madrid, por encontrar en ellas la oportunidad de formarme y aprender cosas significativas, para mi desarrollo profesional.
- Al Programa para el Desarrollo Profesional Docente, para el tipo superior (PRODEP) del Gobierno de México, la Universidad de Guadalajara y la Universidad Politécnica de Madrid, por financiar mi estancia del doctorado durante estos años.
- A los alumnos del practicum y de fin de grado de la Escuela Técnica Superior de Ingenieros Informático de la Universidad Politécnica de Madrid: Diego, Roberto, Javier, Adrián, Álvaro, Pablo y David, por su excelente ayuda en la implementación de los algoritmos y la aplicación de experimentos.

A mis padres Virginia y Rogelio

Por su esfuerzo, ejemplo, cariño y apoyo que me han brindado en la vida y en todos los proyectos que emprendo.

Ama esto es más tuyo y de la Mtra. Adriana Corona, que mío. GRACIAS por todo!!!!

A mis hermanos Blanca, Ernesto y Rogelio (Roy)

Por apoyarme siempre en mis propósitos, con su comprensión y sus buenos deseos.

A sobrinos (mis gorditos) Betito y Esme

Por sus oraciones, cariño y comprensión en los momentos que me acompañan. Mi niños lindos lo quiero mucho!!!!

A las Hermanas Benedictinas Misioneras de Tutzning: Pilar (†), Teresa, Timotea, Visi (†), Ana, Margarete, Gratia, Esmeralda, Hilda, Jania.

Por abrirme la puerta de su casa, por sus oraciones, su amistad y su hermosa compañía durante estos años.

A todos mis familiares y amigos de México y España

Gracias por sus oraciones y palabras de ánimo.

“Lo maravilloso de aprender algo es que nadie puede arrebatarémoslo” B. B. King

Abstract

This doctoral thesis presents a computational model that simulates the process of giving directions for the location of objects within virtual environments. The model allows the automatic generation of directions in natural language, using spatial and perceptual aspects, with reference objects.

In order to assist users in the process of location of objects, this work is supported by three pillars:

- *The proposal of a computational model of perceptual saliency for 3D objects applicable to Virtual Environments.* The computational model of saliency facilitates the evaluation and selection of the best reference object to be used in directions. It is based on the object's features with the major stimulus for the human visual system; three basic features of 3D objects (i.e. color, size and shape) are evaluated from two perspectives: individual and with respect to the context.
- *A user model and an explicit representation of virtual world semantics.* In order to make the directions for the location of objects as relevant to the user as possible, a model to represent previous education and experience of the user as well as his/her cognitive and perceptual attributes is presented. The perceptual characteristics allude to how the user interprets information through the visual system. The cognitive characteristics reflect somehow the processes of the human brain, i.e. how the user saves and manages information throughout time. Likewise, a semantic representation of the virtual environment is necessary for the generation of directions, in which properties and relations between objects in the virtual environment are captured by an ontology.
- *The automatic generation of directions to locate objects in natural language,* this takes into account reference frames and reference objects. For the selection of the best reference object three criteria were taken into account: (1) The perceptual saliency of objects; (2) The probability that the user remembers the location of the object, and (3) Prior knowledge of the user about the class of object.

This computational model was implemented as a software component that can be integrated into any virtual environment in which there is a need to guide users in the

location of objects, for example when intelligent pedagogical agents within virtual training environments demand users to interact with objects that previously have to be located by the user.

The results obtained through a set of experiments allowed us to collect evidence on the effectiveness and usefulness of the proposal, and to verify that the model is adaptive to user characteristics and to the context or situation.

Resumen

Esta tesis doctoral presenta un modelo computacional que simula el proceso de dar indicaciones para la localización de objetos en entornos virtuales. El modelo permite la generación automática de indicaciones en lenguaje natural, utilizando los aspectos espaciales y perceptivos, con objetos de referencia.

Con el propósito de ayudar a los usuarios en el proceso de localización de objetos, este trabajo se sustenta en tres pilares:

- *La propuesta de un modelo computacional de saliencia perceptual aplicable a objetos 3D en Entornos Virtuales.* El modelo computacional de saliencia facilita la evaluación y selección del mejor objeto de referencia, que se utiliza en la indicación. El modelo se basa en las características del objeto con mayor estímulo para el sistema visual humano; tres características básicas de los objetos 3D (es decir, el color, tamaño y forma), estas son evaluadas desde dos perspectivas: individual y con respecto al contexto.
- *Un modelo de usuario y una representación explícita de la semántica de entornos virtuales.* Con el fin de hacer que las instrucciones para la localización de objetos sean lo más claras y relevantes para el usuario, un modelo para representar la formación previa y la experiencia del usuario, así como sus características cognitivas y perceptivas se presenta. Las características perceptivas aluden a cómo el usuario interpreta la información a través del sistema visual. Las características cognitivas reflejan de alguna manera los procesos del cerebro humano, es decir, cómo el usuario guarda y gestiona la información en todo momento. Del mismo modo, una representación semántica del entorno virtual es necesaria para la generación de indicaciones, en la que las propiedades y las relaciones entre los objetos en el entorno virtual son capturados por una ontología.
- *La generación automática de indicaciones en lenguaje natural para localizar objetos,* esto toma en cuenta los marcos de referencia y objetos de referencia. Para la selección del mejor objeto de referencia se tuvieron en cuenta tres criterios: (1) La saliencia perceptual de los objetos; (2) La probabilidad de que el usuario recuerde la ubicación del objeto, y (3) el conocimiento previo del usuario acerca de la clase de objeto.

Este modelo computacional fue puesto en práctica como un componente de software que puede ser integrado en cualquier entorno virtual en el cual hay una necesidad de dirigir a usuarios en la posición de objetos, por ejemplo cuando agentes inteligentes pedagógicos dentro de entornos de educación virtual guían a usuarios para interactuar con objetos, estos antes tienen que ser localizados por el usuario.

Los resultados obtenidos a través de una serie de experimentos nos permitieron reunir pruebas sobre la eficacia y utilidad de la propuesta, para verificar que el modelo es adaptable a las características del usuario y al contexto o situación.

Table of Contents

LIST OF FIGURE.....	18
LIST OF TABLES	21
1 INTRODUCTION.....	24
1.1 Research Objectives.....	26
1.2 Structure of the Thesis.....	27
 2 THEORETICAL GROUNDWORK	30
2.1 Application of Virtual Environments	30
2.2 Spatial Language	35
2.3 Perceptual saliency	43
2.4 Color Space Models	46
2.4.1 RGB Color Model.....	46
2.4.2 LMS Color Model.....	47
2.4.3 CIELAB Color Model.....	51
2.5 Shape Descriptors	52
2.5.1 Methods for representing shape	55
2.5.2 Taxonomies of shape descriptors	55
2.6 User Modeling.....	62
2.7 Data ontologies.....	65
2.8 Conclusion.....	66
 3 STATE OF THE ART.....	68
3.1 Computational Models of Perceptual Saliency.....	68
3.2 Computerized spatial language generation for object location.....	78
3.2.1 Spatial language systems.....	78
3.2.2 Computer systems with spatial language	81
3.3 Conclusion.....	96
 4 COMPUTATIONAL MODEL OF PERCEPTUAL SALIENCY FOR 3D OBJECTS IN VIRTUAL ENVIRONMENTS.....	98
4.1 General Overview of the Computational Model of Perceptual Saliency	98
4.2 Modelling the Individual Saliency	101
4.3 Modelling the Contextual Saliency	118
4.4 General Saliency	140
4.5 Empirical Validation of the Saliency Model.....	141
4.5.1 Experimental evaluation of the shape saliency metric	141
4.5.2 Experimental evaluation of the computational model of individual perceptual saliency	155
4.5.3 Experimental evaluation of the computational model of contextual saliency	167

4.6 Conclusion.....	173
5 USER MODELING AND VIRTUAL WORLD SEMANTICS	177
5.1 Semantics of virtual worlds	177
5.1.1 Identification of the characteristics of an object within a virtual environment	178
5.1.2 Creation of the conceptual model of the world ontology	178
5.2 User modeling	180
5.2.1 Identification of the characteristics of the user	180
5.2.2 Creation of the conceptual model of the user ontology	181
5.3 Implementation of the ontologies with the Protégé tool	190
5.4 Conclusion.....	192
6 AUTOMATIC GENERATION OF DIRECTIONS TO LOCATE OBJECTS IN VES	194
6.1 Algorithm for the selection of the best reference object (SBRO)	202
6.2 Algorithm for the Selection of Spatial Relations (SSR).....	215
6.2.1 Exploratory study	217
6.2.2 Criteria for the computation of the spatial relations	219
6.3 Algorithm for the generation of directions when the OL is in the user's field of view	226
6.5 Conclusion.....	244
7 CONCLUSION AND FEATURE WORK	246
7.1 Discussion.....	246
7.2 Contributions	247
7.3 Future research lines	248
7.4 Scientific dissemination	249
REFERENCES	252

List of Figure

Figure 2. 1 Virtual Reality System Components (inspired by Mazuryk et al., 1996)	31
Figure 2. 2 Intrinsic frame of reference.....	38
Figure 2. 3 Relative frame of reference.....	39
Figure 2. 4 Absolute frame of reference.....	40
Figure 2. 5 Egocentric and allocentric frames of reference.....	41
Figure 2. 6 Vase with respect to flower	41
Figure 2. 7 RGB color model (inspired by Gossett et al., 2004).....	47
Figure 2. 8 Normalized relative sensitivity of people with the normal color vision (cone cell LMS wavelengths)	48
Figure 2. 9 Relative sensitivity of people with protanopia (inspired by Ku T., 2006).....	49
Figure 2. 10 Normal vision and colorblind vision (inspired by Ku T., 2006).....	50
Figure 2. 11 CIELAB coordinate system	51
Figure 2. 12 System process.....	64
Figure 2. 13 Example of RDF elements within a data ontology for knowledge representation .	66
Figure 3. 1 The assessment of the saliency of spatial or geographic features based on the trilateral relationship between observer, environment, and geographic feature (inspired by Caduff et al., 2008).....	72
Figure 4. 1 Scheme of computational model of perceptual saliency.....	100
Figure 4. 2 Sequence of transformations of the color of a 3D object.....	103
Figure 4. 3 Characterization of color	105
Figure 4. 4 Saliency by color	107
Figure 4. 5 (a) status and (b) position of an object.....	108
Figure 4. 6 Measures for the volume of an object.....	108
Figure 4. 7 View of a voxelized 3D object	109
Figure 4. 8 Characterization of size	110
Figure 4. 9 Saliency by size	112
Figure 4. 10 Contrast in size.....	112
Figure 4. 11 Other characteristics of an object (a) the shape, (b) the orientation and (c) the distance.....	113
Figure 4. 12 Shape factor: high and low	113
Figure 4. 13 Equal Shape factor	114
Figure 4. 14 Characterization of shape.....	115
Figure 4. 15 Empty and full space in the bounding box of a 3D object.....	116
Figure 4. 16 Saliency by shape.....	117
Figure 4. 17 Individual Saliency	118
Figure 4. 18 3D vision system (Inspired in (Cruz, 2007))	119
Figure 4. 19 View of a 3D scene perceived by the observer in 2D, with obstructions	120
Figure 4. 20 Color distributions within the context of figure 19(b)	123
Figure 4. 21 Coordinates of the average color in Cie-Lab	125
Figure 4. 22 Average of the color in the context.....	125
Figure 4. 23 Distance in color	127
Figure 4. 24 Standard normal distribution	128

Figure 4. 25 Saliency by color with respect to context	130
Figure 4. 26 Average size in the context	131
Figure 4. 27 Distance of size	132
Figure 4. 28 Saliency of size with respect to the context	133
Figure 4. 29 View of a scene with regular and irregular objects.....	134
Figure 4. 30 Average shape in the context	136
Figure 4. 31 Distance of the shape	137
Figure 4. 32 Saliency of shape with respect to the context	139
Figure 4. 33 Saliency with respect to the context.....	140
Figure 4. 34 General saliency	141
Figure 4. 35 View of an experimental system trial	143
Figure 4. 36 Dispersion diagrams showing the correlation between (a) the simple error and the standard deviation of the error in movements; (b) the simple error and the standard deviation of the error in saliency	152
Figure 4. 37 Dispersion diagram showing the correlation between the simple error and the diversity in saliency.....	154
Figure 4. 38 The set of nine objects selected for the experiment	156
Figure 4. 39 The colors used in the objects	157
Figure 4. 40 Three slides of the presentation of the objects for the participants.....	159
Figure 4. 41 View of the areas where participants wrote their answers.....	160
Figure 4. 42 Results of the multiple regression model based in the participants' answers in the test	165
Figure 4. 43 Dispersion diagram showing the correlation between the value of individual saliency and the average value of votes	166
Figure 4. 44 Scenes used in the study	169
Figure 4. 45 View of a sentence displayed to the users	170
Figure 4. 46 Set of objects used in the study	170
Figure 4. 47 Dispersion diagram showing the correlation between the value of contextual saliency and the average time of locating of the participants.....	173

Figure 5. 1 Hierarchy of classes of the world ontology	179
Figure 5. 2 Conceptual map of the world ontology.....	179
Figure 5. 3 Hierarchy of classes of the user ontology	181
Figure 5. 4 Conceptual map of the user ontology	182
Figure 5. 5 Rey-Osterrieth's complex figure	184
Figure 5. 6 Snellen's table	186
Figure 5. 7 Sequence of transformations for the conversion of the color of a 3D object in the "normal" RGB color space into its dichromatic versions.....	188
Figure 5. 8 Image of the Editor - Exporter in Unity.....	191

Figure 6. 1 Cases considered for automatic directions generation	195
Figure 6. 2 (a) The OL is in the same room that the user, but not in the user's field of view ...	196
Figure 6. 3 The OL is in another room.....	197
Figure 6. 4 The OL is in the user's field of view	198
Figure 6. 5 Scheme of a final sentence.....	200
Figure 6. 6 Syntactic structure of a sentence	200

Figure 6. 7 Clarity of vision	208
Figure 6. 8 Field of vision of a user	210
Figure 6. 9 Central and peripheral vision	211
Figure 6. 10 Percentage of spatial relations used by the users	217
Figure 6. 11 Representation of the spatial relation “close to”	220
Figure 6. 12 Representation of the spatial relations “left / right / above / below / in front of / behind”	221
Figure 6. 13 Spatial Cartesian coordinate system	222
Figure 6. 14 Representation of the spatial relation “inside”	223
Figure 6. 15 Representation of the ray collisions with a box	224
Figure 6. 16 Distances of two OR to an OL	224
Figure 6. 17 Representation of the spatial relation “first, second, third, etc.”	225
Figure 6. 18 Direction where the object to be located (OL) is inside another one	227
Figure 6. 19 Flow diagram of the algorithm to generate directions	228
Figure 6. 20 Flow diagram of the case c.1	229
Figure 6. 21 Listener centred intrinsic direction where the OL is salient in itself	230
Figure 6. 22 Flow diagram of the case c.2	231
Figure 6. 23 Direction where the user is very familiar with the environment and the RO has a high probability of being remembered	232
Figure 6. 24 Flow diagram of the case c.3.1 (a)	234
Figure 6. 25 Direction where the OR has not probability of been remembered	235
Figure 6. 26 Flow diagram of the case c.3.1 (b)	236
Figure 6. 27 Direction where the color of OR can help to locate the object	237
Figure 6. 28 Direction where the size of OR can help to locate the object	237
Figure 6. 29 Flow diagram of the case c.3.2 (a)	239
Figure 6. 30 Direction where the OR is selected by GS, PR and KE	240
Figure 6. 31 Flow diagram of the case c.3.2 (b)	241
Figure 6. 32 Direction based in the seen surface of OL	241
Figure 6. 33 Flow diagram of the case c.3.2 (c)	243

List of Tables

Table 2. 1 Different approaches to spatial language expressions.....	42
Table 2. 2 Types of color blindness	49
Table 2. 3 Categories of shape descriptors.....	58
Table 3. 1 Examined aspects of the analyzed models of saliency.....	77
Table 3. 2 Computational systems applying spatial language.....	83
Table 3. 3 Other areas addressed by computational systems	93
Table 4. 1 List of salient colors according to Choungourian	106
Table 4. 2 Relative frequency and Percentage of color for the scene in figure 19(b)	122
Table 4. 3 Mathematical representation of the colors in figure 4.19(b)	124
Table 4. 4 Color difference	126
Table 4. 5 Standard scores and saliency in color for the scenario of Figure 19(b)	129
Table 4. 6 Size difference.....	132
Table 4. 7 Values of standard scores and saliency in size.....	133
Table 4. 8 Values of 3D Zernike descriptors.....	135
Table 4. 9 Values of Euclidean distances in shape for the scenario of figure 4.29	137
Table 4. 10 Values of standard scores and saliency of shape.....	138
Table 4. 11 Order and value of saliency given by our metric, to the four objects of the twenty-five trials.....	144
Table 4. 12 Sample of the order given for a participant, to the four objects of the twenty-five trials.....	145
Table 4. 13 Sample of the order given by a participant, to the four objects of a trial and their values of EM and ES.....	147
Table 4. 14 Calculation of the diversity in saliency of a trial	148
Table 4. 15 Order of the objects in each trial: (a) based on our metric of saliency and (b) based on the mode. (When the order is the same in both, the cell is highlighted in green)	149
Table 4. 16 Values for the correlations between (a) the standard deviation of the error in movements (σ_{EM}) and the simple error and (SE), and the standard deviation of the error in saliency (σ_{ES}) and the simple error (SE).....	151
Table 4. 17 Values for the correlations between the simple error and (SE), and the diversity in saliency by trial	153
Table 4. 18 Colors and sizes used in the set of objects	157
Table 4. 19 Evaluation of how outstanding is a feature	161
Table 4. 20 Results obtained in the model of individual saliency by color, size, shape ordered by individual saliency	163
Table 4. 21 Contextual saliency of each object and its average time to be located	172
Table 5. 1 Criteria for evaluating the ROCF.....	185

Chapter **1**: Introduction

Computational model for the generation of directions for object location in virtual environments: spatial and perceptual aspects

1 Introduction

Virtual reality-based simulations are ever more present in different areas of research, such as modeling, design and planning, training and education, telepresence and teleoperation, medicine, psychological treatments, rehabilitation and entertainment.

Today's computer technology has made it possible to recreate, with the help of virtual reality, more or less any scene and real activity, sometimes programmed with the aim of providing support to human beings in tasks that are hazardous or processes that are too costly.

Reality can be simulated to represent something that is not there or project what something that still does not exist might be like. Simulation and virtual reality have helped human beings to solve many real-world challenges.

It is recognized that the human being is by nature an active entity who performs daily activities in response to needs and desires, such as personal hygiene, food, functional mobility—i.e., moving from one place to another—, object location in a physical space, etc.

In the real world, humans interact with objects that may be familiar or unfamiliar. People often spend a lot of time looking for objects, which may be lost, forgotten or hidden inside a room or perhaps elsewhere. Some factors that might complicate the execution of this task are: disorder/untidiness, the object having been moved from one place to another, the object being unknown or unfamiliar to the user, object visibility being poor because it is occluded by other objects and the user being unable to recall the object location. People might require clear and precise help in order to carry out some of these activities.

For example, a person who is being trained to operate machines for manufacturing and assembly might not be able to completely identify the physical structure of each machine. Therefore, they may require directions to help them locate a particular part. An easy solution

for automating help would be to give a predefined instruction, such as "please press the white button on the control table to switch on the machine". However, the on/off switch might not be in the same place, of the same color, or of the same shape on all machines. Precise directions may call for the combination of all possible elements, of which there may be a great many, and it would not be practical to predefine all possible cases.

The task of locating objects in the real world is very complex and variable. It largely depends on the position of the elements at the time, as well as previous knowledge of the environment and the skill of the person performing the task. A predefined solution could fail in an unforeseen situation. For example, a person who is going to prepare a cup of coffee cannot find the coffee machine because it has been moved and is not in its usual place.

A technique that is commonly used in the real world to help locate an object is to use additional elements in the environment of the object to be located. These objects are identified as reference objects, and they can be used as support for giving directions to the object location. For example, a book could be located using the shelf on which it is placed as reference object.

Reference objects are an effective resource, but what makes an object a good candidate for use as a reference object? The selection of a reference object can depend on several variables, such as:

- The physical characteristics of the object (color, size, shape).
- The spatial location of the object with respect the point of observation (to the right, to the left side, behind, etc.).
- Previous knowledge of objects or the environment in which they are located.
- The user's perceptual and cognitive characteristics.

On the other hand, when people learn object locations in a new environment, they interpret the spatial structure of that environment in terms of a spatial reference system (McNamara, 2003). Furthermore, Gapp (1996a) stated that object location is a task in which the integration of language and vision is a necessity per se.

In this respect, academic training environments, as well as staff virtual training in industry and virtual guided tours of museums are some of the activities that require object location and information exchange by conveying messages that can improve user learning and knowledge building.

Virtual tutors are a potentially useful resource for providing users with guidance and support for performing these activities: content learning, industrial equipment operation, museum tours, etc. The development and use of these virtual environments open up an infinite range of possibilities for user learning.

However, some of these activities require very specific information exchange. For example, a user touring the Dali Theatre and Museum may have a particular interest in viewing the objects on display in particular museum rooms that may not be easily viewable. This poses a problem unless the virtual system is intelligent. This problem can be solved by opting for the development of an intelligent virtual museum environment, that is, a system that uses natural language expressions to dialogue with the user and give simple and clear directions as to where the target object is.

In response to this need, this thesis proposes a mechanism enabling a system (intelligent virtual environment) to answer questions about where a particular object is in such a manner as is intelligible for users. This mechanism combines different elements, like reference objects, perceptive saliency, spatial reference systems, etc.

1.1 Research Objectives

The aim of this doctoral thesis is to propose a computational model for the automated generation of directions to help users to perform tasks involving object location in virtual environments. This model considers the different factors that are part of this task, specifically spatial and perceptual aspects.

The proposed model will provide users with support by providing natural language directions, simulating what a human being would do. The form of these directions is complex, because they are highly variable. The selection and use of a reference object is proposed as an effective and rapid mechanism in many situations. Some user cognitive and perceptual characteristics are also represented and considered in order to generate directions that are adaptive to specific user needs.

In the pursuit of this general objective, the thesis addresses the following practical goals:

- Definition and design of a computational model of object perceptual saliency, which is applicable to 3D virtual environments.
- Definition and design of mechanisms for the extraction of the features of a 3D object in order to measure its saliency.
- Definition and design of mechanisms for modeling the user and the adaptation of the system according to the user profile.
- Definition and design of mechanisms for selecting reference objects, intrinsically considering their key features depending on the context and the user.
- Definition and design of mechanisms for automatically generating natural language directions for object location.
- Development of a set of tools, which gather and implement all the elements established in the computational model proposed in this thesis for application to any virtual environment.

1.2 Structure of the Thesis

This thesis is divided into seven chapters.

Chapter 1: Introduction.

Chapter 2: This chapter describes a number of theoretical concepts applied in the main lines of research reported in this thesis, such as spatial language, frames of reference, reference objects, perceptual saliency, color space models, shape descriptors, user modeling, data ontologies.

Chapter 3: This chapter reviews the state of the art reported in the most significant work related to the proposal presented in this thesis.

Chapter 4: This chapter describes the proposal of a computational model of perceptual saliency for 3D objects based on an individual and contextual approach. It also describes a number of mechanisms for extracting the characteristics of a 3D object. They determine the saliency of a 3D object within a virtual environment in a numerical format. Also, in this chapter describes the process followed in order to empirically validate the proposed computational model perceptual saliency for 3D objects. Three controlled experiments were run to conduct the validation, which yielded some preliminary results.

Chapter 5: This chapter describes the representation of the virtual environment semantics required for this computational model. Additionally, this chapter describes the user modeling approach that will be used to represent the key characteristics of the users interacting with the proposed intelligent virtual environment. Two types of characteristics —perceptual and cognitive— are custom modeled. Both the semantic description of the virtual environment and the user model are represented as data ontologies.

Chapter 6: This chapter presents the mechanisms for generating natural language user directions to facilitate object location in a virtual environment. The directions conform to different criteria, such as user previous knowledge, user perceptual and cognitive characteristics, virtual environment scene layout, semantics associated with the scene objects, previous system use history, as well as other key user and context factors.

Chapter 7: This chapter summarizes the main contributions and the general conclusions of this thesis. It also describes the opportunities and lines of future work derived from this research.

Finally, the literature referenced throughout the thesis is listed.

Part I: Theoretical Groundwork & Relate Work

Chapter 2: Theoretical Groundwork

2 Theoretical Groundwork

The computational model proposed in this thesis is based on several fields of study, ranging from virtual reality (VR), through spatial language, spatial reference systems, perceptual saliency, color representation models and methods for object shape representation, to the field of the user modeling using a data ontology, also bearing in mind all the key concepts within a virtual environment. Therefore, the purpose of this chapter is to define all the concepts, theories and methods that are a formal bedrock of the computational model proposed in this doctoral thesis.

2.1 Application of Virtual Environments

The concept of virtual reality (VR), also known as virtual environments, synthetic experience, virtual worlds, artificial worlds or artificial reality (Mazuryk and Gervautz, 1996), can be defined as a paradigm that uses computers and human-computer interfaces (HCI) to create the effect of a three-dimensional (3D) world in which the user interacts directly with virtual objects. Virtual reality is computer-generated, three-dimensional, interactive and immersive. Virtual reality is an effect, not an illusion (Bryson, 1996).

This technology began to become popular and fashionable as of the 1990s. However, Ivan E. Sutherland presented the idea of virtual reality in 1965, as “*make that (virtual) world in the window look real, sound real, feel real, and respond realistically to the viewer’s actions*” (Sutherland, 1965). Sutherland is considered the father of this technology. Nevertheless, the term *virtual reality* is credited to Jaron Lanier, founder of VPL Research (Botella et al., 1998; Machover and Tice, 1994). Earlier experimenters, like Myron Krueger in the mid-1970s, used phrases like *artificial reality* (Machover and Tice, 1994).

Systems require several crucial components to get the effect of immersive virtual reality, that is, a sense of presence, like “the participant's sense of *being there* in the virtual environment” (Bowman and McMahan, 2007; Brooks, 1999; Bryson, 1996; Mazuryk and Gervautz, 1996):

- A usually stereoscopic, head-tracked and head-mounted display that presents the virtual world from the user's current head position and has the visual cues required so that the virtual scene is perceived as independent of the user, that is, the object stays still while the user moves about, for example. This requires a tracking system that continually reports the position and orientation of the user's head and limbs.
- A high-performance computer graphics system that computes and renders the virtual world. This is a graphics rendering system that generates 20 to 30 frames per second.
- A database construction and maintenance system for building and maintaining detailed and realistic models of the virtual world; and
- Three-dimensional input devices (3D mouse, wand, joystick, etc.) enabling the user to interact with the system directly in three dimensions (see Figure 2.1).

Figure 2. 1 Virtual Reality System Components (inspired by Mazuryk et al., 1996)

Brooks (1999) cited some important but not so crucial auxiliary technologies:

- Synthesized sound, including directional sound and simulated sound fields, for auditory perception;

- Synthesized forces and other haptic sensations for kinesthetic perception;
- Devices, such as wired gloves with pushbuttons, used by the user to specify interactions with virtual objects, and substitutes for real interaction techniques; and
- Interactions with the physical world, exploring augmented reality.

VR has captured people's imagination, providing a powerful, easy and intuitive form of human-computer interaction. With the development of this technology, the user can see and manipulate a simulated environment in more or less the same way as they interact with the real world (Mazuryk and Gervautz, 1996).

In the 1990s, people began to regard virtual reality as a new medium, a new form of entertainment, and a new and very powerful art form. In this context, the behavior of virtual agents, the dramatic content, and the presentation style vary according to explicit artistic models built into the world by their creators (Bates, 1991).

VR has been applied in different areas, such as teleoperation, entertainment, vehicle simulation, vehicle design, training, medicine (Brooks, 1999; Mazuryk and Gervautz, 1996), and all sorts of interactive applications (Gutierrez et al., 2005). There follows a brief description of some of its applications.

- Modeling, Designing and Planning

The manufacturing industry is an economic activity responsible for transforming a wide variety of raw materials into different products for consumption by people. Mujber et al. (2003) stated that it is becoming increasingly difficult to competitively meet customer demands. The advances in virtual reality technology have boosted the application of VR to different engineering applications, such as product design, modeling, shop floor controls, process simulation, manufacturing planning, training, testing and verification.

With regard to modeling, virtual reality offers the possibility of looking in real time and in real space at what the modeled object will look like. Users can change the colors, textures and positions of the objects, immediately seeing what the whole surroundings would look like. A big advantage of this technology is that users can see and even feel the fashioned surface under their fingertips (Mazuryk and Gervautz, 1996).

The use of virtual prototypes has become an interesting product development strategy. The automotive and aerospace industries have pioneered the application of virtual reality. These industries have adopted VR technology for design applications, assembly verification and the maintenance process. VR has also been considered a good tool for getting quick and intuitive answers within business and product planning processes (Gomes and Zachmann, 1999).

- Training and Education

Military training provided one of the first applications of immersive VR. The military can train infantry in urban combat tactics in a virtual city filled with computer-generated enemies and friendly troops. Training in a virtual world is a good option bridging the traditional alternatives of classroom-based training and real-world training exercises. Virtual training provides a level of realism that is out of the question in the classroom and is more flexible and less costly than real-world exercises. Recently, the success of VR military training has led to the adoption of VR technologies for other types of training (Bowman and McMahan, 2007).

VR has become an effective means for automotive industry employees to access training services and interactive manuals (Gomes and Zachmann, 1999). Flight simulators are another example of virtual training. They have been considered as the precursors of VR. Many civilian companies are offering this type of low-cost training (Mazuryk and Gervautz, 1996). Nowadays this type of training provides increasingly real experiences. Users have the option of choosing the type of plane that they want to fly and to which airport, as well as the real weather conditions, including turbulence, thunderstorms, clear skies and other options. They can also choose to be military, commercial, independent or aerobatic pilots (Simuladordevuelo.net, 2014).

NASA has used other training applications with VR. This training is very valuable, and there are few and poor alternatives to VR technology. Extra-vehicular activity is very hard to simulate. Moving around on the outside of a space vehicle is another skill that cannot be trained on earth. VR systems provide astronauts with simulated practice (Brooks, 1999).

- Telepresence and Teleoperation

Teleoperation is a technology enabling people to operate machines in remote environments by means of VR user interfaces. In many cases, this form of remote control is the only option: the distant environment may be hazardous to human health or life, and no other technology supports such a high level of operational dexterity (Mazuryk and Gervautz, 1996).

- Psychological Treatments

Rothbaum et al. (2000) claimed that any method capable of activating and modifying the fear structure could be predicted to improve the symptoms of anxiety. In this sense, virtual reality exposure therapy (VRE) has been proposed as a new medium to treat disorders. VR systems are applicable to the treatment of phobias, such as fear of spiders (arachnophobia), fear of flying (aerophobia), fear of open spaces (agoraphobia), fear of being confined in a closed space (claustrophobia), and fear of heights (acrophobia). Botella et al. (1998) applied VR for the treatment of claustrophobia. The therapy involved presenting patients with multiple virtual environments for several five-minute sessions. The patients were encouraged to interact with the virtual environments for a time to decrease their anxiety. The fear measures were drastically reduced, providing evidence of the effectiveness of the therapy. According to clinical outcomes, the treatment of phobias with the support of VR has been found to be a great success. Some other advantages are that virtual reality therapies are less expensive, less time consuming, less risky, and less embarrassing.

- Rehabilitation

VR proposes clinical assessment and rehabilitation options that are not available with traditional methods. VR also offers a variety of new options within rehabilitation, including the potential to objectively measure behavior in challenging but safe and ecologically valid environments, while maintaining experimental control over stimulus delivery and

measurement. The advantages of VR in clinical assessment and rehabilitation applications are: a more natural or “real-life” environment, control of stimulus presentation and response, measurement, safe assessment of hazardous situations, increased generalization of learning, increased standardization of rehabilitation protocols, and increased user participation (Schultheis and Rizzo, 2001) .

- Entertainment

In recent years, lower prices and more powerful hardware have finally brought VR to the masses, and it is being applied in entertainment. The entertainment industry has successfully brought networked multi-player game systems on to the market. Apart from these complicated systems, the home entertainment market has rapidly expanded since the 1990s (Mazuryk and Gervautz, 1996).

These are just a few examples from a wide range of VR applications that have been developed over the last thirty years.

2.2 Spatial Language

Spatial language is the syntax used for object or place locations. Object location is a task that necessarily combines spatial language and spatial vision. Albeit a seemingly simple task, it calls for a mix of human knowledge theories and an accessible visual representation for the linguistic system.

An important feature of living beings and their surrounding objects, all of which have a place in space, is undoubtedly their spatial position (Moratz and Tenbrik, 2006). Accordingly, people develop spatial knowledge, a basic skill that is helpful for the location process. The representations underlying object recognition, object search and spatial navigation are fundamental components of this spatial knowledge. Through our ability to use these representations to express our spatial experience, we can formulate a spatial language in order to discuss where objects are located (Landau and Jackendoff, 1993).

Regier and Carlson (2001) pointed out that “the linguistic categorization of space serves as an interface between language and the perceptual world”. (Mark and Frank (1989) described

spatial language as the terms in human language that people use to refer to spatial situations and as an important indicator of the major ways in which people conceptualize space. Spatial language consists of expressions that serve the purpose of describing the spatial relation, naturally offering specific details about objects and their environment, whereas a linguistic expression is considered to be a prepositional phrase of location that implicitly contains an indication of where an object or place is located (Skubic et al., 2002). The linguistic concepts of syntax and semantics are able to structure these natural language expressions by identifying the spatial relations between entities.

Research on the use of spatial language has focused on the representation of the relation between two objects. The description of a spatial relation consists of one entity, a “*locatum*”, or object to be located (OL), relative to another entity, a “*relatum*”, or reference object (RO), and a “spatial term” that describes the relation between the “*locatum*” and the “*relatum*” (Tenbrink and Ragni, 2012). These three roles may be played out in a variety of ways depending on several factors. In the formulation of spatial language, the spatial reasoning task is based mainly on a small subset of relational terms, with a large collection of linguistic expressions offered by each language.

Gapp (1995, 1996a) pointed out that the object location task is divided into three steps: 1) identify the object to be located (target object); 2) select a useful object that serves as a reference (reference object); and 3) develop a linguistic expression in relation to both objects.

The location of an object is inherently relative and cannot be referenced without establishing a frame of reference (Mou and McNamara, 2002; Shelton and McNamara, 2001). For decades, frames of reference (FoRs) have been a topic of interest to the research community in different scientific and technological areas. According to Rock (1990), the notion can be reduced from Gestalt theory to “a unit or organization of units that collectively serve to identify a coordinate system with respect to which certain properties of all objects, including the phenomenal self, are gauged”. Thus, spatial reference frames, also known as spatial reference systems (we use both expressions in this thesis), consist of a spatial coordinate system used in multiple ways to specify the location of objects (Levinson, 1996; Majid et al., 2004; O'Meara and Pérez, 2011; Tenbrink, 2005; Wraga et al., 1998).

Frames of reference are considered important for the generation of spatial language. In object location, it is important to understand that the space cannot be described or specified using a reference object without an implicit spatial reference frame. For example, a possible simple expression for answering a question like "Where is the book?" would be "It is to the left of the lamp". This is the kind of everyday expressions encoded by human beings in order to locate objects in physical space. However, people from other cultures might express this scenario differently, like, for example, "The book is beside the lamp". The scenario is the same, but there are clear differences as to how the directions given to locate the object are encoded and represented (Majid et al., 2004). In another example, the location of a chair in a room can be expressed in terms of the room itself, as "The chair is in the middle of the room", or from the viewpoint of the speaker, as "The chair is in front of me" (Majid et al., 2004; Vorweg, 2003).

As mentioned above, frames of reference act as internal systems that help to project space regions in order to provide a simple description for locating whatever we are looking for. Following Levinson (1996), we can describe different types of frames of reference.

- *Intrinsic frame of reference.* This is when the spatial reference is based on a landmark's origin and its canonical position (Stoia, 2007). According to Barclay (2010), the intrinsic reference system focuses on the reference object (RO) and adopts a reference frame according to the arrangement of the cardinal axes of the object. On the other hand, when people learn a spatial layout, they interpret the spatial structure in terms of an intrinsic reference system, which is defined by the actual layout. Likewise, the chosen intrinsic reference in each case can be determined by the spatial or the non-spatial properties of the objects, the structure of the surrounding environment and the viewpoint of the observer (Mou and McNamara, 2002; Shelton and McNamara, 2001).

Here, the object's features, such as shape, orientation, movement or use, are essential for determining the coordinates. For example, considering the objects illustrated in Figure 2.2, a possible expression for locating the nurse would be "The nurse is in front of the hospital" or, alternatively, "The nurse is beside the hospital",

with an intrinsic reference to the front of the hospital (Coventry and Garrod, 2004; Levinson, 2003).

Figure 2. 2 Intrinsic frame of reference

where

- X is the origin of the coordinate system.
- F is a figure or referent with center point at volumetric center F_c .
- G is ground or relatum, whose volumetric center is G_c , and whose surrounding region is r .

- *Relative frame of reference.* The relative frame of reference focuses on the listener and the viewpoint of the speaker. This frame of reference is usually established by means of the orientations of the planes through the human body, leading to a series of half-lines, like up/down, back/front, and left/right, which are typically centered on the dorsal spine or the principal axis of the body (Levinson, 1996, 2003).

This is a commonly used frame of reference for object location, although the three-point Hermann location system is also in widespread use. This system implies using the

resource of the position of a third object instead of an intrinsic frame of reference with respect to the speaker. Looking at Figure 2.3, for example, the expression would be “Looking from the church, the nurse is on the left of the hospital”. In this case, the position of the object is specified in relation to an object from the speaker’s viewpoint but targeting the listener (Moratz et al., 2003). This frame of reference can also be based on human projections, such as “in front of me”, or “to my left” (Pederson et al., 1998).

Figure 2. 3 Relative frame of reference

where

V is the viewpoint of observer

- *Absolute frame of reference.* The absolute frame of reference is described as a fixed bearing provided by cardinal points or the visual horizon, that is, focused on the environment and its salient features. In this frame of reference, the cardinal points correspond to directions that are related to geographical points, such as north, south, east, west, etc. This type of frame of reference, then, considers neither a third object nor the intrinsic characteristics of the object to be located (Coventry and Garrod, 2004). Using the same example of the nurse and the hospital, you could say "The nurse is to the north of the speaker, the listener or the hospital" (see Figure 2.4). If the nurse

were located in a group of nurses, you could say that she is located to the north of the whole group. Absolute reference systems are a special case in that there is no way of labeling “origins” or “relatum” in a way that is consistent with the other kinds of reference systems, as directions behave differently than entities (Moratz et al., 2003).

Figure 2. 4 Absolute frame of reference

- *Egocentric and allocentric frames of reference.* Another way to classify frames of references is to distinguish whether they are represented with respect to the egocentric or the allocentric space. The egocentric frame of reference represents the location and orientation of an object using peripheral aspects in reference to the observer’s body (Mou and McNamara, 2002). The origin of the coordinate axes for the egocentric reference frame is inside the body or mentally inside the self (Ma et al., 2012; McNamara, 2003); see Figure 2.4 (left). In the allocentric frame of reference, the object location is represented with respect to the environment and independently of the observer (Ma et al., 2012; Mou and McNamara, 2002; Shelton and McNamara, 2001). The origin of the coordinate axes for this frame of reference is located outside the self; see Figure 2.5 (right). This reference frame is also called exocentric or geocentric (Klatzky, 1998).

The egocentric and allocentric approaches can be associated with the above intrinsic, relative and absolute frames of reference. Because the egocentric frame of reference is based on the viewpoint of the interactants, it is equivalent to a relative frame of reference when the expression is centered on the speaker. The allocentric reference frame is based on the environment and can therefore be related to both the intrinsic and absolute frames, and, if the expression is centered on the listener, even the relative reference frame.

Figure 2. 5 Egocentric and allocentric frames of reference

Linguistic expressions, then, may vary depending on descriptions in relation to other objects and their location (Majid et al., 2004). For example, using different types of frames of reference to describe the location of the vase with respect to the flower in Figure 2.6, we can say “The vase is to the left of the flower” (relative FoR), “The vase is to the north of the flower” (absolute FoR) and “The vase is at the tip of the flower” (intrinsic FoR).

Figure 2. 6 Vase with respect to flower

Table 2.1 shows other examples of simple expressions of location using two different approaches; the listener is facing the located object in both cases (expressions based on (Barclay, 2010)).

Table 2. 1 Different approaches to spatial language expressions

First approach		Second approach	
Frame of reference	Example	Frame of reference	Example
<i>Intrinsic / egocentric (speaker centered)</i>	“The chair is in front of me”	<i>Intrinsic (object centered)</i>	“The chair is in front of the desk”
<i>Intrinsic (listener centered)</i>	“The flowers are in front of you”	<i>Intrinsic (object centered)</i>	“The flowers are on the table”
<i>Relative (speaker centered)</i>	“The cat is in front of the house”	<i>Relative (listener centered)</i>	“The cat is in front of the house (from your point of view)”
<i>Relative (third party centered)</i>	“The museum is in front of the hospital from James's point of view”	<i>Relative (speaker centered)</i>	“The museum is in front of the hospital”
<i>Absolute (using the cardinal points)</i>	“The museum is north of the city”	<i>Absolute (fixed bearing)</i>	“The museum is downtown”

There is not one answer to the question, which of these spatial reference frames is likely to be most useful? Several factors influence the preference for a specific frame of reference, including, to name but a few, the functional relation of the objects in the environment, the features of the objects or their movement, the specific purpose of each particular task, the need for the listener to liaise with the speaker or the perspective adopted in the scene (Carlson-Radvansky and Logan, 1997).

Moreover, several cognitive criteria can be used to select the reference objects. In some cases, these may be the distance between the target and the reference object, functional dependencies of objects, prior knowledge about the objects, the perceptual saliency of object, etc. (Gapp, 1995). The concept of perceptual saliency, explained below, is essential for understanding which characteristics or which objects are important for qualification as a reference object.

2.3 Perceptual saliency

Perceptual saliency, also known as visual saliency or prominence, is unquestionably an important concept in the development of this thesis.

An object may have many different characteristics, but some of them are more prominent than others. When people observe an image or an object, they focus on certain areas to which they are attracted, areas that contain information that they regard as relevant (Undurraga and Mery, 2011). Generally speaking, saliency is a key concept of psychology, mentioned by Titchener (1908) (Titchener, 1908), one of the first writers in this field. In recent years, this concept has been applied in information technology for object analysis and computer vision (Huang and Pashler, 2005).

Some of the most cited definitions of perceptual salience are:

- The visual saliency of an object depends on the interaction of its basic features, color, size, and shape mainly (Gapp, 1995; Hoffman and Singh, 1997; Spotorno et al., 2013; Stoia, 2007), with respect to other objects.

- Perceptual saliency is the term used to refer to a bottom-up task with independent factors to which human beings pay more attention (Itti et al., 1998).
- Visual saliency is a broad term that refers to the idea that certain parts of a scene are pre-attentively distinctive and create some form of immediate significant or visual arousal within the early stages of the human visual system. The term ‘pop-out’ is used to describe the visual saliency process occurring at the pre-attentive stage. Certain visual primitives are immediately perceivable—they ‘pop-out’. Naturally, saliency implies rarity (Kadir and Brady, 2001).
- According to Hall et al. (2002), perceptual saliency intuitively corresponds to the rarity of a feature.
- Oliva et al. (2003) (Oliva et al., 2003) defined perceptual saliency from the viewpoint of the likelihood of finding a set of local features, or regions of interest in an image.
- According to Kapur, visual saliency is a process of association of objects and their representation in order to attract attention and capture thinking and behavior.
- Caduff and Sabine (2008) stated that perceptual saliency refers to the effect of drawing attention to the saliency of a feature.
- Frintrop and Rome (2010) consider, based on Egeth and Yantis (1997) (Egeth and Yantis, 1997), that regions of interest that attract our attention in a bottom-up way are called salient, and the feature responsible for this reaction must be sufficiently discriminative with respect to surrounding features. Besides bottom-up, this attention-grabbing mechanism is also called exogenous, automatic, reflexive or peripherally cued.

- Katramados and Breckon (2011) built upon the definition given by Achanta et al. (2009) in order to specify that visual saliency is the perceptual quality that makes a group of pixels stand out with respect to its neighbors.
- Objects can certainly attract the perception of viewers due to their absolute, natural or distinctive features (Vargas and Lahera, 2011).
- Röser et al. (2011; 2012; 2013) define visual saliency as referring to all visual features of an object, such as size, shape, color, texture, etc. The visual features of landmarks (visual saliency) must have a high contrast with their environment. Strictly speaking, an object is perceptually salient if it is an outlier, meaning that it is quite different from the other existing objects.
- According to Lahera et al. (2013), perceptual saliency is the automatic and subliminal process of bottom-up visual discrimination, whereby certain stimuli stand out from the perceptual field and attract attention. In other words, it is the higher-order mental process by means of which certain perceived or mentally represented objects attract the focus of attention, including thinking and behavior.
- Li et al. (2013) define visual saliency as the local complexity. They consider that, generally speaking, there are two different processes that influence visual saliency: one is top-down, and depends on the task at hand, and the other is bottom-up, which is driven by the input image.

However, a good perception of an object depends on factors such as visual acuity, clarity of vision and the viewpoint of the observer. On the other hand, the psychological impact that the features of an object might have on the viewer also have an influence on whether an object is prominent.

2.4 Color Space Models

The color of the objects is especially important in this research. It is used in the design of several algorithms described in Chapters 4 and 5. For this reason, we give a brief description of the mathematical models that can be used to characterize color.

The three color models (also called color spaces) used to characterize the color of the objects that were analyzed in this research were: the RGB model (red, green and blue), the LMS model (long, middle and short wavelengths), (color model used when performing chromatic adaptation on a sample under a different illuminant) and the CIELAB model (color model normally used to describe all the colors perceivable by the human eye).

2.4.1 RGB Color Model

The RGB system is a color space originated from the combination of three colored rays (red, green and blue). This color model is one of the most widely used for processing and storing data from digital images.

The RGB color model is a common and popular color space, mainly within the computer graphics community, television and for coding images on the Internet. Any color within the computer monitor color gamut can be produced using a combination of red, green and blue (primary colors). This usually encompasses most of the colors that are perceivable by the human visual system. The RGB model is an aggregate color model and uses transmitted light to display colors. Thus the color combination in RGB is handled additively. Accordingly, when red and green combine, the result is yellow, when red and blue combine, the result is magenta, and when green and blue combine, the result is cyan. The RGB system captures all the color space in a cube comprising the orthogonal RGB Cartesian coordinates (see Figure 2.7). Each color is represented by a point in the cube. In a normalized RGB system, the point (0, 0, 0) means the absence of color, that is, black, while point (1, 1, 1) represents the presence of all the colors, that is, white (Gossett and Chen, 2004; Yam and Papadakis, 2004). All grey colors are present in the main diagonal from black ($R=G=B=0$) to white ($R=G=B=255$) (Viscarra et al., 2006). This model is closely related to the way in which color is perceived in

the human retina. Furthermore, this model is device dependent, since its range of color varies with the display device (Yam and Papadakis, 2004).

Figure 2. 7 RGB color model (inspired by Gossett et al., 2004)

The RGB model is best for multicolored object recognition where all image conditions are controlled (Gevers and Smeulders, 1999). However, RGB is a not very good option for color analysis and color-based recognition algorithms because of the high correlation between channels and significant perceptual non-uniformity mixing of chrominance and luminance data (Vezhnevets et al., 2003). Also the model is device dependent, since the color range varies with the display device (Yam and Papadakis, 2004).

2.4.2 LMS Color Model

The LMS system is a color space that specifies colors in terms of the relative excitations of long-wave sensitive (L), middle-wave sensitive (M) and short-wave sensitive (S) cones (Anagnostopoulos et al., 2007; Doliotis et al., 2009).

This idea requires some knowledge of the acquisition of visual information by the human visual system. The human retina has two types of photoreceptor cells: rod cells that only work in the dark, and cone cells that only work in the light (Ku and Aoyama, 2006; Tkalčič and Tasič, 2003). Trichromatic theory states that there are three types of cone photoreceptors, approximately sensitive to the red, green and blue regions of the light spectrum. The three

types of cones are usually referred to as L (red), M (green) and S (blue) cones (Ku and Aoyama, 2006; Neitz and Neitz, 2005; Tkalčič and Tasič, 2003). People with normal color vision have all three types of cone cells (see Figure 2.8).

Figure 2. 8 Normalized relative sensitivity of people with the normal color vision (cone cell LMS wavelengths)

A degree of affectation in the cells can generate some sort of color blindness. Acromacy, monochromacy, dichromacy and anomalous trichromacy are forms of color blindness characterized by the presence of none, one, two or three functional types of cone, respectively. They are all caused by mutations that lead to an altered complement of functional cone photopigments expressed in the retina (Neitz and Neitz, 2005).

There are two types of trichromatic anomaly: protanomaly and deuteromaly (Jefferson and Harvey, 2007). There are three possible types of dichromatic deficiency: protanopia, deuteranopia and tritanopia. People with protanopia (protanopes) can be divided into two groups: people who do not have L cone cells that mainly detect red light (strong protanopia) among the three types of cone cells; and people who have L cone cells with their spectral sensitivity shifted more toward M cone cells (weak protanopia) (see Figure 2.9).

Figure 2. 9 Relative sensitivity of people with protanopia (inspired by Ku T., 2006)

The vision of people with strong protanopia is significantly different from normal vision, whereas the vision of people with weak protanopia is somewhere between the two extremes. Similarly, people with deuteranopia (deuteranopes) can be divided into two groups: people who have no M cone cells that mainly detect green light (strong deuteranopia) and people who have M cone cells whose spectral sensitivity is shifted more toward L cone cells (weak deuteranopia) (Jefferson and Harvey, 2007; Ku and Aoyama, 2006). Table 2.2 below summarizes these points.

Table 2. 2 Types of color blindness

Name		Cause
Acromacy		Missing or abnormal L, M and S cones
Monochromacy		Presence of only one of the three cone pigments
Dichromacy	Protanopia	Missing L cone
	Deuteranopia	Missing M cone
	Tritanopia	Missing S cone
Trichromacy	Protanomaly	Abnormal L cone
	Deuteranomaly	Abnormal M cone

The LMS color space is used as a procedure for adapting colors in an image or a sequence of images in order to improve the color perception of an observer with some sort of color blindness (see Figure 2.10).

Figure 2. 10 Normal vision and colorblind vision (inspired by Ku T., 2006)

The dichromat observers are missing one class of cone photopigment and confuse colors that differ with respect to the excitation of the missing class of photopigment only. Trichromat observers require color specifications from three components, whereas two components are sufficient to specify color for a dichromat. In these cases, a rule may be constructed to reduce any set of confused colors to a single three-component color specification (Anagnostopoulos et al., 2007).

2.4.3 CIELAB Color Model

CIE $L^*a^*b^*$ (CIELAB) was recommended in 1976 by the CIE (Commission International de l'Eclairage) international standard color specification system (Saleh et al., 2003). The CIELAB model was designed to more closely match human vision, and it is also perceptually more uniform than the RGB and the hue-saturation-value (HSV) color spaces (Liu et al., 2012). This color space is useful for colorimetric assessments of all kinds of objects or images. CIELAB is recognized for having the most optimal gamut of colors for the representation of graphic shapes (Hill et al., 1997). It is mainly used as a uniform color space derived from the tristimulus values of X, Y, Z with L^* , a^* and b^* coordinates. The L^* axis describes lightness, which ranges from black (0) to white (100), while the a^* axis represents red ($+a^*$) to green ($-a^*$), and the b^* axis represents yellow ($+b^*$) to blue ($-b^*$) (see Figure 2.11) (Yuan et al., 2007). These are two chromatic components (ranging from -120 to +120) (Viscarra et al., 2006; Yam and Papadakis, 2004).

Figure 2. 11 CIELAB coordinate system

On the other hand, the HSV color space is a popular choice for manipulating color that is defined in the HSV color space. This color space is developed to provide an intuitive representation of color and to approximate the way in which humans perceive and manipulate color. RGB to HSV is a nonlinear, but reversible, transformation. The hue (H) represents the dominant spectral component—color in its pure form, as in green, red or yellow. Adding white to the pure color changes the color: the color is more saturated the less white there is. This represents the saturation (S). The value (V) represents the color brightness.

This color space is represented by cylindrical coordinates, where the hue is represented as a degree of the angle, the saturation, as a distance to the black/white axis, and the value represents the height on the axis as a percentage. HSV is used in progressions of color. On this ground, it was not used within the model of saliency proposed in Chapter 4.

2.5 Shape Descriptors

The purpose of this section is to analyze the representation of the shape of an object, using computational methods called shape descriptors. We will then explain how a shape descriptor is used to characterize the shape of a 3D object from a virtual environment in Section 4.2.

Given the difficulty in characterizing saliency according to object shape, a wide-ranging study of the state of the art regarding shape descriptors was conducted in order to identify the best candidates for use in this research. This study, plus the proposal of a measure of saliency for object shape, is a significant contribution of this thesis. The results of this have been published in (Lara et al., 2015, 2016a).

The shape of an object expresses its appearance. Shape can convey ideas, as well as attract attention. Therefore, it is a salient feature. Humans are capable of identifying an object from its shape alone. Technological interest in implementing this human capability has enabled the extraction of shape-semantic information, usually through of a process of segmentation. However, the description of object shape is still a difficult task, with a number of limitations.

A shape descriptor can be defined as a mapping from the 2D or 3D object space to some high-dimensional vector space (Akgül, 2007). The main goal of shape description is to get

feature vectors or data structures with the maximum amount of shape information and the lowest possible dimensionality (Dos Santos, 2007), a suitable numerical representation (Laga et al., 2006), and the extracted numerical characteristics that describe the shape of an object. In this respect, Guo et al. (2015) considered that a good descriptor should be descriptive, compact, and robust. Vranić and Saupe (2002) defined four important criteria for a 3D object shape descriptor:

1. Invariance with respect to translation, rotation, scaling, and reflection of the 3D object;
2. Robustness with respect to level of detail;
3. Efficient feature extraction and search; and
4. Multiresolution feature representation.

These criteria were widely discussed in (Lara et al., 2016a).

Shape descriptors are a helpful mechanism for classifying, retrieving, clustering, matching and establishing similarities between objects. They play an important role in different areas such as computer-aided design/computer-aided manufacturing (CAD/CAM), virtual reality, entertainment, medicine, molecular biology, physical simulation, and e-commerce (Akgül, 2007; Bustos et al., 2005; Dos Santos, 2007; Dutağaci et al., 2005; Laga et al., 2006; Schmitt et al., 2015; Zhang et al., 2004) :

- In CAD/CAM, shape descriptors are especially applicable in physical anthropology. Physical anthropology plays an important role in industrial design, for example, clothing design or ergonomics (Paquet et al., 2000), as well as in the matching of solid models of 3D mechanical parts (Dos Santos, 2007). Local features can significantly improve the manufacturing costs, manufacturing process selection, production and functional parameters of 3D objects in the CAD tool field (Bespalov et al., 2006), as well as finding similarities, which is useful in furniture design (Suzuki et al., 2000) and image reconstruction (Mitsumoto et al., 1992).

- For virtual reality and entertainment, the use of 3D models improves the realism in film and video game production. In this industry, 3D objects can be reused and adapted based on their similarity to reduce production costs (Bustos et al., 2005).
- In medicine, object similarities are useful for detecting organ deformation. For example, they have been used in a specific part of the brain called the hippocampus to help diagnose diseases like epilepsy (Keim, 1999).
- In molecular biology, shape descriptors have been applied to analyze molecular surfaces (Ankerst et al., 1999; Venkatraman et al., 2009) and molecule stability (Chung, 1997).
- In physical simulation, (Novotni and Klein, 2001a) applied shape descriptors to find the best-fitting shoe for a given 3D foot scan, for example.
- In e-commerce, a customer can start with a few typical style options and then use a search engine to retrieve similar styles, in furniture shopping for example (Passalis et al., 2004).

Shape descriptor analysis has also been used for computer vision and texture analysis (Lazebnik and Ponce, 2005; Peleg et al., 1989) in order to represent articulated objects (Mateus, 2010), or to compute the similarity between object deformations (Taimouri and Hua, 2014). It has been applied in aerial images to distinguish and categorize areas such as parking, residential or schools (Zhang et al., 2013; Zhang et al., 2014), and fine-grained image recognition for example for insects (Zhang et al., 2015).

It was stressed above that, apart from size or color, object shape can be a salient feature. Likewise, shape is a useful resource for finding similarities between objects, independently of the class to which they belong, that is, two objects can be similar in shape, without belonging to the same class of objects. For example, an airplane and a bird are two completely different objects that are, however, very similar in shape.

2.5.1 Methods for representing shape

Zhang and Lu (2004) divided the methods for representing object shape into two categories: contour-based methods, which represent objects/shapes as a whole, and region-based methods, which represent segments/sections. Both categories have two subdivisions: structural and global methods. They include a large set of shape description techniques that are described later: chain code, polygon, B-spline, invariants, perimeter, compactness, eccentricity, shape signature, Hausdorff distance, Fourier descriptors, wavelet descriptors, scale space, autoregressive, elastic matching, area, Euler number, geometric moments, Zernike moments, pseudo-Zernike moments, Legendre moments, grid method, shape matrix, convex hull, media axis, and core. Ling and Jacobs (2007) pointed out two methods for object shape representation: one based on local features extraction and the other based on the extraction of global features. Tangelder and Veltkamp (2008) also introduced a set of shape representation methods based on the volume and surface of the 3D models, identified as implicit surfaces, constructive solid geometry (CSG), binary space partitioning (BSP) trees, octrees, boundary representation (B-rep) and free-form surfaces.

These methods were organized into different categories of shape descriptors, which were then integrated into the different taxonomies proposed. There follows a description of the six taxonomies proposed for different purposes.

2.5.2 Taxonomies of shape descriptors

There is no universally accepted method for building a shape descriptor taxonomy. Furthermore, each descriptor taxonomy serves a different purpose and addresses a different perspective, albeit with the same goal of identifying object shape. There are taxonomies that include shape descriptors for 2D image and 3D objects.

Zhang L. et al., (2004) proposed a classification divided into three categories: (1) feature-based shape descriptors, (2) graph-based descriptors, and (3) other methods. Their taxonomy is based on the most popular shape descriptors for 3D object classification and retrieval. In particular, this taxonomy considers the spatial partition and the representation of the features

of the 3D models, where the 3D shape can be discriminated by its geometric features and topological properties. Zhang L. et al., (2004) discriminated the shape by measuring and comparing their features. Furthermore, their taxonomy targets the design of 3D object space methods to keep all possible information on an object in a low-dimensional vector. Five sub-categories were compared based on the following criteria: original shape features, spatial partition methods, pose normalization, transformation invariance, and advantages and disadvantages.

Bustos et al. (2005) divided the shape descriptors into five categories: (1) statistics, (2) extension-based, (3) volume-based, (4) surface geometry and (5) image-based methods. This taxonomy targets the retrieval of shapes across 3D objects. They also conducted a qualitative comparison of some of the proposed shape descriptors based on the technical description published in the literature using the following criteria: dimension, invariance, object representation, object consistency and metric (the measure of similarity).

In his doctoral thesis, Akgül (2007) presented a taxonomy for shape descriptors divided into five categories: (1) histogram-based, (2) transform-based, (3) graph-based, (4) 2D image-based and (5) other methods. He focused on a general approach based on the geometric or topological information contained in the 3D object. He also considered similarity for object retrieval and compared the retrieval performance resulting from the fusion of two descriptors against other well-known 3D shape descriptors.

Also in his doctoral thesis, Dos Santos (2007) presented a five-category shape descriptor classification: (1) histogram-based, (2) transform-based, (3) graph-based, (4) 2D image-based and (5) other methods. This taxonomy is very similar to the proposal by Bustos et al. (2005), although Dos Santos compared the behavior of some shape descriptors in order to identify the most suitable options for his thesis. He developed a prototype to compute 3D shape descriptors and evaluate shape-matching performance based on these descriptors and some of their combinations, rating the accuracy and general performance of the 3D models retrieved in queries.

Tangelder et al.'s taxonomy (2008) is organized into three main groups: (1) feature-based, (2) graph-based, and (3) geometry-based methods. This taxonomy focuses on the use of matching methods for content retrieval based on the 3D shape considering the surface and volume of the 3D models. They compared the matching methods according to the following criteria: shape model, triangle inequality, efficiency, discriminative power, partial matching, robustness, and pose normalization requirement.

EINaghy et al. (2013) proposed a taxonomy, again with five categories, as follows: (1) view-based, (2) graph-based, (3) geometry-based, (4) statistics-based and (5) general methods. Like the above four, the purpose of this taxonomy is to recover 3D objects based on similarities. Their comparison was based on 3D object representation requirements, efficiency, discriminative power, partial matching, robustness and sensitivity, and pose normalization. This comparison is very similar to the one proposed by (Tangelder and Velkamp, 2008)

Table 2.3 presents the nine descriptor categories found in these six taxonomies (Akgül, 2007; Bustos et al., 2005; Dos Santos, 2007; EINaghy et al., 2013; Tangelder and Velkamp, 2008; Zhang et al., 2004). The categories are histogram-based, transform-based, graph-based, 2D image-based, feature-based, geometry-based, extension-based, volume-based, and other methods. The histogram-based, 2D image-based, geometry-based, and other method categories are similar to the statistics-based, view-based, surface geometry and general methods, respectively. Therefore, they have been grouped into just four categories and are highlighted in Tables 2.3 with two asterisks between parentheses (**).

Note that some descriptors included in these taxonomies use the transformation of a 3D object into a set of 2D images.

Table 2. 3 Categories of shape descriptors

Category	Proposed by
(**) Histogram-Based or Statistics Based	Bustos et al. (2005) Akgül (2007) Dos Santos (2007) EINaghy et al. (2013)
Transform-Based	Akgül (2007) Dos Santos (2007)
Graph-Based	Zhang L. et al., (2004) Akgül (2007) Dos Santos (2007) Tangelder et al. (2008) EINaghy et al. (2013)
(**) 2D Image-Based or View-Based	Bustos et al. (2005) Akgül (2007) Dos Santos (2007) EINaghy et al. (2013)
Feature-Based	Zhang L. et al. (2004) Akgül (2007) Tangelder et al. (2008)
(**) Geometry-Based or Surface Geometry	Tangelder et al. (2008) EINaghy et al. (2013)
Extension-Based	Bustos et al. (2005)

Volume-Based	Bustos et al. (2005)
(**) Other Methods or General	Zhang L. et al. (2004) Dos Santos (2007) EINaghy et al. (2013)

A description of the nine categories of shape descriptors that were recovered of from the six taxonomies analyzed and compared in by (Lara et al., 2016a).

- **Histogram-Based Descriptors**

This category includes all the shape descriptors that adopt a histogram, even if they are not used in a rigorous statistical sense. With regard to shape descriptors, the histogram is typically an accumulator or container that collects the numerical values of certain features calculated from the shape representation (Akgül, 2007; Dos Santos, 2007); it maintains the neighboring points or their properties (Behley et al., 2012). In this category, descriptors are partitions of certain spaces in a 3D model, where the complete space is decomposed into disjoint cells that correspond to the histogram bins (Huang et al., 2007). These shape descriptors describe the distribution of points in the model across all rays from the origin (Kazhdan et al., 2003b). Histogram-based descriptors have been widely used in computer vision tasks, such as matching, image retrieval (Lowe, 2004; Mikołajczyk and Schmid, 2006; Mortensen et al., 2005), and texture analysis (Lazebnik and Ponce, 2005; Peleg et al., 1989).

- **Transform-Based Descriptors**

These descriptors capture the surface points on a 3D voxel or spherical grid by means of a scalar-valued function which is processed by transformation tools such as the 3D Fourier transform, angular radial transform, spherical trace transform, spherical harmonics or wavelets (Akgül et al., 2009; Zhao and Lu, 2013). An important advantage of the transform-based methods is descriptor compaction due to the fact that the feature vector retains only a few transform coefficients. Furthermore, these descriptors consider the invariance that can be achieved by omitting the transform coefficient phase at the expense of some extra shape

information (Zhao and Lu, 2013). In other words, the shapes are described in a transformation invariant manner, so that any transformation of a shape will be described in the same way. This yields the best measure of similarity in any transformation (Kazhdan et al., 2003b).

- **Graph-Based Descriptors**

Graph-based descriptors aim to get a geometric meaning of the shape of a 3D object using a graph to project how the components of the shape are interconnected (Tangelder and Velkamp, 2008). These descriptors are considered to be more complex and sophisticated than those based on feature vectors. Also, they have the advantage of more accurately encoding the properties of the geometric shape of the object. The descriptors use spectral graph theory tools. The information contained in a graph can be represented as numerical descriptions (Akgül, 2007; Akgül et al., 2009). The graph-based descriptors have the advantage of reducing the problem of shape dissimilarity through graph comparison. These descriptors are used especially for retrieving articulated objects (Laga et al., 2006).

- **2D Image-Based Descriptors**

This category of descriptors represents and compares the shape of a 3D object as a collection of its 2D projections taken from different viewpoints. A standard descriptor for 2D images, like Fourier descriptors and Zernike moments, is considered in order to describe each projection (Laga et al., 2006). The descriptors of this category are designed for similarity methods. Multiple images of a 3D object are captured from several positions with a camera and are stored in a database. The images are processed to find the similarity between the views of the query object and the models in the database (EINaghy et al., 2013; Huang et al., 2007). A particular characteristic of these shape descriptors is that they are a summary of the values of the pixels of a digital image, containing information on the silhouette of an object. Therefore, the shape descriptor is represented with a vector that contains the number of parameters derived in this manner (Latecki et al., 2000).

- **Feature-Based Descriptors**

This category of descriptors was proposed by Zhang L. et al., (2004) (Zhang et al., 2004) and Tangelder et al. (2008) (Tangelder and Velkamp, 2008). This category refers primarily to

descriptors of global and local features, which were the point of reference for the development of many other descriptors grouped in other categories. The descriptors in this category express the geometric and topological properties of the shape of each 3D model. The shape of an object is discriminated by measuring and comparing its features. These descriptor methods aim to represent the shape of a 3D object with the implementation of a compact vector. A simple way to do this is by using functions defined on the unit sphere (Laga et al., 2006). Feature-based descriptors extract the features of the 3D model in a fast and simple way (Bustos et al., 2005).

- **Geometry-Based Descriptors**

Geometry is always specified in 3D models in contrast with other application-dependent features (Bustos et al., 2005). The geometry features usually used to describe the 3D model are volume, surface area or curvature, and ratios, like the surface area to volume ratio; compactness, that is, the non-dimensional ratio of the volume squared over the cube of the surface area; crinkliness, that is, the surface area divided by a model made up of the surface area of a sphere with the same volume as the 3D object; convex hull features; the bounding box aspect ratio; or Euler numbers (EINaghy et al., 2013).

- **Extension-Based Descriptors**

These descriptors are created from samples of features taken along certain spatial directions with a starting point in the center of the object (Bustos et al., 2005). The 3D object is usually treated as functions defined on spheres and described in terms of samples taken from these functions (Barrios and Bustos, 2011).

- **Volume-Based Descriptors**

This category includes all descriptors that represent the shape of a solid object through the volumetric representation obtained from the surface of a voxelized object (Bustos et al., 2005). This representation is computationally expensive, and its accuracy depends on the size of the voxel.

- Other Methods or General

These descriptors usually serve the purpose of improving the retrieval process through integration with other retrieval-oriented 3D object descriptors.

2.6 User Modeling

Frias-Martinez et al. (2006) considered that the key element of a personalized environment is the user model. A user model is a data structure that represents user interests, goals and behaviors. A user model is created by means of a user modeling process in which unobservable information about a user is inferred from observable information about that user. Furthermore, they noted that the user model can be created using a user-driven approach, in which the models are created directly from the information provided by the actual user, or an automatic approach, in which the user is shielded from user model creation process.

According to (Barla, 2010), user modeling "represents all kinds of information related to user and user's context, which are required in order to provide personalized user experience. It can hold various features of the user such as: age, education level, interests, preferences, knowledge etc., or can represent the overall context of the user's work, including platform, bandwidth or location".

Modeling computer system user characteristics and attitudes has been a major challenge and a topic of special interest in the area of computing for more than two decades.

According to (Kobsa, 2001), a user model is a set of information structures designed to represent one or more of the following data:

- Assumptions about one or more types of user characteristics in models of individual users (e.g., assumptions about their knowledge, misconceptions, goals, plans, preferences, tasks and abilities).

- Key common characteristics of users pertaining to specific user subgroups of the application system (the so-called stereotypes).

User modeling acquires the interests of users, their personality, their previous knowledge, their cognitive abilities, their beliefs, their learning preferences and progress. This is a means of generating and updating the user stereotype. According to Fischer (2000), a user model is the representation of the user features with a view to decision making on computer system interaction. He then describes a number of papers, all of which seek to represent individual user characteristics in different contexts.

Peña (2007) presented a student model based on cognitive maps in his doctoral thesis.

The system analyzes students in order to ascertain their interests, skills and attributes, and thus create a mental map. The mental map (also called cognitive map) has the function of representing the student's ideas and their interrelationships.

The user model (in this case, of the student) takes into account their profile, preferences, skills and their evolution over time. Modeling is a two-part process:

- (1) The concepts that integrate the mental image are represented in a cognitive map.
- (2) These concepts are defined in an ontology.

The ontology represents the student's knowledge domains. Seven domains were defined: three for student characterization (cognition, personality and preferences), three more for experience characterization (sequence, content and evaluation), and another one for student knowledge acquisition management.

Ferreira et al. (2010) represented user profiles using ontologies with fuzzy logic. Learning platform users can publish, create and collect specified content. The system constructs an ontology following a five-phase process (see Figure 2.12).

Figure 2. 12 System process

The linguistic preprocessing phase is responsible for collecting and converting the different files uploaded by the user to a common internal format for proper processing in the following steps. The term indexation phase adds a small index later used to build the ontology. This index stores an identifier associated with each term and a list of references to the respective term for each entry. The potentially most interesting terms are selected from the above collection in the key term extraction phase. The user ontology as such is constructed in the following phase. The ontology is considered as a set of related trees where each node represents a topic. The idea is that the resulting tree is a taxonomic representation of previously filtered concepts. The user profile is updated in the last phase. Here new data are added to the ontology as contents published by each user. To do this, the added documents are processed as above, and the ontology is later modified based on the resulting information.

Gascueña et al. (2010) presented a student-centered model using two ontologies that work simultaneously. One ontology stores learning material, and the other contains the student profile. The main purpose of this model is to reuse learning materials and adapt teaching

methods based on the user. An agent operates on both ontologies, selecting and showing students the items of the educational material that better match their profile.

The student ontology organizes data around a class called "student". This class contains student data like visual preferences, devices, personal data, courses in which the student is enrolled and connection speed. The idea of this organization is to adapt learning as closely as possible to the student.

2.7 Data ontologies

Data ontologies have been widely used in recent years. A data ontology is defined as a concrete representation of knowledge. Likewise, a data ontology is a specification of a conceptualization. Data ontologies are represented as an organized knowledge structure. A data ontology represents domain knowledge by means of declarative formalisms, that is, defines the terms and the basic relationships of a set of objects in order to understand an area of knowledge. This represents the universe of discourse. Ontologies are also like database system conceptual schemata. A conceptual schema provides a logical description of shared data, which enables application programs and databases to interoperate without having to share data structures. While a conceptual schema defines relations with respect to *data*, an ontology defines the terms used to represent *knowledge*. Ontologies contain additional information on how to make inferences about data, that is, how to establish formal axioms that may then be applied in different domains to process the stored knowledge (Gruber, 1993). According to the Resource Description Framework (RDF) language, an ontology is composed of three types of elements:

- (1) Classes or concepts: the main components of ontologies. Classes represent the concepts of the domain of discourse. A class can contain subclasses. Subclasses represent more specific concepts than their super-class (Noy and McGuinness, 2005). For example, "Boss" or "Employee" in corporate knowledge. In ontologies, classes are treated as unary relations to help unify object- and relation-centered representation styles (Gruber, 1993).

- (2) Slots, roles or properties: descriptions of various features and attributes of the concept. Properties can be the relations or predicates that are established between the classes (object properties) or actual properties (data properties). An example of an object property is "isEmployeeOf", whereas a data property might be "age".
- (3) Instances: specification of the above abstractions for definite known real-world. "John" is a possible example of the "Boss" class instance. This is represented with triplets in RDF.

Figure 2.13 illustrates the above three concepts.

Figure 2. 13 Example of RDF elements within a data ontology for knowledge representation

2.8 Conclusion

The described in this chapter represents a conceptual scenario composed of different areas of research, such as virtual reality, spatial language, frames of reference, perceptual saliency, color space models, descriptors of shape, user modeling, and data ontologies. These fields of knowledge form the theoretical groundwork of a computational model for the generation of directions for object location in virtual environments consider

Chapter 3: State of the Art

3 State of the Art

This chapter provides an overview of work directly related to the proposal of this thesis. It presents several computational models of perceptual saliency and the purpose for which they were developed. Similarly, computer systems for the generation of natural language directions for object location are analyzed.

Section 3.1 details a set of computational models of perceptive saliency developed in the last few years.

3.1 Computational Models of Perceptual Saliency

The human visual system is capable of detecting the visual saliency of an object or a set of objects extremely rapidly and reliably, and the computational modeling of this basic intelligent behavior still poses a great challenge (Hou and Zhang, 2007). Nevertheless, as Kadir et al. (2001) mentioned, the fields of cognitive psychology and computer vision have proposed numerous models of human visual saliency (sometimes also called visual search or attention). The following is an overview of the proposed computational models of perceptual saliency.

- Gapp (1995) (Gapp, 1995) presented a computational model for selecting reference objects for the object location process. This model is based on the physical characteristics of the object and the context in which the object is located. The model is capable of measuring the Euclidean distance between a vector of scaled attributes derived from the feature variables and the attribute vector of an optimum reference for all candidate reference objects. The object with the minimum distance is the best reference.

Likewise, Gapp presents a study analyzing the most relevant aspects of an object, such as its saliency with respect to color, size, shape and orientation, functional dependencies, prior knowledge, mobility and distance, all of which may cause the object to stand out. Furthermore, he describes aspects of the context surrounding the object. The model selects a reference object based on its visual saliency. To calculate this visual saliency,

Gapp suggested that each object should be given a graduated rating in the interval $[0,1]$, according to the evaluated aspects of the object.

Projects like CITYTOUR and VITRA described in Section 3.2 can be considered the groundwork of this model for supporting the generation of human acceptable descriptions of image sequences and routes.

- Itti et al. (1998) proposed a computational model of saliency based on the behavior and the neuronal architecture of the primate visual system. It is useful for selecting locations in order of decreasing saliency through a dynamic neural network. A combination of the features —size, intensity, color and orientation— of a multiscale image are used to create a topographical saliency map. This map represents the saliency of each location in the visual field by means of a scalar quantity that drives the selection of visited locations based on the spatial distribution of saliency.

Itti et al. used a set of static color images that are digitized with a resolution of 640 x 480 as input. Nine spatial scales are obtained through Gaussian pyramids, which progressively low-pass filter and subsample the input image, yielding horizontal and vertical image-reduction factors ranging from 1:1 (scale zero) to 1:256 (scale eight) in eight octaves. Each feature is measured by a set of linear center-surround operations akin to visual receptive fields.

The resulting maps are combined to output three conspicuity maps — $\bar{I}$ for intensity, $\bar{C}$ for color, and $\bar{O}$ for orientation— on the scale ($\sigma = 4$) of the saliency map. This means that the maps are obtained by means of a scale addition, which consists of reducing each map to scale four and point-by-point addition.

These three conspicuity maps are created to check that similar features compete strongly for saliency, while different modalities contribute independently to the saliency map. Finally, the three conspicuity maps are normalized and summed into the final saliency map input S .

The principal goal of this model was to provide a parallel method for the fast selection of a small number of interesting image locations to be analyzed by more complex and time-consuming object-recognition processes.

- Klippel et al. (2005) presented a model of structural saliency of the objects and landmarks along a route. They combined the formalization of object saliency and the conceptualization of the wayfinding action. The development of this model of saliency contributed to the cognitive groundwork for supporting navigation and formalizing geosemantics.

The model provides a measure of saliency for all identified objects within a street network. The measures can be used to choose the most salient objects along a specific route. Furthermore, this model can provide support at the decision points when there is a change of direction in order to enhance route directions. If these are measures of static objects, they can be stored as their parameters.

Likewise, the model considers three visual, semantic and structural aspects. Itti et al. use a quality value for representing a normalized saliency measure with values from 0 to 1. This value is obtained from the visual saliency (vs), semantic saliency (sm), and structural saliency (ss). The three measures of saliency can be combined into a weighted average to represent a joint saliency.

Structural saliency was considered, but not developed. The visual and semantic saliencies were determined by the visual comparison of semantic properties of the objects with neighboring objects.

- Itti (2006) presented a computational model for the detection of saliency, which addressed three aspects: color, intensity and orientation. The model can be embedded in a framework to measure and compare data on human eye movement. The framework was developed under a very realistic simulation to determine the transition of the human eye fixation to the properties of a local image. It therefore evaluates a measure of perceptual saliency.

Itti aimed to operationally define the perceptual saliency based not only on local image properties but also on how they are captured by the retina and how their cortical representations interact over visual space and time. Itti wanted to find out what part of an image a human being focuses on.

The model was tested with 15 video clips of natural scenes rather than static imagery. The model considers three stages. The first stage outputs the set of saliency maps from the video frames that are shown to human observers. The human eye movement recordings are then used to compare the model-predicted saliency at human eye positions compared to random locations within the video frames. The second stage considers a foveation filter by means of which each input frame becomes increasingly blurred the further it is away from the current human eye position before it is processed by the model. Finally, the third stage incorporates the video frames into a background photograph of the experimental room and computer monitor, shifts the resulting image to focus on human eye position, crops the shifted image to simulate a retinal field of view, and applies a foveation filter to the field of view before it is processed by the model.

The model proposes a measure of absolute saliency composed of a human scan path for comparison with the same measure compounded over a random scan path. This measure is difficult to interpret because of all the factors that influence eye movements.

- Hou et al. (2007) developed a method for detecting visual saliency that it is not related to the features, categories, or other forms of prior knowledge of the objects. To do this, their proposal is to analyze the log-spectrum of an input image by extracting the spectral residual of an image and then using a fast method to construct the corresponding saliency map in the spatial domain. They determine two factors from the generated the saliency map: the hit rate (HR) and the false alarm rate (FAR). Their saliency map was based on the model proposed by (Itti, 2006). They compute the value of FAR and HR by modifying the saliency map output according to Itti's method (1998).

The images used to test the model were also a 320 x 240 sample from Itti's method. The images were natural pictures and artificial images, such as psychological patterns. Such patterns are used in attention experiments, whose aim is to explore the mechanisms of visual attention search and test the effectiveness of the saliency map. The results suggested that the method is fast and robust. Even though this method is limited to the analysis of static images, the saliency map can be calculated for each frame of sequences, as in a video, without regard to continuity.

- Caduff et al. (2008) presented a framework for evaluating the saliency of spatial features: this model targets the navigation domain. They represented a vector of saliency as a trilateral relationship between the observer, the observed geographic feature and the environment (see Figure 3.1) in terms of perceptual, cognitive and contextual saliency. Their model is based on visual sensory input and theories of visual attention as they consider vision to be the most important sensory input for discriminating salient features for navigation. With the support of a Bayesian network, the interdependency between components is analyzed for their integration into one computational model.

Figure 3. 1 The assessment of the saliency of spatial or geographic features based on the trilateral relationship between observer, environment, and geographic feature (inspired by Caduff et al., 2008)

The model has the advantage of having been designed with flexibility and adaptability in mind. Additionally, the system provides automatic generation of route instructions based on (geo-) metric properties with a Bayesian network that requires an assessment of the spatial features available in the surrounding environment. This model of saliency has

no a priori knowledge of the semantics of spatial features, but has a structured framework that provides for the incorporation of additional factors more easily.

First they conceptualized saliency and introduced the terms *perceptual saliency*, *cognitive saliency*, and *contextual saliency*. They constitute a vector of saliency.

a) *Perceptual saliency*: They presented a perceptual saliency model with a bottom-up guidance of attention which is derived from the part of the environment that is perceived by the navigator from one specific position. Bottom-up guidance is a process of searching for a conjunction of features using an attention capture method (Proulx, 2007). They analyzed the continuous stream of stimuli based on a series of criteria (e.g., auditory, olfactory). They also examined a snapshot of the visual stream of stimuli, and they noted that the restriction of the analysis to one stream of stimuli does not affect the basic assumption of the framework. The restriction is derived from the results of spatial cognition and psychology, specifying that the main factor for identifying landmarks in the context of navigation for people who are not blind is the visual stream.

The principal goal for the attention-based assessment of landmarks is the simple hypothesis that landmarks attract attention. Here they specified two dominant divisions of visual attention theories that investigate this hypothesis. The first theory is based on Treisman's model of space- or location-based attention and the second is the developing object-based attention theory.

The main difference between these two divisions is that location-based attention and object-based attention use different fundamental units of attention. The focus of location-based attention is on continuous spatial areas of the visual field, while the theory of object-based attention holds that visual attention can directly select discrete objects. In addition to location- and object-based attention, research has shown that attention is also dependent on the concept of the scene, which defines the structure and global semantic characteristics of the scene.

b) Cognitive Saliency: Cognition refers to an individual's mental processes. However, this does not apply in the context of navigation insofar as the mind has an internal representation of the spatial environment and objects are retrieved from a representation based on the degree of recognition and the idiosyncratic relevance of individual objects. The degree of recognition measures how well an object can be identified by an observation, while the idiosyncratic relevance indicates the object's personal importance to the observer. They assume that objects with a high degree of recognition are more likely to be used as points of reference than objects with a low recognition value. Likewise, they also assume that well-known objects are preferred over unknown objects.

In the process of reasoning on the saliency of spatial objects, stored instances of mental objects are taken into account depending on the degree of recognition and idiosyncratic relevance. Recognition is achieved when some pattern or object recurs. There is recognition if the current observation matches the previously stored attributes of that spatial object and vice versa. For the purposes of recognition, a spatial object must be familiar in the sense that it must be linked to at least one observation. The degree of recognition and familiarity may be different. This recognition consists of a single observation and a description obtained from a stored instance of a mental spatial object. This recognition is considered as a measure of the degree to which observations from specific points of view support the identification of previously observed objects.

c) Contextual Saliency: They consider that context plays an important role during navigation because it defines how much attention can be paid to the recognition and assessment of potential landmarks. In their framework, they distinguish between two types of context: (1) task-based context, which includes the type of task to be performed in the assessment process, and (2) modality-based context, which considers the mode of transportation and the amount of resources that need to be allocated for discriminating potential landmarks. Also, in this framework, they consider that navigation itself is the task based on which the saliency of spatial objects should be assessed.

These three component types (perceptual, cognitive, and contextual) define the vector of saliency for expressing the overall potential of a spatial feature for attracting navigator attention. Furthermore, a set of auxiliary vector components capture the important aspects of saliency in terms of attention (like location- and object-based attention, scene context, degree of recognition, and idiosyncratic relevance), and a set of low-level components (contrast, size, distance, etc.) that provide support.

This model was analyzed in depth because it includes aspects like perceptual saliency, contextual saliency and the color of objects, which were of special relevance for the model of saliency that is proposed in Chapter 4.

- Guo et al. (2008) proposed a method called PQFT (phase spectrum of quaternion Fourier transform) to obtain a spatiotemporal saliency map. This map considers features like color and spatial saliency for orientation in a single frame, and temporal features between frames for motion to detect salient objects in natural images or videos. These features are represented as a quaternion image. This means that they are processed in parallel. When compared with images stained by white-colored noise, PQFT performance is robust to noise (such as color) because it does not conflict with the main feature of the salient object. Furthermore, it does not depend on parameters and prior knowledge.
- Li et al. (2011) proposed a model of saliency based on local and global aspects of an image. Their proposal considers: (1) the detection of the saliency of both small and large regions, where the size of salient regions is considered to vary in quantity; (2) the detection of saliency in cluttered scenes, where the information based on local saliency models with strongly textured regions may stand out from the rest; and (3) the inhibition of repeated patterns. It is believed that all the objects of the scenes seen by the human visual system compete with each other to focus attention selectively.

In this proposal, they modeled salient pixels and regions locally. To do this, they used independent components of natural scenes as the center-surround filters, where difference of Gaussian (DOG) and Gaber were used to measure the local contrast. They then used

the entropy to assign a weight to each map in order to obtain a number of response maps. Finally, a single saliency map is obtained from a weighted sum.

They demonstrated through different experiments that the proposed model has the ability to highlight both small and large prominent regions to inhibit pattern repetition in the cluttered scenes. Table 3.1 presents a summary of the aspects that distinguish each of the analyzed models.

Author and year	Used in	Visual features considered	Other aspects considered
Klaus-Peter Gapp (1985)	- 2D images	<ul style="list-style-type: none"> - color - size - shape and - orientation 	<ul style="list-style-type: none"> - functional dependencies - prior knowledge - mobility - distance - the context
Laurent Itti, Christof Koch, and Ernst Niebur (1998)	- 2D images	<ul style="list-style-type: none"> - size - intensity - color and - orientation 	<ul style="list-style-type: none"> - the behavior of the primate visual system - an individual normalization
Alexander Klippel and Stephan Winter (2005)	- 2D images	- Not specified	<ul style="list-style-type: none"> - the characterization of spatial structures - adaptation to context
Laurent Itti (2006)	- 2D images	<ul style="list-style-type: none"> - color - shape and - orientation 	<ul style="list-style-type: none"> - use in video clips of natural scenes
Xiaodi Hou and Liqing Zhang (2007)	- 2D images	- Not specified	<ul style="list-style-type: none"> - analysis of the log-spectrum of an input image - use in video
David Caduff and Sabine Timpf (2007)	- 2D images	<ul style="list-style-type: none"> - color - intensity - orientation - contrast 	<ul style="list-style-type: none"> - analyzed perceptual saliency, cognitive saliency and contextual saliency
Chenlei Guo, Qi Ma and Liming Zhang (2008)	- 2D images	<ul style="list-style-type: none"> - intensity - color - motion 	<ul style="list-style-type: none"> - salient areas in natural scene and video
Jian Li, Martin D. Levine, Xiangjing An and Hangen He (2011)	- 2D images	<ul style="list-style-type: none"> - size - color - shape 	<ul style="list-style-type: none"> - identification of both small and large saliency regions - identification of saliency in cluttered scenes - inhibition of repeating patterns

Table 3. 1 Examined aspects of the analyzed models of saliency

These models analyzed only static components or objects in images, maps or scenes in order to determine the most salient parts and calculate saliency maps. Moving objects can also be detected within a video in order to determine where to direct the focus of attention. Most systems focus on summarizing the properties of a target object rather than a reference object. Other models measure the saliency of a reference object in order to map routes to destinations rather than for locating objects in virtual environments.

This analysis highlighted that none of the above models meet the needs of the target framework, that is, none focus on the aspects that this computational model of saliency is expected to deal with. For this reason, Chapter 4 proposes a new model specifically focusing on the basic features of an object from the perspective of an individual and of its context.

3.2 Computerized spatial language generation for object location

A key aspect of spatial language has always been the generation of references using appropriate linguistic expressions in order to facilitate the location of objects (Barclay, 2010), that is, from a practical point of view, a series of principles intuitively originated in the speaker and then added at the time of the generation of the descriptions of spatial relations. These principles are related to the association of the spatial relations and the representation of linguistic expressions. In this sense, they describe important aspects of spatial language systems.

3.2.1 Spatial language systems

Spatial language systems are mainly intended for the artificial generation of language and not for its interpretation. However, many of the algorithms developed for computer systems can be used for both purposes. Like spatial languages, the generation of spatial language systems requires the combination of a spatial reference frame and linguistic expressions. As mentioned in Chapter 2 (see Section 2.2 - The use of object as a resource for the frame of reference), a possible resource used by the spatial reference frame is selected objects according to perceptual saliency. This is a far from straightforward task and has a number of limitations, such as:

- Absolute references may be ambiguous with respect to relative references. In some cases it is difficult to distinguish between the behavior of the absolute reference frame and its behavior in terms of a relative frame of reference (Pederson et al., 1998). Therefore, relative references can be misinterpreted (Levinson, 2003; Moratz and Tenbrik, 2006). To solve this problem, priority schemes have been designed with alternative references and their own linguistic expressions (Moratz and Tenbrik, 2006).
- A similar problem with regard to the spatial language systems is the ambiguity of the linguistic expressions used to understand the context and generate directions. Ambiguity can be generated when describing one of the possible target objects available to use as a reference. This situation can render the linguistic expressions redundant or imprecise, generating a high computational cost in object description, search and location processes (Stoia, 2007).
- Another potential problem is when the use of reference objects is not taken into account, or when there are several candidates because there is no way of knowing why one object is preferred to others in the described scenes (Barclay, 2010).

Conceptual schemes have been introduced to deal with some of these limitations. These conceptual schemes consist in the design of spatial language that is as close as possible to natural language and has a planned syntax using common semantics with a logical formalism; a solution likely to be applied in virtual environments (Trinh, 2013). In spatial language systems, it is important to generate linguistic expressions with an efficient description, taking advantage of the semantic rules of the language used in order to model the world (Stoia, 2007). To do this, it is necessary to consider semantic theory in order to understand the multiple meanings of words and sentences and be able to generate multiple interpretations, from which the listener will pick the best according to the surrounding environment (Winograd, 1971). Also called visual languages, spatial language systems have been simplified by graphical editors. These editors seek to express spatial relations in a conceptual, formal, and flexible way (Trinh, 2013), which has led to the identification of linguistic

expressions as a key element for the implementation of systems with the generation of spatial language.

Despite limitations, a number of applications have been developed, including, for example, descriptions of scenes from video input for blind people, the generation of automatic comments and the search for images or objects (Barclay, 2010; Kelleher and Costello, 2009), where virtual environments play an important role. Based on Barclay (Barclay, 2010), are now describe several applications with spatial language systems in different research areas:

- *Graphic design and drawing programs.* It might be complex to visually interpret the design of complex 3D graphics with a particular sequence development, but a language system could help by providing advice and an incentive for the creation of this type of representations.
- *Computer Games.* Video game systems were one of the first areas to generate spatial languages because of the need for realistic software agents and their interaction with humans.
- *Navigation aids.* The use of wayfinding instructions within these software solutions helps users to navigate better, possibly by providing auditory cues, such as "turn left and walk 100 meters" or even more specific expressions like "to your right, walk 5 meters, and turn left at an angle of 45°".
- *Robot systems.* This is certainly the easiest type of application to perceive. Like human beings, a robot can be guided by audio messages or wayfinding instructions. Work developed in the robotics field has relied heavily on spatial language systems. Also, natural language is a key resource for interaction between robots and humans, that is, for instructing and programming robots in localization tasks. An important aspect in this scenario is to establish the spatial references to the target, recognizing scene descriptions so that the robot can identify objects or sites. To do this, however, it is necessary to solve two problems: robot perception and the ambiguity of human language (Moratz et al., 2003).

- *Training simulators.* The key to the information required by these applications is the type of support they provide for the users. Imagine, for example, a virtual training exercise, with an online instructor providing a soldier with the support required for locating a wounded comrade in a danger zone.
- *Geographic information system interfaces.* This type of tools is used to map and analyze events that occur in geographic areas, and they give local map references in response to specific queries.

3.2.2 Computer systems with spatial language

In this section, is analyzed a set of computational systems that integrate spatial language. Also is discussed the progress in recent years in this research line. A systematic search was conducted into the Scopus bibliographic database using different search string criteria in order to select the computational systems. The dynamics of the search was as follows:

1. A pilot run was conducted using the search string "Spatial language", which returned far too many results at 9,391.
2. The search was refined in the second run using the expression "Computer systems" and "spatial language", joining the two criteria with the "and" operator. The number of results was reduced to 1,785, which was again too high.
3. The string used in the third run was "Spatial language generation systems", which reduced the number of papers to 208. The resulting papers were carefully reviewed, considering the title, the abstract and the keywords in order to select the papers related to spatial reference frames, spatial language, object location and perceptual salience in computer systems. Based on quality and relevance, the number of papers was reduced to 121; however, the number of results was still considered too high, and we decided to implement a couple more searches.

4. The fourth run used the search string "Computational systems for the generation of spatial language". This search returned 26 results, of which 12 were among the 121 articles output by the previous search.
5. We added the criterion "Virtual environments" to the final search string in order to examine how many systems have been implemented considering the use of virtual environments. The string used was "Systems for the generation of spatial language in virtual environments", it generated only six results and none of them matched previous searches.

The final criteria for selecting computer systems were: the use of real or virtual (2D or 3D) environments, object location, the use of spatial reference frames and reference objects, the implementation of a perceptual saliency model and the use of spatial natural language. From the results of the last three searches, 11 of the papers that covered one or more of the topics of interest (i.e., spatial language, frames of references, object or place location) were selected. Subsequently, we searched the references cited by the selected papers, and another four papers with the same line of research were also chosen because they considered several of the topics analyzed in this review. Again based on some of the cited papers and on the web, five doctoral theses addressing this line of research were identified. This resulted in a total of 20 documents. Of these 20 documents, only papers related to object location were taken into account, that is to say, documents addressing only place localization were discarded. The discarded papers are: GRADD system by Moulin and Kettani (1999), Automatic landmark detection systems by Elias and Brenner (2005) and GRE (Generation of Referring Expressions System) by Stoia (2007). After applying the inclusion and exclusion criteria to computer systems, we selected 17 documents. During the analysis of the documents, however, we found that the VITRA (Visual Translator) System had been extended with a 3D visual integration, so we decided to include one more document. The final list comprises 18 systems.

Table 3.2 singles out three important aspects of the selected documents. Column 1 lists the name and year of publication. Column 2 gives the authors and the paper reference, as well as some authors that have cited the project. This is interesting because it provides an

understanding of the maturity and the impact of the project on the research community. Column 3 gives a general description of the key aspects of the project, like its organizational structure, the programming language for development, advantages and disadvantages, etc. Note that projects like Barclay's Reference Object Choice in Spatial Language: Machine and Human Models (2010) and Thanh-Hai Trinh's Moscarel (2013) were not cited by other projects, but were included because they cover several of the topics of interest in this review.

Table 3. 2 Computational systems applying spatial language

Name and year	Developed by	
	Cited by	Description
SHRDLU (1971)	Terry Winograd (at the MIT Artificial Intelligence Laboratory) (Winograd, 1971)	<p>A system for the computer understanding of English. The system answers questions, executes commands, and accepts information in natural language. Users could carry on a simple conversation about a small number of objects, blocks of a real-world scene. It was an artificial intelligence project with:</p> <ul style="list-style-type: none"> • Syntactic and semantic processing of English. • A database with logical rules for the identification of objects. • Prepositions with a spatial description (e.g., left, front, back). • Based on the centroid of objects through cardinal axes and fixed definitions like "in" and "on". <p>It was written in Micro Planner and Lisp. The expressions that describe the physical context are ambiguous.</p>
	1. (Kelleher, 2003) 2. (Barclay, 2010)	

VITRA (VIsual TRAnslator) Fraunhofer Institute (IITB, Karlsruhe) project		<p>The VITRA project deals with the relation between vision and natural language.</p> <p>Three systems were developed as part of this project: the Citytour and Soccer systems with 2D images, and the Workbench using 3D images.</p>
2D Images	<ol style="list-style-type: none"> 1. (Herzog, 1992) 2. (Kelleher, 2003) 3. (Stoia, 2007) 4. (Barclay, 2010) 	<p>A knowledge-based artificial intelligence system was designed combining vision/image understanding and a natural language interface.</p> <p>The representation was built on three levels: linguistic, cognitive and sensory, where a process at the cognitive level extracted spatial relations and events.</p> <p>For the analysis of the world scene, they applied an approach for the automatic segmentation and cueing of moving objects and image sequences.</p> <p>It used 2D images of the real world with static and dynamic objects from a bird's eye view.</p>
CITYTOUR System (1986 -1988)	<p>Elizabeth André Guido Bosch Gerd Herzog Thomas Rist</p> <p>(André et al., 1986, 1987; André et al., 1989)</p> <p>1. (Barclay, 2010) (André et al., 1988)</p>	<p>CITYTOUR is a computational question answering system, the dialogue of discourse relates to a simulated trip in a city.</p> <p>It used a stationary monocular camera with more than 3,300 digitized (512 by 512 pixels, 8 bit) gray-scale frames processed by a system of actions.</p> <p>Developed in Common Lisp and Flavors on Symbolics 3600. Image processing was performed on VTE Digital Video Disk and VAX-11/780.</p> <p>This system used a 2D environment with static and dynamic objects from several real-world scenes.</p>

SOCCER System (1986 - 1988)	1. (Barclay, 2010)	<p>SOCCER describes short football scenes in natural language narrated in German.</p> <p>Through a generative rather than interpretive language, it used a geometric description of the scenes represented by a sequence of images.</p> <p>Moving objects were segmented and cued by computing and analyzing displacements of vector fields.</p> <p>It was developed in Common Lisp and CLOS, and the user interface was implemented in CLIM.</p>
3D Images Workbench (1992 - 1995)	<p>Herzog, Blocher, Gapp, Stopp, Wahlster, André, Rist, Nagel, Enkelmann, Zimmermann Wazinski.</p> <p>(Herzog, 1992, 1995)</p> <p>1. (Kelleher, 2003)</p> <p>2. (Stoia, 2007)</p> <p>3. (Barclay, 2010)</p>	<p>The VITRA project was extended to 3D visual integration.</p> <p>3D environments with static and moving images were designed using sequences of 2D images of the real world.</p> <p>The main aim of this project was to automatically interpret dynamic images. It managed the generation of short comments at football matches, the communication with autonomous mobile robots, the description of traffic movements and routes in 3D environments.</p> <p>It applied a hierarchical modeling system approach that provided simple geometric representations of solid objects using polygons, disks, rings, cylinders and spheres.</p> <p>This system used a synchronization technique to control the animation of the images (bitmaps) generated from underlying video sequences.</p> <p>Visual saliency was treated from the color perspective only.</p>

Constrained Connectionist System (1996)	Terry Regier (Regier, 1996)	<p>A model developed with a focus on learning in spatial language generation. The modeled system considers the assignment of spatial propositions. Path following is supported by reference points with two main properties:</p> <ul style="list-style-type: none"> • a prediction of sequential intermediates containing the encoded spatial meanings depending on the primary source and destination of the path, and • a prediction of configuration that expresses the meaning of a word, like "until", "outside" or "in", for example, in a language. <p>The system used abstract 2D pixelated images.</p>
	1. (Regier and Carlson, 2001) 2. (Kelleher, 2003) 3. (Barclay, 2010)	
Scene Describer (1999)	Alicia Abella John R. Kender (Abella and Kender, 1999)	<p>Computational system of spatial prepositions that integrates visual and linguistic ideas in order to generate natural language descriptions.</p> <p>The system architecture is organized by modules – image processing, semantic representation, generation of local expressions–, and includes a natural language generation processor.</p> <p>This system makes use of spatial relations, represented by fuzzy predicate logic. It considers the prior knowledge of the images used to make statements in natural language about location and space.</p> <p>The system was programmed to analyze x-ray images where every stain in the image represents an object that has simple geometric properties, such as alignment and distance. It works with 2D images, which contain 10 to 30 objects. For the experiments they used 2D image maps.</p>
	1. (Barclay, 2010)	

Virtual Director System (2000)	Amitabha Mukerjee Kshitij Gupta Siddharth Nautiyal Mukesh P Singh Neelkanth Mishra (Mukerjee et al., 2000)	<p>Computer system that reconstructs scenes in natural language using a set of linguistic descriptions related to an urban park. It handles imprecise descriptions.</p> <p>The system has two components: a database of objects with actions and a set of default constraints that correspond to the domain dependencies.</p> <p>The constraint set is combined with the use of multi-dimensional fuzzy functions called continuous (potential) fields. The parameters of the continuous fields, like the shape, position and movement, were determined experimentally. The fields were used to create default instances of the described objects.</p>
	1. (Kelleher, 2003)	<p>Object location is described by prepositional phrases. The system placed special emphasis on the intrinsic reference frame.</p> <p>The selection of the reference object was based on object size, although distance was also used.</p> <p>The experiments were conducted in 2D environments only.</p>

Attentional Vector Sum Model (2001)	Terry Regier Laura A. Carlson (Regier and Carlson, 2001)	<p>Computer model that predicts spatial linguistic expressions. The reference objects were selected according to their high geometric extensions or irregular shapes.</p> <p>This system has two conceptually distinct elements, respectively determining whether the path towards the reference point is a vertical or horizontal route. Both elements were implemented by functions of sinusoidal equations.</p> <p>Reference object attention is weighted by a sum of vectors between the reference and the target objects.</p> <p>Predictive models are used in the experiments with this system.</p> <p>No mention is made of a formal model of visual salience. However, the center of mass, orientations and distance of objects were used, as proposed by Gapp (1995).</p> <p>Abstract 2D images of objects were used for experiments.</p>
	1. (Barclay, 2010)	
Describer System (2002)	Deb K. Roy (Roy, 2002)	<p>This computer system is able to generate expressions in spoken natural language.</p> <p>The system learns to describe and identify geometric objects in a visual scene. It uses a learning algorithm to build a symbolic representation of the semantics.</p> <p>There is no mention of a formal model of visual salience. However, the description of the target object is based on its shape, color, size and position.</p> <p>The scenes are represented by a set of 2D images (rectangle).</p>
	1 (Barclay, 2010)	

Situated Language Interpreter System (2003)	<p>John D. Kelleher</p> <hr/> <p>1. (Kelleher, 2003) 2. (Stoia, 2007) 3. (Barclay, 2010)</p>	<p>Computer system for the interpretation of spatial language.</p> <p>For the computational interpretation of the expressions, the system develops and implements a model of discourse.</p> <p>It is based on a semantic framework to support natural language with locative prepositions. The semantic model contains a mechanism for the user to select different frames of reference.</p> <p>The system allows users to move in a 3D environment and manipulate objects.</p> <p>The system was developed using Ling, with an object-oriented programming language, similar to C++ or Java.</p> <p>A visual saliency algorithm is applied in this system to try to integrate speech with visual information.</p> <p>The system interacts in a 3D virtual reality scene, projected on a 2D plane, containing different types of objects.</p>
System for Spatial Knowledge Representation for Human- Robot Interaction (2003 - 2006)	<p>Reinhard Moratz Thora Tenbrink John Bateman Kerstin Fischer</p> <hr/> <p>(Moratz et al., 2003) (Moratz and Tenbrink, 2006)</p> <hr/> <p>1. (Stoia, 2007)</p>	<p>A computational model for representing spatial knowledge appropriate for human-mobile robot interaction.</p> <p>The system has an architecture composed of several components: the syntactic component, the spatial reasoning component and action component.</p> <p>The model consists of two parts: a knowledge base of the world and the representation of information about the position of objects.</p> <p>The system includes a text-mode interface, which provides users with written instructions from the computer. Their experiments included tests with expressions of human-human and robot spatial</p>

		<p>communication.</p> <p>Aspects of object visual saliency were employed for this project.</p>
Bishop System (2004)	Peter Gorniak Deb Roy (Roy, 2002)	<p>The system interprets and understands the human visual language through a model for visually-grounded language to describe objects in scenes.</p>
	1.(Barclay, 2010)	<p>The system includes a descriptive spatial language with hundreds of reference expressions based on similar scenes, a syntactic analyzer of spoken expressions and a composition engine managed by an interpreter that combines various lexical units.</p> <p>This system has an algorithm for extracting visual features and uses the visual saliency for a description of the features of objects. It also manages the description of spatial relations.</p> <p>3D scenes were used for its implementation, but spatial relations are interpreted in 2D.</p>
Space Case (2005 - 2006)	Kate Lockwood Ken Forbus Daniel T. Halstead Jeffrey Usher (Lockwood et al., 2006; 2005)	<p>A computational model that uses spatial propositions combining geometric and functional influences.</p> <p>The functional representations of the system are managed by a large database containing knowledge and user skills from other already performed spatial tasks.</p> <p>The system is designed to provide information on the relevance of the scenes.</p> <p>They used sKEA (sketching Knowledge Entry Associate) as a sketch understanding system. A sKEA sketch is a set of pictograms with actual drawings.</p>
	1. (Barclay, 2010)	<p>This system used 2D drawings with a lateral perspective.</p>

GLIDES System (2006)	<p>Paul Williams Risto Miikkulainen Williams and Miikkulainen,</p> <p>1. (Barclay, 2010)</p>	<p>System with neural network architecture developed to describe simple scenes in natural language.</p> <p>The system uses intelligent symbol representations based on visual perception.</p> <p>The system consists of two memory modules. They handle linguistics and visual aspects, respectively, between which associative connections are maintained. The system stores the learned relations.</p> <p>Abstract 2D images were used to develop the experiments.</p>
Situated Artificial Communicators (2006)	<p>Gert Rickheit Ipke Wachsmuth (Rickheit and Wachsmuth, 2006)</p> <p>1. (Barclay, 2010)</p>	<p>System with a computer model for generating propositions in order to identify objects in still images.</p> <p>The results from the experiments on this system were validated in computer simulations of cognitive processing with a Bayesian network.</p> <p>The network contains a cluster node for each object, whose range is equal to the number of object types in the scene.</p> <p>It does not focus on the selection of references.</p> <p>This project uses an object classification process.</p> <p>Experiments were run using 3D images (using a camera) and blocks of, 10 to 20 objects.</p>

Reference Object Choice in Spatial Language: Machine and Human Models (2010)	Michel Barclay (Barclay, 2010)	<p>A machine model for processing scene descriptions which operates in a sufficiently realistic environment; not distinguishable from the mental models used by humans.</p> <p>This project emphasizes the use of references and spatial relations for locating objects, taking into account perceptual salience.</p> <p>It was tested on a large set of representations of developed 3D scenes.</p>
Moscaret (2013)	Thanh-Hai Trinh 1. (Barclay, 2010)	<p>This system implements a semantic model of spatial relations between objects and virtual environments.</p> <p>It specifies the spatial relations at a conceptual level using the VRX-OCL language.</p> <p>The model focuses on the spatial limitations of virtual environments, such as space communication difficulties.</p> <p>There is no mention of visual salience.</p> <p>The system included a tool developed for semantic modeling of VEs as close to reality as possible.</p>

Table 3.3 identifies the specific areas addressed by each computer system for these same projects. Column 1 lists the name of the system; column 2 specifies whether syntax and semantics were used to create the spatial language; column 3 indicates whether the system includes object localization; column 4 is concerned with whether the concept of object salience is used; column 5 refers to whether VEs were built for the system; column 6 describes whether the system interacts with humans or robots and, finally, column 7 shows whether AI was applied for the algorithm.

Table 3. 3 Other areas addressed by computational systems

Name	Syntax and Semantics	Object location	Saliency	Use of Virtual Environment			Interaction		Artificial intelligence
				Text	2D	3D	Human-computer	Robot	
SHRDLU	✓	✓	✗	✓	✗	✗	✓	✗	✓
2D Images	✓	✗	✗	✗	✓	✗	✓	✗	✓
CITYTOUR System	✓	✗	✗	✗	✓	✗	✓	✗	✗
SOCCER System	✓	✗	✗	✗	✓	✗	✓	✗	✗
3D Images Workbench	✓	✗	✓	✗	✓	✓	✓	✗	✗
Constrained Connectionist System	✓	✓	✗	✗	✓	✗	✗	✗	✓
Scene Descriptor	✓	✓	✗	✗	✓	✗	✓	✗	✗
Virtual Director System	✗	✓	✓	✓	✓	✗	✓	✗	✗
Attentional Vector Sum Model	✗	✓	✓	✗	✓	✗	✓	✗	✗
Descriptor System	✓	✓	✓	✗	✓	✗	✓	✗	✗
Situated Language Interpreter System	✓	✓	✓	✓	✗	✓	✓	✗	✗
System for Spatial Knowledge Representation for Human-Robot Interaction	✓	✓	✓	✓	✗	✗	✓	✓	✗

Bishop System	✓	✗	✓	✗	✓	✓	✓	✗	✗
Space Case	✗	✗	✓	✗	✓	✗	✓	✗	✗
GLIDES System	✓	✗	✗	✗	✓	✗	✗	✗	✓
Situated Artificial Communicators	✓	✓	✗	✓	✓	✗	✗	✓	✓
Reference Object Choice in Spatial Language: Machine and Human Models	✓	✓	✓	✗	✗	✓	✓	✗	✗
Moscaret	✓	✓	✗	✗	✓	✓	✓	✗	✓

In most of the analyzed systems, the use of semantics also stands out as a key aspect. Fifteen of the 18 systems developed a semantic model for the correct interpretation of symbols and words. Syntax is another important aspect underlying interest in the selection and combination of the right words.

The main objective of 11 of the systems is object location. However, all systems were analyzed in search of key aspects for describing objects and scenes, such as a semantic model for studying linguistic behavior.

The reviewed papers describe spatial relations based on visual perception and natural language, thereby generating linguistic expressions.

As Table 3.3 shows, only nine systems consider a model of perceptual salience based on the basic characteristics of objects.

All the systems used a virtual environment, albeit with varying dimensionality (i.e., textual, 2D or 3D).

Human-computer interaction is the most common interaction mode (16 systems), and only a couple of the projects used interaction with robots.

Projects, like 2D Images, Abella and Kender's Scene Descriptor, the Virtual Director System, System for Spatial Knowledge Representation for Human-Robot Interaction, the GLIDES System and the Bishop System, explain the design of their architecture and their components, giving a clear description of their responsibilities. However, other systems, like Situated Artificial Communicators, provide very limited information so that it is impossible either to visualize technical details and development tools or to gather evidence from the results of the respective experiments.

Several of these systems consider the use of reference objects, linguistic expressions and/or spatial reference frames. However, only 11 applications, namely the CITYTOUR System, SOCCER System, Abella and Kender's Scene Descriptor, the Virtual Director System, the Descriptor System, Kelleher's Situated Language Interpreter System, the System for Spatial Knowledge Representation for Human-Robot Interaction, the Bishop System, Space Case, Reference Object Choice in Spatial Language: Machine and Human Models, and Moscarel, specify precisely how they are used. Of these applications, seven explicitly specify the implementation process. The other applications, including SHRDLU, 2D Images, 3D Images Workbench, Regier's Constrained Connectionist System, the Attentional Vector Sum Model, the GLIDES System, and Situated Artificial Communicators, describe the use and implementation of these aspects implicitly.

Artificial intelligence is a multidisciplinary area applied in the development of six of these projects, specifically SHRDLU, 2D Images, Regier's Constrained Connectionist System, the GLIDES System, Situated Artificial Communicators, and Moscarel. Through computing and logic, artificial intelligence has helped to implement real dialog between the user and the computer. Furthermore, through the support of neural networks, knowledge bases and Bayesian networks, it has been possible to develop models that simulate human intelligence, which is implicitly linked to the object location process.

In this section was presented a review of computer systems that use spatial language for object location. The process of locating objects has been conducted in both real and virtual, 2D and 3D environments. Borrowing from linguistics and psychology, semantics and syntax are two of the fields involved in the study of frames of reference, perceptual salience and cognitive maps that have supported the comprehension and generation of artificial spatial languages, which is, as mentioned, a seemingly simple but actually highly complex process.

The analyzed systems present some limitations, barriers to a complete and efficient interaction with humans in the localization process. However, important advances have been made in this ongoing research area (Lara et al., 2016b).

3.3 Conclusion

The state of the art described in this chapter focused on two lines of research covering the scope of the topic of this thesis: computational models of perceptual saliency and the spatial language system. Some case studies related to the computational proposal addressed in this thesis were reported for each of these lines of research. Likewise, the analysis of each paper provided a practical insight into several projects, each covering one or more of the areas of study of this research.

This research into related work was conducted separately because no papers addressing all the fields of study covered by the proposal of this model were retrieved.

This confirms that, as mentioned at the beginning of this chapter, there are no projects covering the study areas addressed by the approach presented in this proposal in the literature.

Part II: Contributions

Chapter 4: Computational model of perceptual saliency for 3D objects in virtual environments

4 Computational model of perceptual saliency for 3D objects in virtual environments

The principal purpose of this chapter is to present a computational model of perceptual saliency. This computational model of perceptual saliency emerges from the need to select the best reference object within a 3D virtual environment. The object will be used within a computer system that allows the automatic generation of natural language directions to help a human user to locate specific objects within the same virtual environment. The computational model of perceptual saliency is centred on the visual features of an object such as color, size, and shape. The basic features of a 3D object in a virtual environment are individually extracted with different methods proposed in this thesis. Then, the information obtained is normalized according to several criteria, and the values are combined using a linear function that helps to determine a measure of general saliency for each 3D object within a virtual environment.

An object can attract the perception of the observer by its natural or distinctive features (Vargas and Lahera, 2011), but it also depends on the interaction of its basic features with respect to other objects (Gapp, 1995; Hoffman and Singh, 1997; Spotorno et al., 2013; Stoia, 2007). The model of saliency should be able to measure both the intrinsic and the contextual saliency of each object, at the same time that allows the identification of the most salient object in a scenario.

In the following sections the computational model of perceptual saliency is explained, presenting the mathematical representation of the model for each of its parts (individual saliency, contextual saliency, and general saliency). Finally, the implementation process is described. The evaluation of applying this computational model of saliency within a virtual environment is presented in the sections 4.5.1, 4.5.2 and 4.5.3.

4.1 General Overview of the Computational Model of Perceptual Saliency

As a previous task for the design of this computational model of perceptual saliency different models of saliency were analyzed as described in Chapter 3 (see Section - 3.1 Computational

Models of Perceptual Saliency). According to different studies of the perceptual saliency of objects (Caduff and Sabine, 2008; Frintrop and Rome, 2010; Gapp, 1995; Li et al., 2011; Li et al., 2013; Oliva et al., 2003; Röser et al., 2011; Röser et al., 2012; Röser et al., 2013), saliency is related to the visual characteristics of an object that can capture the attention, such as: color, size, shape, texture, contrast, orientation and motion.

A review of the literature, led us to conclude that the visual features of an object with the greater stimulus for the human visual system are: color, size and shape; this is why these features were selected to be used in this computational model. The order of influence of these features was established according to (Gapp, 1995). Mangold (1986) demonstrated in his experimental studies that the color is more dominant than the size and shape in object identification tasks. Likewise, size is more easily recognizable than shape.

The proposed computational model of perceptual saliency is based principally on three parts: first calculation of the individual saliency of the object independently of its context (environment); second, calculation of saliency of the object with respect to its context, and third, calculation of a general measure of saliency (also called global saliency). Figure 4.1 reflects these three steps:

Figure 4. 1 Scheme of computational model of perceptual saliency

4.2 Modelling the Individual Saliency

Next is described the first part of the computational model of perceptual saliency, from an individual approach to the objects. In this section are explained the processes of characterization (also called "processes of extraction of characteristics") of the three factors (or basic features of the object) chosen for the computational model of perceptual saliency. Characterization is considered as the process of abstraction and quantification of features (color, size and shape) of a 3D object in a virtual environment. The information obtained from this characterization is normalized in the case of size, using a set of functions with several criteria. For the color and the shape of the objects, other criteria are taken into account. The values obtained from the process of normalization or during the feature extraction process are combined using a linear function that allows us to obtain a measure of individual saliency in the interval $[0,1]$ for each 3D object within a virtual environment.

Characterization of Color

An object can stand out from its surroundings on color alone. Take, for example, a white cube in the midst of a group of red cubes. Color is a difficult property to measure, identify and compare, although some studies use perceived light as a complex function of illumination, reflectance/absorption on surfaces of objects, and receptive abilities of the visual sense (Raubal and Winter, 2002).

The human visual system is able to distinguish hundreds of thousands of different colors. There are a number of color models (also called color systems or spaces) useful for the characterization of color as perceived by the human eye, like the RGB, CMYK, CIE-XYZ tristimulus, CIE-Lab, CIE-LUV, among others. RGB is the color model most commonly used in digital image processing (e.g. scanners, cameras and monitors). The RGB color model is additive, that is a color is obtained from the mixture of different proportions of red, green and blue. These components are measured with a scale from 0 to 255 and each color is defined by a set of three values. However, the RGB system presents a number of disadvantages:

- *High correlation between its components:* The RGB system has a dependency on the light intensity of its three components: red, green and blue. The RGB values depend on

the ability and specific sensitivity of each device for capture and visualization; thus, the RGB color space is device-dependent. However, with methods for device calibration, the RGB space can be transformed into a linear perceptually uniform color space (Tkalčič and Tasič, 2003).

- *Not psychologically intuitive:* RGB is not based on any psychological perception model; this psychological non-intuitiveness is a serious problem for humans that present, in some cases, the trouble of a bad visualization of colors defined by its three components (Tkalčič and Tasič, 2003).
- *Not uniformity.* The distance RGB space does not correspond to perceptual distance (Gevers and Smeulders, 1999; Tkalčič and Tasič, 2003); it does not have perceptual uniformity, i.e. there is a low correlation between the perceived difference of two colors and the Euclidian distance in the RGB space. This is an evident problem to take this model for the computation of perceptual saliency.

Because the RGB model is oriented towards the hardware, it is not intuitive, and it is not related to the concepts of hue, saturation and brightness. Therefore, for a color in RGB format to be suitable for our saliency model, it is necessary to resort to the transformation of its values into a uniform color space, where the distance between the points is directly proportional to the difference in the perceived colors. Our target color space is known as CIE-Lab, which was designed to be perceptually uniform (Connolly and Fliess, 1997; López et al., 2005). The CIE-Lab color space (strictly known as CIE 1976 L * a * b *) is oriented towards the human eye; it allows quantifying color differences that can be expressed in terms of human visual perception. Differences (ΔE^*) between two colors can be obtained from CIE 1976 L*a*b* with the following equations (1) (Connolly and Fliess, 1997; Yuan et al., 2007):

$$\Delta E^* = \sqrt{[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]} \quad (1)$$

The color difference (ΔE^*) is often used for the evaluation of color reproduction quality in an image processing system (Saleh et al., 2003).

In order to transform a color of the RGB system to the CIE-Lab system, the RGB system has to be first transformed into the CIE-XYZ system (also called XYZ tristimulus system). The XYZ tristimulus system is used as a platform from which other specifications of color are made, and it is an intermediary for the determination of perceptually uniform color systems as CIE-Lab or CIE-LUV. In the XYZ tristimulus system, CIE defines colors with the variables X, Y and Z, where Y represents the color luminosity parameter, and the variables X and Z represent the virtual components of the primary spectres for the calculation of all colors through its addition (Wyszecki and Stiles, 1982).

Figure 4.2 shows the sequence of transformations for the characterization of the color of a 3D object in a numerical format. The proposed procedure is simple, it includes three steps: (1) the color of the 3D object is obtained in the RGB system; (2) the RGB color values are converted to the CIE-XYZ color system; and (3) the values in CIE-XYZ are used to convert the color to the CIE-Lab system.

Figure 4. 2 Sequence of transformations of the color of a 3D object

To convert a color of RGB standard to the XYZ tristimuli system, the transformation matrix proposed by Wyszecki and Stiles (1982) is used:

RGB to XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Afterwards, the tristimuli value is transformed to the CIELAB Cartesian coordinate system, using the following linear equations (CIE, 1978):

XYZ to CIE-Lab: Equations (2)-(4)

$$L = 116 X \left[\left(\frac{Y}{Y_n} \right)^{1/3} - 16 \right] \quad (2)$$

$$a = 500 X \left[\left(\frac{X}{X_n} \right)^{1/3} - \left(\frac{Y}{Y_n} \right)^{1/3} \right] \quad (3)$$

$$b = 200 X \left[\left(\frac{Y}{Y_n} \right)^{1/3} - \left(\frac{Z}{Z_n} \right)^{1/3} \right] \quad (4)$$

Here X_n , Y_n , and Z_n represent the values of X , Y , and Z . These are the tristimulus values of the illuminant, in this case the illuminant D65 (Connolly and Fliess, 1997). The illuminant D65 represents the normal conditions of diurnal light. It corresponds to daylight with a color temperature of 6500 °K (degrees Kelvin) (Domínguez, 2012). With this series of transformations, the characterization of the color of an object in a numerical format of three digits is obtained (see Figure 4.3). This characterization is represented in a vector of three components. To represent the individual color of an object within this computational model of perceptual saliency is used the variable (C).

Figure 4. 3 Characterization of color

Saliency by color

In order to identify the most salient colors to the users' view, and measure the saliency by color, we perform a deep search. However, it was not possible to find in literature a way of quantifying the degree of saliency by color, so that it could lead to a mathematical function. In this search, we found was a list of colors that stimulate the human visual system proposed by (Choungourian, 1968), which we decided follow to determine the saliency by color. In this sense Kröser and Julesz (1989) cited that the saliency can relate, with the detectability of the stimulus.

Choungourian (1968) stated that the colors with major stimuli to the human visual system are: red, orange, yellow, yellow-green, green, blue-green, blue, and purple. This was demonstrated in a set of experiments and studies conducted with university students from different countries, with equal numbers of men and women. Results showed that there were no differences in sex or culture in the preference of the above-mentioned colors. Table 4.1 presents the list of these colors with their numerical representation in the RGB and CIE-Lab systems.

Table 4. 1 List of salient colors according to Choungourian

Color	R	G	B	CIE		
				L*	a*	b*
Red	255	0	0	53.23	80.10	67.22
Orange	255	112	40	64.16	51.02	62.62
Yellow	255	255	0	97.13	-22.55	94.48
Yellow-green	154	205	50	76.53	-37.99	66.58
Green	0	255	0	87.73	-86.18	83.18
Blue-green	13	152	186	58.11	-20.81	-26.94
Blue	0	0	255	32.30	79.19	-107.86
Purple	102	2	153	27.20	59.25	-56.42

For the computational model to calculate the color saliency of an object, a three step process is followed:

(1) The distances between the color of the object with respect to the group of the 8 most salient colors is calculated using Euclidean distances.

(2) Then, the salient color with less distance to the color of the object is selected, and a saliency value is assigned to each of its 3 coordinates (L, a, b) according to the next ranges:

- '1.00' for distances of less than 5 units
- '0.75' for distances bigger than 5 and less than 10 units
- '0.50' for distances bigger than 10 and less than 15 units
- '0.00' for distances bigger than 15 units

(3) Finally, the three saliency values are averaged to get the saliency by color of the object.

We are aware that this is a rough categorization for color saliency, but to the best of our knowledge there are not alternative solutions that allow for a better graduation of intermediate values. For this reason, in the empirical evaluation, it is contrasted if the measure of saliency

for color that one has proposed, represents adequately the perception of saliency for color that the human beings have.

The value of saliency by color of the object is represented in the computational model by the variable ($Sc_{[0-1]}$), see Figure 4.4.

Figure 4. 4 Saliency by color

In this initial approach, the computational model of saliency is only for 3D objects with one solid color. The RGB color of the 3D object can be recovered from the software tool with which it was modelled. As part of the future work we would like to extend the characterization of the color of this computational model to include objects' texture.

Characterization of Size

An object can be attractive to the view of the human being simply by its size. The size usually comprises the length, width and height of the object. However, for some objects the size refers only to a specific dimension, and this dimension varies between objects. The status (open or closed) or position (horizontal or vertical) of an object is also an important aspect to specify the size. For example the size of an open envelope is not the same as the size of a closed envelope, so also the position where the envelope is placed in the space this can change the relation between the measures of its size (see Figure 4.5 (a) and (b)).

Figure 4. 5 (a) status and (b) position of an object

A solid object represents a three-dimensional geometric body that occupies a place in space. With the volume measurement of an object it can be identified its size or physical extent in three dimensions (i.e. height, length and width), using, for example, cubic centimetres (cm³) or cubic meters (m³) as unit of measure.

For the computation of the volume of a 3D object in a virtual environment, it is usually easy to get the length, width and height of the 3D object in the world coordinate system (see Figure 4.6). The product length * width * height gives us a quick approximation to the volume; however, this mathematical expression can only calculate precisely the space occupied by geometries such as cubes and rectangular prisms. For the calculation of the volume of geometric objects such as pyramids, cylinders, cones and spheres, among others, the formula should vary according to the specific characteristics of each object.

Figure 4. 6 Measures for the volume of an object

However, objects in the real world or in a virtual environment are frequently not regular geometric figures, thus the mathematical expressions that are used to calculate the volume of regular geometric figures are not always appropriate for all objects. An automatic measurement for the volume of regular and irregular 3D objects in a virtual environment is needed, without regard to their orientation and position.

In this sense, we look for a measurement for the size that would fit to the orientation and position of the object in the space in all its possible ways of been placed, without changing its fixed point of reference. An object can be rotated about its point of reference and the measure of size not change. The change of position of an object comprises only rotations on its axes. Therefore, we propose a volume measure through a pre-voxelization of the object, which implies to consider the following three previous steps:

- (1) *The 3D object is voxelized.* The voxelization process is the segmentation of an object into small cubic portions, a unit called voxel, which conform and represent the three dimensional object, as shown in Figure 4.7. The voxelization process will help us to characterize both the size and the shape of an object. Regarding the size of the voxel (v_{si}), in virtual environments it is common to work with sizes in terms of centimetres; then the size of the voxel is here established as one cubic centimetre.

Figure 4. 7 View of a voxelized 3D object

(2) Then, the total number of voxels (nv) for the object is obtained. The number of voxels (nv) of an object is obtained by the voxelization algorithm.

(3) Then, the number of voxels (nv) is multiplied by the voxel size (vsi). This step obtains the object volume (vo) in cubic centimetres with the equation (5):

$$vo = nv * vsi \quad (5)$$

Additionally to the volume, three parameters are considered to fully characterize the size of a 3D object: length, width, and height of its bounding box. A bounding box is a three-dimensional box, that delineates the boundaries of an object; and it also has length, width and height. This method permits to characterize the size of 3D objects both in a simple and precise manner. This characterization is represented in a vector of four components. The characterization of the object's size in this computational model of perceptual saliency is represented by the variable (Si), (see Figure 4.8).

Figure 4. 8 Characterization of size

Saliency by size

After the color, according to the literature, as mentioned, the size of an object is the second characteristic with major influence for determining its saliency. The saliency by size cannot be measured up of totally independent form to the rest of the objects that exist in the

environment. For this, we need to determine, which is the size of the object of major size with the aim to assign the maximum saliency to the above mentioned size (1).

From the list of all the objects in the virtual environment, the largest object in each of its size parameters (i.e. length, width, height, and volume) is used to establish the maximum value of saliency for each of them. For example, if the length of the longest object is 12000 centimetres, this length will be assigned saliency 1.

Then, in order to calculate the saliency by length of an object, a proportional value will be calculated with respect to the longest object. The next equation (6) is used:

$$Sl_{[0-1]} = \frac{l}{hlv} \quad \text{Saliency for length} \quad (6)$$

Where $Sl_{[0-1]}$, represents the normalized value of saliency by length, calculated by dividing the object's length (l) by the highest length value of the objects in the virtual environment (hlv). Likewise are calculated the saliency by width $Sw_{[0-1]}$, height $Sh_{[0-1]}$, and volume $Sv_{[0-1]}$.

The four values are then averaged using the equation (7) to get the saliency by size of the 3D object, represented in the computational model by the variable ($Ssi_{[0-1]}$), see Figure 4.9.

$$Ssi_{[0-1]} = \frac{Sl_{[0-1]} + Sw_{[0-1]} + Sh_{[0-1]} + Sv_{[0-1]}}{4} \quad (7)$$

Figure 4. 9 Saliency by size

According to the measurement of individual saliency obtained by our model, the object with the largest volume will be mathematically most salient than the object with the lower volume; however not always the largest object in an environment is the most salient. In some cases it is possible that a small object could be the most salient, for example if it contrasts with respect to the other objects in the scene, see Figure 4.10. This case is treated with more detail in the section 4.3.

Figure 4. 10 Contrast in size

On the other hand, if all objects are of nearly the same size in the virtual environment, other characteristics such as the shape (see Figure 4.11 (a)), the orientation (see Figure 4.11 (b)) or the distance between objects (see Figure 4.11 (c)), could help to distinguish one object from the others (Gapp 1995). The shape of the object was included in our computational model.

Figure 4.11 Other characteristics of an object (a) the shape, (b) the orientation and (c) the distance

Characterization of the Shape

As mentioned, the shape of an object contributes to its visual attraction. When the visual interest of an object is its shape, the measure could be specified by the deviation of its shape from a rectangle or by its shape factor, which represents the height-to-width ratio. For example, skyscrapers have a high shape factor; whereas long and low buildings have a low shape factor (Raubal and Winter, 2002), see Figure 4.12.

Figure 4.12 Shape factor: high and low

This could be a simple way to specify the shape of a 3D object, although with not enough detail for our purposes. Given that could arise situations like the following: Figure 4.13 shows

a rectangle (a) and a rhombus (b) with the same shape-factor. The two objects have the same length and width. However both objects have a different shape, given this situation the following question arises: really these two objects are equal of salient by shape? An answer to this would be that the rhombus could draw more attention than the rectangle, because the rhombus deviates more with respect to the rectangle. This allows us to assume that the shape factor is not a good way to measure the saliency by shape of a 3D object.

Figure 4. 13 Equal Shape factor

One of the toughest challenges of this computational model of saliency was undoubtedly the characterization of shape, due to the large amount of information it possesses, making it a complex characteristic. Normally, the shape of a 3D object is generally recognized by its vertices, edges or sides, aspects that are generally projected towards the observer's viewpoint in virtual environments. However, the implementation of this appreciation, from a computational approach might be complicated.

In order to find an optimal way to extract and represent the properties of the shape of a 3D object within a virtual environment, that is, to obtain the characterization of the shape, we propose the following two forms of representation:

- (1) A vectorial representation of the object, using the describer of Zernike. The 3D Zernike descriptor will be used, also known as descriptor of geometric moments with orthogonal bases. This descriptor has been used effectively as a method to represent the global shape of a 3D object (Esquivel-Rodriguez and Kihara, 2012; Kihara et al., 2011; Sael et al., 2008; Venkatraman et al., 2009).

The 3D Zernike moments' descriptor allows obtaining the properties of the shape of a voxelized 3D object, as a collection of numeric values, resulting from evaluating the object with the 3D Zernike moments in a specific order. The order is a property of this type of descriptor that allows determining the size of the vector.

This representation, though it is not going to serve us to determine the saliency by individual shape, will help us to calculate the saliency by shape with regard to the context.

- (2) A voxelized representation of the object in its bounding box. This representation allows us to determine the individual saliency of each object within of the virtual environment.

The characterization of the shape is represented in the computational model with the variable (Sh), see Figure 4.14.

Figure 4. 14 Characterization of shape

Then, having examined the methods of how to mathematically represent the shape of a 3D object, and hoping to find some way with which it were possible to calculate the saliency of a certain shape, an investigation was conducted through an exhaustive study (see Chapter 3,

section 3.1 Computational Models of Perceptual Saliency). Not finding what we wanted, it was necessary to propose a metric own to evaluate the saliency by shape, inspired by the idea of calculating the deviation from a rectangle (Raubal and Winter, 2002).

This measure is based on the central idea of the volume of a pre-voxelized object.

Saliency by shape

Assuming that the flatter the surface of an object is, the less salient it is, and inversely, objects with high pointedness tend to be perceptually more salient, we propose to calculate the proportion of empty space and full space in voxels in the bounding box of the 3D object.

For this, first the volume of the bounding box in voxels (*siBBBox*) is obtained, and then the volume in voxels of the object (*vo*) is subtracted from this number. The result of this operation is the "Empty space" (*Es*), it represents the number of free voxels within the bounding box that are not part of the object; conversely, the volume of the object in voxels corresponds to the "Full space" (*Fs*) in the bounding box. Figure 4.15 illustrates the empty and full spaces in the bounding box of a 3D object.

Figure 4. 15 Empty and full space in the bounding box of a 3D object

Finally, the empty space (Es) divided by the total of voxels in the bounding box ($siBBox$). It is the empty space proportion which will be used as measure of shape saliency for the model. The next equation (8) is used.

$$Ssh_{[0-1]} = \frac{Es}{siBBox} \quad (8)$$

With the above calculation, we provide a direct estimation of the saliency of an object's shape. The saliency by shape variable is represented in the model by ($Ssh_{[0-1]}$), and it takes values between [0 - 1], to the being a proportion (see Figure 4.16).

Figure 4. 16 Saliency by shape

Calculating the individual saliency of the 3D object

The calculation of the individual saliency for each object within the scenario is made through a weighted sum. According to Gapp (1995) and Mangold (1996) the color is the most salient factor, followed by the size and shape. Unfortunately, the literature does not provide a measure of the relative importance of each factor. As a tentative solution a weight was assigned to each factor as follows: 0.5 to the color, 0.3 to the size, and 0.2 to the shape. The equation (9) is then proposed to calculate an individual saliency for each 3D object in the scenario.

$$IS = (Sc_{[0-1]} * 0.5) + (Ssi_{[0-1]} * 0.3) + (Ssh_{[0-1]} * 0.2) \quad (9)$$

Where IS represents the individual saliency of each 3D object and $Sc_{[0-1]}$, $Ssi_{[0-1]}$ and $Ssh_{[0-1]}$ correspond to its saliency by color, size and shape respectively, see Figure 4.17.

Figure 4. 17 Individual Saliency

4.3 Modelling the Contextual Saliency

The following describes the second part of the computational model of perceptual saliency, concerned with the saliency of an object with respect to its context. The general strategy of this stage is to calculate the color, size and shape averages in context. After that, it is calculated the distance of the color, size and shape of each object with respect to the average of the color, size and shape in the context. Finally, it is described how to measure the saliency by color, size and shape of objects, using certain determined criteria.

It is important to consider that although the human beings live in a three-dimensional world, we only see in two dimensions. In other words, the human visual system only receives centred projections on plane images of two dimensions, see Figure 4.18. This is why the representation of the color and size of an object with respect to its context was extracted from a 2D projection in this model of contextual saliency.

In this model of contextual saliency, the context is the projected 2D image that the observer can see of the virtual environment at a given time. The three objects of the figure 4.18 (cube, sphere and cylinder), are seen of different form, from the point of view of the observer. The cube although is large, the observer can only see one side of it. The sphere is the smallest object and is totally occluded, so that the observer cannot see him within his context. The cylinder is very large, but the observer can only see a small part of it, because it is also being occluded by the cube.

From this perspective, these cases were addressed in our model of contextual saliency, to handle color and size. However, the shape of objects was treated in a special way, which is also described this section.

Figure 4. 183D vision system (Inspired in (Cruz, 2007))

Average color in the context

The approach consists in obtaining the amount of each color seen by the user in a certain context. If a color is abundant in the scene, the objects with that color may be little salient, or on the contrary, an object with a color that does not appear too much might be salient with respect to color.

To simplify the calculations, and not to have to treat with every color separately, we calculate the average color of the context, which is possible because we have a representation of the color (CIELab) that allows operating arithmetically with the colors.

To obtain the average color in a context, it is necessary to calculate the amount of each color in the virtual environment, from a 2D perspective. It is important to realize that the 2D images that the people see in a computer can be represented as a matrix of pixels, having each pixel a certain color.

Trying to illustrate a little this situation, Figure 4.19(a) presents a scene in three dimensions. In the Figure 4.19(b) is represented the same scene in two dimensions.

Figure 4. 19 View of a 3D scene perceived by the observer in 2D, with obstructions

If a rectangular screen is considered with a resolution of 600 pixels of height by 800 pixels of width, we have a spatial resolution of 480 000 pixels. With the equation (10) we calculate the total number of pixels of a context ($tpctx$).

$$tpctx = h * w \quad (10)$$

Where

h is height (in pixels) of the context.
w is width (in pixels) of the context.

It is also important to realize that the amount of visible surface of a 3D object within the field of vision of the user will depend on the proximity or distance of the user to the object and the occlusions between objects. The amount of visible surface of a 3D object in a specific context can also be measured as a number of pixels in a 2D image.

To calculate the percentage per color in a context, and also to calculate the amount of visible surface for the objects in a context, a technique inspired on the rendering of 3D graphics within a 2D image is proposed. The technique consists in launching several rays, one for each pixel in the screen. A counter is maintained for each object and for each color. When a ray impacts in one object, the counter for the object is increased, as well as the counter for the color of the object. The result of this process is the visible surface for all objects that are visible, discarding those that are not. The number of impacts will vary depending on screen resolution; therefore the measure should be transformed into a percentage in order to make it invariable to screen resolution.

The equation (11) allows calculating the percentage of the visible surface of the object (vso).

$$\% vso_{i=1}^N = \frac{Ni_{i=1}^N}{tpctx} * 100 \quad (11)$$

Where

$\% vso$ is the % of the screen that corresponds to the visible surface of the object, for each object i present in the context.
 Ni is the number of impacts, for each object i present in the context.
 $tpctx$ is the total number of pixels of context (this also corresponds with the total of launched rays).

This solution proves to be effective and efficient to measure the surface of regular and irregular objects, even with occlusions.

Regarding the amount of each color in the context, when two or more surfaces of objects have the same color, the pixels for all of them are accumulated. $tpic_{color}$ represents the total number of pixels per individual color. The idea is simply to accumulate all the impacts on objects of the same color.

The relative frequency of each color in the context is calculated using the equation (12).

$$f_{color} = \frac{tpic_{color}}{tpctx} \quad (12)$$

Table 4.2 shows the color of structural objects and non structural, the number of object, the total number of pixels per individual color (tpc_{color}), the relative frequency per each color (f_{color}), and the percentage of color in the context (w_i), of objects presents for the scene in figure 19(b).

Table 4. 2 Relative frequency and Percentage of color for the scene in figure 19(b)

Color	No. of object	The total number of pixels per individual color ($tpic_{color}$)	Relative frequency of color (f_{color})	Percentage of color in the context (w_i)
Green	1 and 4	70 000	0.146	14.6 %
Yellow	2	10 000	0.021	2.1 %
Red	3	30 000	0.062	6.2 %
Grey	5	25 000	0.052	5.2 %
White	Context	345 000	0.719	71.9 %
Total		480 000	1	100 %

Figure 4.20 shows the color distribution of the figure 4.19(b).

Figure 4. 20 Color distributions within the context of figure 19(b)

With the total percentage of each individual color and the mathematical representation of the colors in three coordinates under the system CIE-Lab, the average color in the context is calculated. For this calculation a weighted average is computed, using the following equation (13).

$$C_{avg} = \frac{\sum_{i=1}^N c_i w_i}{\sum_{i=1}^N w_i} \quad (13)$$

Where

C_{avg} is the weighted average which is calculated for each of the three coordinates in CIE- Lab (L, a, b).

C_i is the numerical value of the coordinate (L, a, and b) in CIE-Lab for each color i present in the context.

w_i is the total percentage of the color i in the context.

N is the total number of colors.

Table 4.3 displays the mathematical representation of the present colors in the context of Figure 4.19 (b), in the format of CIE-Lab.

Table 4. 3 Mathematical representation of the colors in figure 4.19(b)

Color	Green	Yellow	Red	Gray	White
L	87.73	97.13	53.23	76.97	100
a	-86.18	-21.55	80.109	0.004	0.005
b	83.181	94.482	67.220	-0.008	-0.010

The results of the three weighted averages (L_{avg} , a_{avg} and b_{avg}) representing the average color in the context of the Figure 4.19 (b) are:

$$L_{avg} = \frac{(87.73 * 0.146) + (97.13 * 0.021) + (53.23 * 0.062) + (76.97 * 0.052) + (100 * 0.719)}{0.146 + 0.021 + 0.062 + 0.052 + 0.719} = 94.05$$

$$a_{avg} = \frac{(-86.18 * 0.146) + (-21.55 * 0.021) + (80.109 * 0.062) + (0.004 * 0.052) + (0.005 * 0.719)}{0.146 + 0.021 + 0.062 + 0.052 + 0.719} = -8.06$$

$$b_{avg} = \frac{(83.181 * 0.146) + (94.482 * 0.021) + (67.220 * 0.062) + (-0.008 * 0.052) + (-0.010 * 0.719)}{0.146 + 0.021 + 0.062 + 0.052 + 0.719} = 18.28$$

In this case the obtained coordinates of the average color within the context are: 94.05, -8.06 and 18.28. The location of the coordinates of the average color of the context as well as the coordinates of the colors present in the context have been represented graphically in Figure 4.21.

Figure 4. 21 Coordinates of the average color in Cie-Lab

To represent the average color in the context we use the variable (*cctx*), see Figure 4.22.

Figure 4. 22 Average of the color in the context

Distance in color with respect to the context

The following process in the computational model of contextual saliency is the calculation of the distance in color (*dc*), between the average color of the context and the individual color of all content objects in our virtual environment. Once we have obtained the coordinates of the

average color of the context, the saliency in color of an object in the environment with respect to the context will be computed as the Euclidean distance between both colors.

This is

$$dc = cctx - ic \quad \forall \text{ object from the point of view of the user}$$

Where

dc is the distance in color (also called color difference).

$cctx$ is the average color of the context (coordinates in CIE-Lab).

ic is the color of the object (as computed in the color characterization of the object).

The equation (14) allows measuring this difference between two colors.

$$cd = \Delta E_{CIE} = \sqrt{(L_2 - L_1)^2 + (a_2 - a_1)^2 + (b_2 - b_1)^2} \quad (14)$$

Where

$dc = \Delta E_{CIE} = \Delta E^*$ defines the magnitude of the difference between two colors CIE-Lab, that is, the distance between colors.

Table 4.4 shows the obtained distance results for the context of figure 19(b).

Table 4. 4 Color difference

Object	Average color of the context ($cctx$)			Individual color of the object (ic) in CIE-Lab			Color difference (cd)
	L_2	a_2	b_2	L_1	a_1	b_1	ΔE_{CIE}^*
Obj(1)	94.05	-8.06	18.28	87.73	-86.18	83.18	101.76
Obj(2)	94.05	-8.06	18.28	97.13	-22.55	94.48	77.63
Obj(3)	94.05	-8.06	18.28	53.23	80.10	67.22	108.78
Obj(4)	94.05	-8.06	18.28	87.73	-86.18	83.18	101.76
Obj(5)	94.05	-8.06	18.28	76.97	0.004	-0.008	26.29

These results indicate that the color with less difference is the grey, since this one is the closest to the average color of the context. The objects with green color have greater presence in the environment. However, the color with greater distance with respect to the context was undoubtedly the red color. A color that for the human eye is surely very salient. With the calculation of the color distance, the colors with higher contrast in the context can be identified, that is, the colors that stand out the most within the environment. The color distance with respect to the context is represented in the computational model by the variable (cd), see Figure 4.23.

Figure 4. 23 Distance in color

Saliency of color with respect to the context

To determine mathematically the value of the saliency per color, size and shape of the all objects of the virtual environment with respect to the context, we propose a process which is based on the following operations:

- *Standardizing the distribution:* Assuming that the distances in color (size and shape) with respect to the context are normally distributed, we will transform these distances into their corresponding values on a standard normal distribution.

All normal distributions are the same if we use the measurement σ units around its mean μ which is the centre. The process to change our distribution to these variables is

known as standardization. It is important to know that from any variable X that follows a distribution $N(\mu, \sigma)$, we can get another Z characteristic with a standard normal distribution, using the equation (15) (Pértegas and Pita, 2001):

$$Z = \frac{X - \mu}{\sigma} \quad (15)$$

The standard normal distribution (also called typified or reduced) has as half the zero value ($\mu = 0$), and as standard deviation the value one ($\sigma = 1$). The Figure 4.24 illustrates this distribution.

Figure 4. 24 Standard normal distribution

This transformation allows us to determine the degree of abnormality of a certain distance in color (size or shape) among the distances of all present objects within the context. This value will be used in Chapter 6 to determine certain criteria in the process of generation of directions (see section 6.3 Algorithm for the generation of directions when the OL is in the user's field of view).

- *Using a proportional distribution for determining the normalized saliency value.* The object with the highest standard distance value will have the value of saliency '1', the object with the lowest standard distance value will have the value of saliency '0', and the rest will be weighted in proportion to their values. To apply this criterion of normalization we will use the following equation (16):

$$K_x = \frac{x + |min|}{|max| + |min|} \quad (16)$$

Where

K or $(ScC_{[0-1]})$ is the normalized value of saliency of an object with respect to the context.

x is the value of the standard distance score to be normalized.

min is the minimum value of the standard scores.

max is the maximum value of the standard scores.

Table 4.5 shows the obtained results for the standard scores and the saliency values for the color differences of the objects 1 to 5 of figure 4.19 (b). Mathematically the object 3 is the most salient and the object 5 is the least salient.

Table 4. 5 Standard scores and saliency in color for the scenario of Figure 19(b)

Object	Color difference (cd)	Standard score	Saliency in color with respect to the context
Obj(1)	101.76	0.61	0.91
Obj(2)	77.63	-0.18	0.62
Obj(3)	108.78	0.84	1.00
Obj(4)	101.76	0.61	0.91
Obj(5)	26.29	-1.88	0.00

The value of saliency in color with respect to the context is represented with the variables $(ScC_{ss[-3 - 3]})$ when the saliency is considered from the standard score, and $(ScC_{[0-1]})$ when the saliency is considered from the proportional distribution, see Figure 4.25.

Figure 4. 25 Saliency by color with respect to context

This procedure is used to determine the saliency of an object with respect to the context, for its three physical characteristics (color, size, shape).

Average size of the context

To calculate the average size of the objects within the context, we will be using the same pseudo-rendering process that we proposed for the calculation of the visible surface of each object. The size or visible surface of an object is obtained from the number of rays that impact in that object. With the same equation (11) we compute the size of each object within the context as a percentage of the pixels in the screen.

Then, the average size of the structural objects in the context is calculated. The result of the average size of objects in Figure 4.19 (b) is:

$$sictx = \frac{(12.5) + (2.1) + (2.1) + (6.2) + (5.2)}{5} = 5.6$$

To represent the average size in the context we use the variable (*sictx*), see Figure 4.26.

Figure 4. 26 Average size in the context

Distance in size with respect to the context

The following process in the computational model of contextual saliency is the calculation of the distance in size (dsi), between the average size of the context and the individual size (in percentage) of all content objects in our virtual environment. Once we have obtained the average size of the context, the saliency in size of an object in the environment with respect to the context will be computed as the Euclidean distance between both sizes.

This is

$$dsi = sictx - isi \quad \forall \text{ object from the point of view of the user}$$

Where

dsi	is the distance in size (also called size difference).
$sictx$	is the average size of the structural objects in the context.
isi	is the individual size of the object (is equal to the visible surface of object number (vso)).

The same equation (14) allows measuring the difference between two sizes.

Table 4.6 shows the obtained distance results for the context of figure 19(b).

Table 4. 6 Size difference

Object	Size difference (dsi)
Obj(1)	6.9
Obj(2)	3.5
Obj(3)	3.5
Obj(4)	0.6
Obj(5)	0.4

The variable representing in the model the distance in size of an object with respect to the context is (*dsi*), see Figure 4.27.

Figure 4. 27 Distance of size

Saliency of size with respect to context

The value of saliency of size of an object with respect to the context is calculated with the transformation into standard scores and then using equation (16).

Table 4.7 shows the obtained results for the standard scores and the saliency values in size for the objects 1 to 5 of figure 4.19 (b).

Table 4. 7 Values of standard scores and saliency in size

Object	Size (visible surface) of object in context	Standard score	Saliency in size with respect to context
Obj(1)	6.9	1.65	1.00
Obj(2)	3.5	0.22	0.23
Obj(3)	3.5	0.22	0.23
Obj(4)	0.6	-1.00	-0.42
Obj(5)	0.4	-0.19	-0.46

Mathematically the object 1 is the most salient and the object 5 is the least salient, within the context. $(SsiC_{ss[-3 - 3]})$ represents the saliency value by size with respect to context from the standard score, and $(SsiC_{[0-1]})$ represents the saliency of size with respect to the context value from the proportional distribution, see Figure 4.28.

Figure 4. 28 Saliency of size with respect to the context

Average shape of the context

In this model of contextual saliency we decided to treat the shape of an object considering its three dimensions, unlike the color and size features which were treated from a two dimensional perspective. In this sense the Gestalt theory mentions that the human mind transforms and organizes the elements perceived of something to be integrated into a whole. This theory also extends to the organization of shapes and their meanings (Pinna, 2010).

The shape of an object is a basic feature that allows identifying and distinguishing one object from another, from an individual perspective and in a context. The shape of an object can give us information on their function and use. There are objects like a box or a car whose functional properties are clearly related to its perceptible shape and structure.

On the other hand, the shape of natural objects such as trees, flowers, rocks, etc., slightly change within their environments, at least those that tend to grow, change position or get worn out physically. In summary, the shape an object provides important information to recognize objects on future occasions or to determine if an object is similar to others within a real or virtual environment.

To calculate the average shape of the context, the vector of the shape of the context is obtained with the following method, which includes three steps:

- (1) The objects of the virtual environment are voxelized.
- (2) The 3D Zernike descriptor is applied to all voxelized objects with an order 8, to generate a vector of 25 numeric values.
- (3) Then, the values of all vectors are added and averaged in each of its components. The final result of this process is the vector of the context (V_{cont}) with 25 values.

Figure 4.29 illustrates a scene with six objects.

Figure 4. 29 View of a scene with regular and irregular objects

Having voxelized the six objects of the Figure 4.29, the Table 4.8 presents in the first six columns the set of 25 numeric values obtained for the 3D Zernike descriptor of each object. The seventh column shows the average shape vector for the context.

Table 4. 8 Values of 3D Zernike descriptors

Descriptor components	Obj(1) V_1	Obj(2) V_2	Obj(3) V_3	Obj(4) V_4	Obj(5) V_5	Obj(6) V_6	Vcont
1	2.31E+37	2.08E+37	5.99E+26	1.95E+37	1.08E+37	4.78E+26	1.24E+37
2	3.30E+21	5.25E+21	8.14E+10	1.75E+21	2.55E+21	1.35E+10	2.14E+21
3	2.57E+37	3.34E+37	1.37E+27	3.33E+37	2.06E+37	1.09E+27	1.88E+37
4	2.58E+37	3.34E+37	1.37E+27	3.33E+37	2.06E+37	1.09E+27	1.89E+37
5	4.14E+36	1.18E+36	2.21E+24	6.73E+33	1.45E+33	1.30E+24	8.89E+35
6	4.48E+36	1.31E+36	2.41E+24	9.72E+33	2.66E+33	1.42E+24	9.68E+35
7	5.14E+36	2.58E+37	2.13E+27	2.88E+37	2.25E+37	1.70E+27	1.37E+37
8	6.61E+36	2.63E+37	2.13E+27	2.88E+37	2.25E+37	1.70E+27	1.40E+37
9	7.71E+36	2.63E+37	2.13E+27	2.89E+37	2.25E+37	1.70E+27	1.42E+37
10	8.12E+36	3.34E+36	1.00E+25	2.10E+34	5.46E+33	6.28E+24	1.91E+36
11	9.76E+36	3.89E+36	1.08E+25	3.12E+34	1.02E+34	6.78E+24	2.28E+36
12	9.86E+36	3.89E+36	1.08E+25	3.14E+34	1.02E+34	6.79E+24	2.30E+36
13	7.53E+36	8.96E+36	2.88E+27	1.25E+37	1.68E+37	2.30E+27	7.62E+36
14	8.36E+36	1.25E+37	2.88E+27	1.25E+37	1.68E+37	2.30E+27	8.34E+36
15	1.29E+37	1.30E+37	2.88E+27	1.30E+37	1.68E+37	2.30E+27	9.28E+36
16	1.30E+37	1.30E+37	2.88E+27	1.30E+37	1.68E+37	2.30E+27	9.30E+36
17	4.43E+36	3.75E+36	2.46E+25	2.73E+34	9.85E+33	1.64E+25	1.37E+36
18	7.58E+36	4.90E+36	2.58E+25	4.43E+34	1.94E+34	1.74E+25	2.09E+36
19	8.13E+36	4.96E+36	2.59E+25	4.63E+34	1.97E+34	1.74E+25	2.19E+36
20	8.16E+36	4.96E+36	2.59E+25	4.63E+34	1.97E+34	1.74E+25	2.20E+36
21	4.87E+36	3.09E+36	3.61E+27	3.03E+36	7.21E+36	2.88E+27	3.03E+36
22	5.46E+36	9.63E+36	3.61E+27	3.03E+36	7.21E+36	2.88E+27	4.22E+36
23	1.07E+37	1.19E+37	3.61E+27	8.05E+36	7.69E+36	2.88E+27	6.38E+36
24	1.13E+37	1.19E+37	3.61E+27	8.09E+36	7.69E+36	2.88E+27	6.49E+36
25	1.13E+37	1.19E+37	3.61E+27	8.09E+36	7.69E+36	2.88E+27	6.50E+36

The average shape of the context is characterized with the vector of the context and is represented in this model with the variable (*shctx*), see Figure 4.30.

Figure 4. 30 Average shape in the context

Distance in Shape with respect to the context

To obtain the distances in shape the Euclidean distance is used. This distance is calculated between the components of the obtained 3D Zernike descriptor vectors of each object, and the components of the vector of the context. To perform this calculation the following equation is applied (17):

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_{25} - q_{25})^2} =$$

(17)

$$\sqrt{\sum_{i=1}^{25} (p_i - q_i)^2}$$

Where

d_E is the absolute value of the standard Euclidean distance between two vectors (object and context).

p is the set of numerical values of a vector that represents the shape of an object within the context.

q is the set of numerical values of the vector that represents the average shape of the context.

Table 4.9 presents the results of the Euclidean distances for the six objects in the scenario of figure 4.29.

Table 4. 9 Values of Euclidean distances in shape for the scenario of figure 4.29

Distance in shape (<i>dsh</i>)	
$d_1 = V_1 - V_{cont} $	2.76E+37
$d_2 = V_2 - V_{cont} $	3.37E+37
$d_3 = V_3 - V_{cont} $	4.40E+37
$d_4 = V_4 - V_{cont} $	3.52E+37
$d_5 = V_5 - V_{cont} $	2.36E+37
$d_6 = V_6 - V_{cont} $	4.40E+37

The variable used to represents this value is (*dsh*), see Figure 4.31.

Figure 4. 31 Distance of the shape

The values of table 4.8 also allow the calculation of the average distance in shape. For this process is proposed the following equation (18):

$$d_{avg} = \frac{\sum_{i=1}^n d_i}{n} \quad (18)$$

Where

d_{avg} is the average distance in shape between the objects of the context with respect to the vector of the context.

$\sum_{i=1}^n d_i$ is the sum of all distances of the objects in the context with respect to the vector of the context.

n is the number of objects in the context.

The average distance in shape for the set of objects in the scenario of figure 4.29 is:

$$d_{avg} = \frac{(2.76E+37) + (3.37E+37) + (4.40E+37) + (3.52E+37) + (2.36E+37) + (4.40E+37)}{6} = 3.47E+37$$

Saliency of shape with respect to the context

The value of saliency of shape of an object with respect to the context is calculated with the transformation into standard scores and then using the equation (16).

Table 4.10 shows the obtained results for the standard scores and the saliency values for the shape distance of the objects 1 to 6 of figure 4.29.

Table 4. 10 Values of standard scores and saliency of shape

Object	Distance in shape	Standard scores	Saliency of shape with respect to context
Obj(1)	2.76E+37	-1.84	0.20
Obj(2)	3.37E+37	-1.04	0.50
Obj(3)	4.40E+37	0.31	1.00
Obj(4)	3.52E+37	-0.85	0.57
Obj(5)	2.36E+37	-2.37	0.00
Obj(6)	4.40E+37	0.31	1.00

Mathematically the objects 3 and 6 are the most salient in shape and the object 5 is the least salient. ($SshC_{ss_{l-3-3j}}$) represents the value of saliency of shape with respect to the context

from the standard scores, and $(SshC_{[0-1]})$ represents the value of saliency of the shape with respect to the context from the proportional distribution, see Figure 4.32.

Figure 4. 32 Saliency of shape with respect to the context

Calculating the Contextual saliency of the 3D object

The calculation of the saliency with respect to the context for each object within the scenario is made also through a weighted sum. It was decided to establish a weight of 0.5 for the color, 0.3 for the size, and 0.2 for the shape. The following equations (19a and 19b) are then proposed to calculate the contextual saliency of a 3D object based on the standard normal distribution and the proportional distribution, respectively.

$$SoC_{SS} = (ScC_{SS[-3-3]} * 0.5) + (SsiC_{SS[-3-3]} * 0.3) + (SshC_{SS[-3-3]} * 0.2) \quad (19a)$$

$$SoC = (ScC_{[0-1]} * 0.5) + (SsiC_{[0-1]} * 0.3) + (SshC_{[0-1]} * 0.2) \quad (19b)$$

Where SoC_{SS} represents the saliency with respect to the context from the standard score, which includes $ScC_{TP[-3-3]}$, $SsiC_{TP[-3-3]}$ and $SshC_{TP[-3-3]}$ corresponding to the saliency of color, size and shape of each object; SoC represents the saliency with respect to the context

from the proportional distribution, which also includes $ScC_{[0-1]}$, $SsiC_{[0-1]}$ and $SshC_{[0-1]}$ corresponding to the saliency of color, size and shape of each object, see Figure 4.33.

Figure 4. 33 Saliency with respect to the context

4.4 General Saliency

Finally, the general saliency (also called global saliency) of an object, in this computational model of saliency, is calculated as the sum of one third of the value obtained in the individual saliency plus two thirds of the value in the saliency with respect to the context. This is best expressed by the following linear equations (20a and 20b).

$$GS_{ss} = \left(\left(\frac{1}{3} \right) (IS) + \left(\frac{2}{3} \right) SoC_{ss} \right) \quad (20a)$$

$$GS = \left(\left(\frac{1}{3} \right) (IS) + \left(\frac{2}{3} \right) (SoC) \right) \quad (20b)$$

Where GS represents the general saliency of a 3D object in a virtual environment in this computational model of saliency (see figure 4.34).

Figure 4. 34 General saliency

4.5 Empirical Validation of the Saliency Model

In this section are presented the results of the validation processes carried out to test the validity of the computational model of perceptual saliency that was proposed in this chapter. Three experiments were designed and are described, including their motivation and goals, the design of the experiment, the compilation of empirical information, the analysis of the collected information and the discussion of the results. At the end of each experiment conclusions are exposed.

4.5.1 Experimental evaluation of the shape saliency metric

Given that, this metric is an entirely novel contribution and is not based on previous results of other authors, it is necessary to verify that the metric is really valid, before going on to check the validity of the model as a whole. For this, an experiment was designed and carried out with the aim to investigate if people's perception of saliency of shape in 3D objects is similar to the measure provided by the shape saliency metric proposed in this work (see Lara et al., (2015)).

a) Method

Participants. Forty undergraduate students of the *Escuela Técnica Superior de Ingenieros Informáticos* of the *Universidad Politécnica de Madrid*, thirty male and ten female, with ages in the range of 18 to 25 years, voluntarily participated.

Materials, devices and situation. The experiment was carried out in a laboratory with suitable lighting condition. Each participant's session was run in a SONY laptop computer, model VGN-CS270T, with an Intel ® Core ^(TM)2 Duo processor, CPU P8600, 2.40 GHz, and 4.00 GB memory, using a mouse.

A computational application was developed to implement and test the metric using the Unity 3DTM platform, with some scripts created in the C# programming language. The voxelization process of the 3D objects was made through a script of voxelization in Unity 3DTM adjusted to extract the number of voxels and so calculate the volume of each object.

Results were automatically recorded in a .csv (comma-separated values) file; afterwards statistical analysis was conducted using the SPSSTM (Statistical Product and Service Solutions) application.

Design and procedure. The designed experimental system consisted of twenty-five scenarios (also called trials), each with four 3D objects. One hundred 3D objects were extracted randomly from the Princeton Shape Benchmark (PSB), which contains 1,814 3D models and is publicly available in the World Wide Web (Shilane et al., 2004). The selected objects for this experiment were adjusted to have the same size. Each of the objects was voxelized and the number of voxels was counted, the time for the voxelization process of each object varied from 0.405 to 1.097 seconds.

Participants were informed that their task during the experiment was to place the four objects provided on each trial on an empty platform in front of them, as shown in Figure 4.35, ordering them from left to right according to their shape saliency; this concept was explained to them as “the capability of the object's shape to attract their attention”. Therefore, the most

shape striking object should be placed to the left. Participants were given a brief demonstration of the system on how to place each object on the platform. Also, it was explained to participants that they could make as many changes as they required, before they confirmed the final order of the objects on each trial. Participants were asked to provide basic personal information as their age and gender, within the system. Each person lasted about 8 minutes to complete the twenty-five trials.

Figure 4. 35 View of an experimental system trial

b) Statistical Analysis of Results

In the first experiment several statistical analyses were performed in order to evaluate how valid is the metric with respect to the perception of saliency of the participants. The following three aspects were analysed:

1. The choice of the first or second object as the most salient. Because the saliency of the shape of an object is a very subjective characteristic, and given that there is no reference order with which to compare the order given by our metric (see Table 4.11), the comparison with each of the given orders of the participants (see Table 4.12 with a sample of a participant) required to be flexible enough to admit variations, but at the same time capable to give information about the performance of the metric. It was decided to measure the extent to which the first or second most salient objects, according to the metric of saliency, matched the object placed by the participants in the first position. This condition

helped to identify if the most salient objects to the human visual system correspond with high-valued objects by our metric.

Table 4. 11 Order and value of saliency given by our metric, to the four objects of the twenty-five trials

Number of trial	First object	Second object	Third object	Fourth object
1	(*) Lamp	Sextant	Antenna	Dice
	(**) 1	0.97	0.88	0.12
2	ShaveMachine	Apollo13	Lollypop	Diskette
	1	0.82	0.81	0.54
3	Mallet	Apple	Shovel	Door
	1	0.98	0.93	0.87
4	Sofa	MayaPiramid	Arrow	Drum
	1	0.99	0.99	0.86
5	Spider	Microscope	Axe	Drumbell
	1	0.97	0.88	0.72
6	EnterpriseShip	MobilePhone	Bat	SteakKnife
	1	0.82	0.79	0.73
7	MobilePhone	Extinguisher	BeerBottle	Spray
	1	0.9	0.52	0.47
8	Bicycle	F16Plane	Motorcycle	StreetLantern
	1	0.98	0.94	0.88
9	Submarine	PailCube	Binoculars	FaxMachine
	1	0.92	0.87	0.86
10	Fence	PanelScreen	Table	Biotank
	1	0.99	0.98	0.83
11	FerrariF380	Pear	TeaCup	BitDrill
	1	1	0.89	0.83
12	BoeingPlane	Pencil	FishTank	Television
	1	0.51	0.5	0.5
13	TeremeoLamp	Piano	Book	FlashLight
	1	0.86	0.72	0.71
14	Flute	PicnicBench	TheetBrush	Briefcase
	1	0.97	0.76	0.3
15	PipeSmoke	TieFighter	C64Chip	GarbageCan
	1	0.96	0.63	0.41
16	PositionalLight	ToyBear	GasMark	Camera
	1	0.93	0.92	0.73
17	Glasses	Candle	PsxController	Train

	1	0.96	0.73	0.72
18	Hammock	Tricycle	CandyPot	Revolver
	1	0.94	0.87	0.84
19	Handbell	Truck	CanpyBed	Dice
	1	0.99	0.67	0.11
20	Hat	Umbrella	RoadsterCar	CashRegister
	1	0.97	0.93	0.63
21	Headset	Unicycle	RocketMMA-1	CasketBox
	1	0.99	0.93	0.53
22	Hind24H-D	Chair	Violin	Rollerball
	1	0.93	0.81	0.78
23	Rose	CleaverKnife	Wagon	Hourglass
	1	0.75	0.73	0.73
24	Hydrant	WashingMachine	CPU	RugbyBall
	1	0.92	0.74	0.61
25	SaberSword	X-Wing	Keyboard	Cross
	1	0.94	0.79	0.52

(*) Order and (**) Value in saliency

Table 4. 12 Sample of the order given for a participant, to the four objects of the twenty-five trials

Number of trial	First object	Second object	Third object	Fourth object
1	Sextant	Lamp	Antenna	Dice
2	Lollypop	ShaveMachine	Apollo13	Diskette
3	Door	Shovel	Mallet	Apple
4	Arrow	Sofa	Drum	MayanPyramid
5	Spider	Microscope	Axe	Drumbell
6	EnterpriseShip	MobilePhone	SteakKnife	Bat
7	MobilePhoneShell	Extinguisher	Spray	BeerBottle
8	F16Plane	Motorcycle	Bicycle	StreetLantern
9	Submarine	Binoculars	FaxMachine	PailCube
10	Biotank	PanelScreen	Fence	Table
11	FerrariF380	BitDrill	Pear	TeaCup
12	BoeingPlane	Pencil	Television	FishTank
13	FlashLight	TeremeoLamp	Piano	Book
14	Flute	TheetBrush	PicnicBench	Briefcase
15	TieFighter	PipeSmoke	C64Chip	GarbageCan
16	GasMask	ToyBear	PositionalLight	Camera
17	PsxController	Candle	Glasses	Train

18	Tricycle	Revolver	Hammock	CandyPot
19	CanpyBed	Handbell	Truck	Dice
20	RoadsterCar	Umbrella	CashRegister	Hat
21	Unicycle	RocketMMA-1	Headset	CasketBox
22	Hind24H-D	Chair	Violin	Rollerball
23	Rose	Wagon	Hourglass	CleaverKnife
24	Hydrant	RugbyBall	WashingMachine	CPU
25	X-Wing	SaberSword	Keyboard	Cross

2. Simple error, defined as the comparison of the statistical mode with respect to the metric. In this second statistical analysis, it was compared for each trial the order given by our metric. The mode, or the most voted object in each position for each trial, was obtained, counting the number of times that an object was placed in each of the four positions within each trial. Then, was computed the distance between each pair of objects (the one given by the metric and the most voted one) in each position of each trial, according to the value assigned to each object by our metric. Finally, to get the value of the simple error (SE) we add the distances in saliency for all positions of each trial. We propose simple error as a measure of the distance between the order provided by the saliency metric and the most general opinion of participants.

In order to further evaluate the validity of the metric, two additional variables were computed: (EM) the error in movements (or jumps) when comparing the order given by each participant in each trial, with respect to the order given by our metric; and (ES) the error in saliency when comparing the order given by each participant in each trial, with respect to the order given by our metric.

For obtaining these two values we measured the distance in absolute value of the saliency between each pair of objects of each trial, according with the value of our metric of saliency and the order given by participants. Table 4.13 presents a sample of a trial with these two measures corresponding to a participant.

Table 4. 13 Sample of the order given by a participant, to the four objects of a trial and their values of EM and ES

Number of trial	First object	Second object	Third object	Fourth object	Values	
1	Lamp	Sextant	Antenna	Dice		Order given by our metric
	Sextant	Lamp	Antenna	Dice		Order given by the participant
	1	0.97	0.88	0.12		Saliency given by our metric
	0.97	1	0.88	0.12		Saliency according to the order of participant
	1	1	0	0	2	Error in movements (EM)
	0.03	0.03	0	0	0.06	Error in saliency (ES)

The values (EM) and (ES) are calculated once per participant in his 25 trials of the test. Therefore 40 participants x 25 trials = 1 000 values obtained.

Based on these variables we made two correlation analysis: (1) between the value of the simple error (SE) of each trial and the standard deviation of error in movements (EM) for all participants in each trial; and (2) between the value of the simple error of each trial and the standard deviation of the error in saliency (ES) for all participants in each trial. We want to explore if higher diversity between participants leads to higher simple error (in other words if we can expect that trials in which the distance between the participant's order and the metric's order is more variable will also be the trials in which the order generated by the metric is less prototypical).

3. Comparison between the simple error and the possible diversity in saliency of all trials. This last statistical analysis sought to evaluate a new hypothesis: the greater the diversity of the trial in saliency, the lower the simple error. It was expected that participants would adhere more closely to the reference order provided by the metric when there was a significant difference in saliency between the objects presented in the trial. To test this hypothesis the

correlation between these two variables was calculated. Results are discussed in the following section.

For obtaining the value of the diversity in saliency of a trial, we calculated the sum of all possible differences in saliencies (see Equation (21)).

Diversity in Saliency

$$= ((IS_1 - IS_2) + (IS_2 - IS_3) + (IS_3 - IS_4) + (IS_2 - IS_3) + (IS_1 - IS_4) + (IS_2 - IS_4)) \quad (21)$$

Where

$IS_1 \dots IS_4$ is the individual saliency of the objects of a trial.

For example diversity in saliency of the trial 1 is 2.73 (see Table 4.14).

Table 4. 14 Calculation of the diversity in saliency of a trial

Number of trial	First object	Second object	Third object	Fourth object	
1	Lamp	Sextant	Antenna	Dice	
	1	0.97	0.88	0.12	Saliency given by our metric
	$(1 - 0.97) + (0.97 - 0.88) + (0.88 - 0.12) + (0.97 - 0.88) + (1 - 0.12) + (0.97 - 0.12) = 2.73$				(diversity in saliency of trial 1)

c) Results

The results of the first statistical analysis indicated that the metric effectively predicts the objects that humans tend to perceive as salient.

The number of possible orders for the 4 objects of each trial is 24, but only half of them (12) would fit with the restriction established in the first statistical analysis, that is, to include the first or second most salient objects (according to the metric) in the first position. Assuming that participants provided random orders for each trial, it would be expected that 50% of the

orders provided would fit this condition. However, with the orders given by the 40 participants in each of the 25 trials, a mean of 16.5 matching orders was obtained. These 16.5 matches represent a 66% of all provided orders, exceeding the expected 50% value in case of randomness. Therefore it can be said that our metric certainly measures the shape saliency of objects.

The results in the second statistical analysis showed a degree of 37% agreements between the order based on the mode and the reference order given by the metric. This percentage demonstrates that the metric is a good representative of the saliency perception in a prototypical person (Table 4.15 (a) order for each trail based on our metric and (b) order based on the mode).

Table 4. 15 Order of the objects in each trial: (a) based on our metric of saliency and (b) based on the mode.
(When the order is the same in both, the cell is highlighted in green)

(a)				
Trial	Pos_1	Pos_2	Pos_3	Pos_4
Trial_1	Lamp	Sextant	Antenna	Dice
Trial_2	ShaveMac	Apollo13	Lollypop	Diskette
Trial_3	Mallet	Apple	Shovel	Door
Trial_4	Sofa	MayaPira	Arrow	Drum
Trial_5	Spider	Microsco	Axe	Drumbell
Trial_6	Enterprise	MobilePh	Bat	SteakKni
Trial_7	MobilePh	Extinguis	BeerBottl	Spray
Trial_8	Bicycle	F16Plane	Motorcyc	StreetLan
Trial_9	Submarine	PailCube	Binocular	FaxMach
Trial_10	Fence	PanelScre	Table	Biotank
Trial_11	FerrariF3	Pear	TeaCup	BitDrill
Trial_12	BoingPlan	Pencil	FishTank	Televisio
Trial_13	TeremeoL	Piano	Book	FlashLigh
Trial_14	Flute	PicnicBe	TheetBru	Briefcase
Trial_15	PipeSmok	TieFighte	C64Chip	GarbageC
Trial_16	Positional	ToyBear	GasMark	Camera
Trial_17	Glasses	Candle	PsxContr	Train
Trial_18	Hammock	Tricycle	CandyPo	Revolver
Trial_19	Handbell	Truck	CanpyBe	Dice
Trial_20	Hat	Umbrella	Roadster	CashRegi
Trial_21	Headset	Unicycle	RocketM	CasketBo
Trial_22	Hind24H-	Chair	Violin	Rollerball
Trial_23	Rose	CleaverK	Wagon	Hourglass
Trial_24	Hydrant	Washing	CPU	RugbyBal
Trial_25	SaberSword	X-Wing	Keyboard	Cross

(b)

Trial	Pos_1	Pos_2	Pos_3	Pos_4
Trial_1	Sextant	Antenna	Lamp	Dice
Trial_2	Apollo13	ShaveMac	Lollypop	Diskette
Trial_3	Mallet	Door	Shovel	Apple
Trial_4	Drum	MayaPira	Arrow	Sofa
Trial_5	Spider	Microsco	Axe	Drumbell
Trial_6	Enterprise	MobilePh	SteakKni	Bat
Trial_7	MobilePh	Spray	Extinguis	BeerBottl
Trial_8	F16Plane	Motorcyc	Bicycle	StreetLan
Trial_9	Binocular	Submarine	FaxMach	PailCube
Trial_10	Biotank	PanelScre	Fence	Table
Trial_11	FerrariF3	Pear	TeaCup	BitDrill
Trial_12	BoingPlan	Televisio	Pencil	FishTank
Trial_13	FlashLigh	Piano	TeremeoL	Book
Trial_14	Flute	TheetBru	PicnicBe	Briefcase
Trial_15	TieFighte	PipeSmok	C64Chip	GarbageC
Trial_16	Positional	GasMark	ToyBear	Camera
Trial_17	PaxContr	Candle	Glasses	Train
Trial_18	Tricycle	Revolver	Hammock	CandyPo
Trial_19	CanpyBe	Truck	Handbell	Dice
Trial_20	CashRegi	Roadster	Hat	Umbrella
Trial_21	Unicycle	CasketBo	RocketM	Headset
Trial_22	Hind24H-	Violin	Rollerball	Chair
Trial_23	Rose	Wagon	Hourglass	CleaverK
Trial_24	Hydrant	CPU	Washing	RugbyBal
Trial_25	X-Wing	Keyboard	SaberSword	Cross

It was also employed Pearson's correlation coefficients as indicators of the relationship between (1) the standard deviation of the error in movements (σ_{EM}) and the simple error (SE), and (2) the standard deviation of the error in saliency (σ_{ES}) and the simple error (SE). The values for these two correlations can be seen in Table 4.16 (a) and (b). According to these coefficients, both correlations are positive, but low. The sets of points in the scatter diagrams of the Figure 4.36 (a) (b) allow appreciating the linear relationship in both cases. In the case of Figure 4.36 (a) the correlation index was ($r = 0.14$), and in Figure 4.36 (b) of ($r = 0.27$).

Table 4. 16 Values for the correlations between (a) the standard deviation of the error in movements (σ_{EM}) and the simple error and (SE), and the standard deviation of the error in saliency (σ_{ES}) and the simple error (SE)

(a)			(b)		
Trial	σ_{EM}	SE	Trial	σ_{ES}	SE
T1	1.418	0.24	T1	0.608	0.24
T2	1.889	0.36	T2	0.295	0.36
T3	1.471	0.22	T3	0.056	0.22
T4	0.900	0.28	T4	0.044	0.28
T5	1.889	0	T5	0.187	0
T6	1.809	0.12	T6	0.170	0.12
T7	1.617	0.86	T7	0.392	0.86
T8	1.008	0.12	T8	0.040	0.12
T9	1.646	0.28	T9	0.090	0.28
T10	0.911	0.34	T10	0.051	0.34
T11	1.797	0	T11	0.143	0
T12	1.566	0.02	T12	0.399	0.02
T13	0.778	0.58	T13	0.109	0.58
T14	1.707	0.42	T14	0.487	0.42
T15	1.446	0.08	T15	0.324	0.08
T16	2.134	0.4	T16	0.178	0.4
T17	2.584	0.54	T17	0.311	0.54
T18	1.319	0.32	T18	0.075	0.32
T19	1.594	0.66	T19	0.556	0.66
T20	1.562	0.82	T20	0.202	0.82
T21	2.143	1.01	T21	0.412	1.01
T22	1.594	0.3	T22	0.133	0.3
T23	1.663	0.04	T23	0.236	0.04
T24	1.494	0.36	T24	0.218	0.36
T25	1.539	0.42	T25	0.300	0.42

Figure 4. 36 Dispersion diagrams showing the correlation between (a) the simple error and the standard deviation of the error in movements; (b) the simple error and the standard deviation of the error in saliency

For the third statistical analysis it was also used Pearson's correlation coefficients between (1) the simple error (SE) and the diversity in saliency by trial. The values for this correlation can be seen in Table 4.17 The result showed a weak to moderate value of correlation ($r = 0.28$) with a positive trend, as shown in Figure 4.37. This indicates that the relationship between variables is positive and low, and therefore the greater the diversity of the trial in saliency, the higher the value in the simple error. Here, our hypothesis has to be rejected since it was assumed that when the difference in saliency (also called diversity in saliency) between the objects of one trial is high, it would be easier to find agreement between the order of participants and the order generated by the metric, which is not happening.

Table 4. 17 Values for the correlations between the simple error and (SE), and the diversity in saliency by trial

Trial	SE	Diversity in saliency by trial
T1	0.24	2.73
T2	0.36	1.39
T3	0.22	0.44
T4	0.28	0.42
T5	0	0.93
T6	0.12	0.84
T7	0.86	1.97
T8	0.12	0.4
T9	0.28	0.47
T10	0.34	0.52
T11	0	0.62
T12	0.02	1.51
T13	0.58	1.01
T14	0.42	2.31
T15	0.08	2.1
T16	0.4	0.82
T17	0.54	1.07
T18	0.32	0.55
T19	0.66	2.99
T20	0.82	1.15
T21	1.01	1.47
T22	0.3	0.78
T23	0.04	0.83
T24	0.36	1.35
T25	0.42	1.59

Figure 4. 37 Dispersion diagram showing the correlation between the simple error and the diversity in saliency

4.5.2 Experimental evaluation of the computational model of individual perceptual saliency

With the aim to investigate if people's perception of saliency is similar to the computational model of saliency results it was designed and carried out the next experiment.

a) Method

Participants. A group of thirty people, undergraduate students, professors and administrative employees of the *Escuela Técnica Superior de Ingenieros Informáticos* of the *Universidad Politécnica de Madrid*, seventeen male and thirteen female, with ages in the range of 20 to 67 years, voluntarily participated.

Materials, devices and situation. The experiment was carried out in a laboratory with suitable lighting condition. Each participant's session was run in a SONY™ laptop computer, model VGN-CS270T, with a processor Intel ® Core™2 Duo CPU P8600, 2.40 GHz, and 4.00 GB memory, with a mouse.

An application was developed to measure 3D objects saliency according to the computational model here proposed. The game engine Unity 3D™ and a number of scripts in C# programming language were used. Particularly, for the 3D objects voxelization process, a script of Unity 3D™ was adjusted to extract the number of voxels and so calculating the volume of each object. Through the semantic editor-exporter (described in the Chapter 6) data was collected for the 3D objects in the scenes and exported to the ontology.

Data was statistically analyzed using the SPSS™ (Statistical Product and Service Solutions) application.

Design and procedure. From the Princeton Shape Benchmark (PSB), which contains 1,814 3D models and is publicly available (Shilane et al. 2004), one hundred 3D objects were randomly extracted for the previous experiment related to the saliency of shape (see (Lara et al., 2015)). Nine of them were selected as follows: three evaluated as with high saliency, three as having medium saliency, and three with low saliency, according to their shape (as measured by our metric). In Figure 4.38, the selected models and their given name in the PSB can be seen. In the first row are the ones with high saliency in shape (Ferrari F380, Spider and Mobile

Phone), in row two the medium ones (GarbageCan, C64Chip and BeerBottle) and in the last row the less salient in shape objects (Pencil, Dice and BriefCase).

Figure 4.38 The set of nine objects selected for the experiment

The nine selected 3D objects were resized, adjusting each of them to three different scales: 1 for a large size, 0.65 for medium and 0.25 for a small size, getting a set of 27 different objects which were used twice, making a total of 54 objects.

Two groups of colors were used in these 54 objects: the group formed by the eight most salient colors identified by (Choungourian, 1968); and a second group of eight colors randomly selected from the RGB system. In Figure 4.39 are presented the 16 colors, with their RGB and CIE-Lab systems number specifications.

Figure 4. 39 The colors used in the objects

The two same figure and size objects were colored: one with a salient color (from the first group of colors) and one with a random color (from the second group of colors), taking care that no color was repeated in the same figure object. Table 4.18 summarizes this information: in the first column is the name of the object, in the second column its size, and in the third and fourth columns the salient and the random colors used in the two same figure and size objects.

Table 4. 18 Colors and sizes used in the set of objects

Name	Size	Colors	
		Salient	Random
GarbageCan	Large	Purple	Random 8
	Medium	Red	Random 2
	Small	Yellow-green	Random 3
BeerBottle	Large	Blue	Random 1
	Medium	Purple	Random 5
	Small	Green	Random 2
C64Chip	Large	Orange	Random 3

	Medium	Red	Random 8
	Small	Yellow	Random 7
Dice	Large	Blue-green	Random 2
	Medium	Yellow-green	Random 4
	Small	Purple	Random 6
Ferrari F380	Large	Blue-green	Random 6
	Medium	Yellow	Random 4
	Small	Purple	Random 3
Pencil	Large	Blue	Random 1
	Medium	Orange	Random 5
	Small	Red	Random 4
BriefCase	Large	Orange	Random 5
	Medium	Red	Random 8
	Small	Purple	Random 4
Mobile Phone	Large	Green	Random 5
	Medium	Purple	Random 6
	Small	Yellow	Random 7
Spider	Large	Red	Random 4
	Medium	Blue-green	Random 2
	Small	Orange	Random 8

The snapshots of the 54 figures were randomly placed in slides of a PowerPoint™ presentation. Three different sequences were made, each presented to three groups of 10 persons. The experiment was organized of this way because we hoped that as objects are presented to the participant, he could go making comparisons with other objects that he has seen previously, so that the evaluations of the objects that he sees in the end, these will be more influenced by the objects previously seen. Figure 4.40 shows three of the slides, from the slide 20 to 22 of one of the three presentations sequences.

Figure 4.40 Three slides of the presentation of the objects for the participants

Participants were asked to provide basic personal information as their age and gender. Then they wrote their answers in a designed paper format with two parts: a semantic differential test and a selection test.

The participants were informed that their task was to observe a set of objects. Then they were instructed as follows: *“Imagine each object within a room among other objects, and then answer: ‘To what extent do you think that the object would attract your attention?’ by marking with a vertical line within the horizontal line of the object in a scale of 0-no attraction to 10-maximum attraction (the line was 10 cms long and had a 0 on the left and a 10 on the right, see Figure 4.41; which constitutes a semantic differential test). Then please mark with an X the feature or features you think are the most outstanding of the object (the options were: color ‘C’, size ‘Si’ and shape ‘Sh’), if you think it is none of them, then do not mark any”*. Figure 4.41 shows a part of the area where participants wrote their answers. Then the presentation was shown to them, and it was explained to use the forward arrow to see the next figures with the next instruction: *“Please do not go back on the presentation”*.

Each person lasted about 15 minutes to complete the test.

First part		Second part		
1)	MIN (0) _____ MAX (10)	☐	☐	☐
		C	Si	Sh
2)	MIN (0) _____ MAX (10)	☐	☐	☐
		C	Si	Sh
3)	MIN (0) _____ MAX (10)	☐	☐	☐
		C	Si	Sh

Figure 4. 41 View of the areas where participants wrote their answers

b) Data analysis

Data was treated for the two parts of the participants' answers as follows:

A set of continuous data from the semantic differential part of the test in the experiment was obtained. The lines marked by the participants were measured in centimetres and millimetres; with these values the participants expressed their opinions in relation to the saliency for each observed object. The average of these measures (AV) was then calculated for each object.

The frequency of each feature (color, size and shape) of the object that was marked by the participants was obtained from the second part of the test.

Additionally, the 54 objects were processed by the developed application to determine their saliency by color (Sc), size (Ssi), and shape (Ssh) and to obtain their individual saliency (IS) according to the given weights for each feature.

c) Results and Discussion regarding the saliency of object's features

First, are analysed the results of the second part of the test. In order to understand the measure of saliency given by the participants to each object and its relation with the outstanding objects' characteristics they marked, frequencies were obtained and separated in three ranges; with a minimum value of 2 and a maximum of 27, the ranges were labelled as:

- From 2 to 10 = "Low" range;

- From 11 to 18 = “Medium” range; and
- From 19 to 27 = “High” range

Table 4.19 summarizes the results of the second part by ranges of frequencies. In the first column are the names of object’s features and in the second the total number of objects in which this feature was present; columns 3, 4 and 5, present the number of objects for which the feature was selected as outstanding in a certain frequency range. For example, in the case of the Red color, for the 5 objects colored in red all the times there were at least 19 participants that marked the color feature as outstanding.

Table 4. 19 Evaluation of how outstanding is a feature

Most Salient Colors	Total	Low	Medium	High
Red	5	-	-	5
Orange	5	-	4	1
Yellow	2	-	1	1
Yellow - green	2	1	1	-
Green	2	-	1	1
Blue - green	3	1	2	-
Blue	2	-	1	1
Purple	6	-	3	3
Total of objects (27)		2 (7.4 %)	13 (48. 2 %)	12 (44. 4 %)
Less Salient Colors				
Random 1	2	2	-	-
Random 2	4	1	3	-
Random 3	3	3	-	-
Random 4	5	2	3	-
Random 5	4	3	1	-
Random 6	3	2	-	1
Random 7	2	1	-	1
Random 8	4	3	1	-
Total of objects (27)		17 (63 %)	8 (29. 6 %)	2 (7.4 %)
Total of objects with colors most and less salient		19(41.3 %)	21(28.3 %)	14(30.4 %)
Size				
Large	18	-	4	14
Medium	18	12	6	-
Small	18	11	7	-

Total of objects (54)		23 (42.6 %)	17 (31.5 %)	14 (25.9 %)
Shape				
GarbageCan	6	4	2	-
BeerBottle	6	6	-	-
C64Chip	6	5	1	-
Dice	6	5	1	-
Ferrari F380	6	1	5	-
Pencil	6	6	-	-
BriefCase	6	6	-	-
Mobile Phone	6	6	-	-
Spider	6	1	4	1
Total of objects (54)		40 (74.1 %)	13 (24.1 %)	1 (1.8 %)

Based on the information in Table 4.19, we can state that:

- In general, the shape is the characteristic that was found by the participants as less attractive, followed by size, and with the color as the most attractive, which is consistent with the literature review (Mangold, 1986). The shape of the objects was the feature with the highest number of objects valued in the low range (40 of 54 objects); the only object in which the shape was marked as outstanding with a high frequency was the Spider, but this occurred when it had large size and red color. The Ferrari F380 and the Spider had the major medium range values regarding the shape, while the rest of objects were mainly in the low range categorization. Moreover, these two objects were precisely those objects with more saliency value, given by the metric proposed.
- The eight colors proposed by (Choungourian (1968) were clearly more prominent to the participants compared with the Random colors. Most of them were marked in the medium or high range of frequencies (92,6%). The color that highly stands out was the red one, with no regard of the size or shape of the object. Two of the randomly selected colors had one object valued in the high range: Random 6 and Random 7; however, the objects in which the participants found them attractive were the Spider and the Mobile Phone, which have a high salient shape, with a large and medium size respectively as can be seen in Table 4.20. This suggests that participants find it

difficult to judge the saliency of object properties in isolation, and they tend to value all properties of a highly outstanding object as outstanding.

Table 4. 20 Results obtained in the model of individual saliency by color, size, shape ordered by individual saliency

Name of object	Color	Size	Saliency by color	Saliency by size	Saliency by shape	Individual Saliency
Ferrari F380	Blue-Green	Large	1.00	0.870	0.862	0.933
Spider	Red	Large	1.00	0.574	0.966	0.865
Mobile Phone	Green	Large	1.00	0.667	0.777	0.855
Dice	Blue-Green	Large	1.00	0.685	0.717	0.849
Ferrari F380 2	Yellow	Medium	1.00	0.429	0.850	0.799
GarbageCan	Purple	Large	1.00	0.520	0.696	0.795
Spider 2	Blue-Green	Medium	1.00	0.278	0.959	0.775
Mobile Phone 2	Purple	Medium	1.00	0.310	0.745	0.742
Dice 2	Yellow-Green	Medium	1.00	0.300	0.714	0.733
BeerBottle	Blue	Large	1.00	0.479	0.428	0.729
Spider 4	Orange	Small	1.00	0.110	0.939	0.721
Ferrari F380 4	Purple	Small	1.00	0.169	0.817	0.714
C64Chip	Orange	Large	1.00	0.150	0.833	0.712
Pencil	Blue	Large	1.00	0.384	0.462	0.708
BriefCase	Orange	Large	1.00	0.541	0.226	0.708
GarbageCan 2	Red	Medium	1.00	0.209	0.689	0.701
C64Chip 2	Red	Medium	1.00	0.073	0.787	0.679
Mobile Phone 4	Yellow	Small	1.00	0.121	0.708	0.678
Dice 4	Purple	Small	1.00	0.104	0.706	0.672
GarbageCan 4	Yellow-Green	Small	1.00	0.078	0.674	0.658
BeerBottle 2	Purple	Medium	1.00	0.219	0.426	0.651
Pencil 2	Orange	Medium	1.00	0.191	0.441	0.646
C64Chip 4	Yellow	Small	1.00	0.029	0.676	0.644
Pencil 4	Red	Small	1.00	0.115	0.528	0.640
BriefCase 2	Red	Medium	1.00	0.219	0.241	0.614
BeerBottle 4	Green	Small	1.00	0.085	0.441	0.614
BriefCase 4	Purple	Small	1.00	0.082	0.326	0.590
Ferrari F380 3	Random 4	Medium	0.33	0.695	0.850	0.545
Ferrari F380 1	Random 6	Large	0.17	0.870	0.862	0.517
Spider 1	Random 4	Large	0.33	0.380	0.966	0.474
GarbageCan 1	Random 8	Large	0.33	0.520	0.696	0.462
Mobile Phone 1	Random 5	Large	0.17	0.667	0.776	0.439

Dice 1	Random 2	Large	0.17	0.685	0.717	0.432
Dice 3	Random 4	Medium	0.33	0.300	0.714	0.400
BeerBottle 1	Random 1	Large	0.33	0.479	0.428	0.396
Spider 5	Random 8	Small	0.33	0.110	0.939	0.387
Pencil 1	Random 1	Large	0.33	0.384	0.462	0.374
Spider 3	Random 2	Medium	0.17	0.278	0.959	0.358
C64Chip 3	Random 8	Medium	0.33	0.073	0.787	0.346
Mobile Phone 3	Random 6	Medium	0.17	0.310	0.745	0.325
Pencil 5	Random 4	Small	0.33	0.115	0.528	0.307
BriefCase 1	Random 5	Large	0.17	0.541	0.226	0.291
GarbageCan 3	Random 2	Medium	0.17	0.209	0.689	0.284
BriefCase 3	Random 8	Medium	0.33	0.219	0.241	0.281
Mobile Phone 5	Random 7	Small	0.17	0.121	0.708	0.261
BriefCase 5	Random 4	Small	0.33	0.082	0.326	0.256
Dice 5	Random 6	Small	0.17	0.104	0.706	0.256
BeerBottle 3	Random 5	Medium	0.17	0.219	0.426	0.234
Pencil 3	Random 5	Medium	0.17	0.191	0.441	0.229
C64Chip 5	Random 7	Small	0.17	0.029	0.676	0.227
Ferrari F380 5	Random 3	Small	0.00	0.169	0.817	0.214
C64Chip 1	Random 3	Large	0.00	0.150	0.833	0.212
BeerBottle 5	Random 2	Small	0.17	0.085	0.441	0.197
GarbageCan 5	Random 3	Small	0.00	0.078	0.674	0.158

- Regarding the size of the objects, as expected, the large size had higher frequency ranges followed by the medium size and the small one.

To corroborate that the measure given by the participants to the objects' saliency was related with the marked features a multiple linear regression model was calculated, with the average (AV) of the measure given by the participants in the first part of the test as the dependent variable, and the frequencies with which the features (color, shape and size) were marked as attractive as independent variables. Results are shown in Figure 4.42; with an adjusted R^2 of .63, the variation in the measure given by the participants can be assumed as explained by the independent variables. The ANOVA corroborates that the correlation is not due to chance ($Sig < .000$ or $.003$). And the coefficient values indicate that the three

independent variables are significant for the model. However, based on their standardized coefficients, the size has a higher relative value than the color, contrary to what was expected.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.806 ^a	.650	.629	.78146

a. Predictors: (Constant), Shape, Size, Color

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	56.770	3	18.923	30.987	.000 ^a
	Residual	30.534	50	.611		
	Total	87.304	53			

a. Predictors: (Constant), Shape, Size, Color

b. Dependent Variable: AV

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.109	.540		.201	.841
	Color	.104	.020	.494	5.273	.000
	Size	.176	.019	.887	9.493	.000
	Shape	.076	.025	.266	3.088	.003

a. Dependent Variable: AV

Figure 4. 42 Results of the multiple regression model based in the participants' answers in the test

d) Computational model compared with the participants' appreciation of saliency

A Pearson's correlation coefficient was calculated to compare the average of the participants measure of saliency for each object (AV) with the metric value, giving a correlation index of ($r_p = 0.60$), Sig. = .000 significant at the 0.01 level 2-tailed; a scatter diagram is presented in Figure 4.43 to appreciate their linear relation. This metric value was calculated based on the weights given by us (see Equation (9)).

Figure 4.43 Dispersion diagram showing the correlation between the value of individual saliency and the average value of votes

Even though the participants marked the color as the more prominent characteristic of the objects, the statistical model reflects that the size had a higher weight in the correlation with their evaluation of saliency for the objects. When we adjust the weights in the computational model of individual saliency for the three characteristics, to the relative weight of the independent variables in the regression model, the correlation between the participants' measure of saliency and the computational model increases to ($r_p = 0.85$).

This better fit of the adjusted model with the participants' perception of saliency, probably obeys to the experimental conditions. Only one object was showed at a time to the participants in a white rectangular background (see Figure 4.40) with the intention to isolate it from distractions. This probably leads the participants to compare the size of the object to the size of the rectangular background, which acts as a context and helps the participants to evaluate the saliency of this feature. Moreover, the size is an ordinal variable that can be easily compared with the other objects presented in the experiment.

4.5.3 Experimental evaluation of the computational model of contextual saliency

With the aim to investigate if to greater salience of an object in context, less time it takes for the participants to locate it.

a) Method

Participants. Twenty eight people, undergraduate students and professors of the *Escuela Técnica Superior de Ingenieros Informáticos* of the *Universidad Politécnica de Madrid*, twenty two male and six female, with ages in the range of 20 to 45 years, who voluntarily participated.

The 28 users were divided into two groups (A) and (B), and every group was organized into seven different series of scenes, avoiding repetitions. Our experiment was organized in two groups of participants to increase the reliability of results. The order of the scenes presented in each series was varied as follows:

Series 1:

1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7

Series 2:

2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1

Series 3:

3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2

Series 4:

4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3

Series 5:

5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4

Series 6:

6 > 7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5

Series 7:

7 > 1 > 2 > 3 > 4 > 5 > 6 > 7 > 1 > 2 > 3 > 4 > 5 > 6

Materials, devices and situation. The experiment was carried out in a laboratory with suitable lighting condition. Each participant's session was run in a SONY laptop computer, model VGN-CS270T, with an Intel ® Core ^(TM)2 Duo processor, CPU P8600, 2.40 GHz, and 4.00 GB memory, using a mouse.

A computational application was developed to implement and test the experiment using the Unity 3DTM platform, with some scripts created in the C# programming language. This application allowed us to control the experiment, enabling and disabling the interface, monitoring the different times of the indications generated by voice in natural language and the time that really the user lasted in finding (to give click to) the object to be located.

The fourteen objects those were located in seven different scenes, together with other objects. Both objects and scenes were designed in BlenderTM and manipulated Unity 3DTM (see Figures 4.44 and 4.46).

Results were automatically recorded in a .csv (comma-separated values) file; afterwards statistical analysis was conducted using the SPSSTM (Statistical Product and Service Solutions) application.

Design and procedure. The test consisted on asking every user to locate fourteen different objects in seven different scenes (see Figure 4. 44).

Figure 4. 44 Scenes used in the study

Every trial was accompanied of a sentence that indicated the object to be located in the next scene (see Figure 4.45) and an image of the object to be located (see in Figure 4.46). The image and the sentence facilitated the location of the object within the scene. Each person lasted about 20 minutes to complete the fourteen trials.

Figure 4. 45 View of a sentence displayed to the users

Figure 4. 46 Set of objects used in the study

b) Statistical Analysis of Results

Comparison between the contextual saliency of each object and the average time of location. This statistical analysis allows us to evaluate the hypothesis: to greater salience of an object in context, less time it takes for the participants to locate it. In this hypothesis we have three independent variables:

IV_1 = No. of scenes (7)

IV_2 = No. of object (2)

IV_3 = No. of positions (2)

IV_4 = Contextual saliency

and one dependent variable:

DV_1 = Time

For obtaining the value of the contextual saliency we used the equation (19b), and for obtaining the average time of locating we calculated the time of locating of each object within of each scene, according with the two positions that had the object on the same scene.

c) Results

For this statistical analysis it was used Pearson's a correlation coefficients between (1) the contextual saliency of each object in its different positions of a scene and (2) the average time of locating of the participants. The values for calculating this correlation can be seen in Table 4.21.

Table 4. 21 Contextual saliency of each object and its average time to be located

Object	Position	Scene	Contextual Saliency	Average time
Big box	1	1 (Bath)	0.786	2.544
Big box	2	1 (Bath)	0.853	2.328
Hairband	1	1 (Bath)	0.932	3.229
Hairband	2	1 (Bath)	0.946	4.903
Bike	1	2 (Street)	0.865	1.446
Bike	2	2 (Street)	0.743	1.595
Mobile	1	2 (Street)	0.120	21.238
Mobile	2	2 (Street)	0.140	6.882
Microwave	1	3 (Kitchen)	0.640	1.586
Microwave	2	3 (Kitchen)	0.635	1.802
Salt shaker	1	3 (Kitchen)	0.490	3.804
Salt shaker	2	3 (Kitchen)	0.510	2.584
Keys	1	4 (Living room)	0.030	2.504
Keys	2	4 (Living room)	0.030	4.292
TV remote control	1	4 (Living room)	0.040	2.784
TV remote control	2	4 (Living room)	0.050	2.973
Big cube	1	5 (Classroom)	0.875	3.592
Big cube	2	5 (Classroom)	0.871	1.873
Big sphere	1	5 (Classroom)	0.821	3.551
Big sphere	2	5 (Classroom)	0.824	2.966
Pen	1	6 (Office)	0.470	2.397
Pen	2	6 (Office)	0.460	2.952
Pencil sharpener	1	6 (Office)	0.150	10.159
Pencil sharpener	2	6 (Office)	0.160	9.508
Beach cube	1	7 (Beach)	0.730	2.529
Beach cube	2	7 (Beach)	0.724	2.768
White towel	1	7 (Beach)	0.832	1.935
White towel	2	7 (Beach)	0.853	2.543

The result showed a correlation index of ($r_p = -0.460$) with a negative perfect trend, as shown in the scatter diagram of the Figure 4.47. The index indicates a total dependence between the two variables, which is known as inverse relationship. Here it is observed that when one

variable increases, the other decreases in constant proportion. With this, there is demonstrated that to lower saliency of the object, greater it is the time to locate it.

Figure 4. 47 Dispersion diagram showing the correlation between the value of contextual saliency and the average time of locating of the participants

4.6 Conclusion

This chapter proposes a computational model of perceptual saliency for 3D objects in virtual environments based on the color, size and shape; the visually salient features of the objects, the better perceived by the human visual system. The model was proposed from two approaches: individual and contextual. In the individual approach we followed three steps: (1) characterization of color, size and shape, that is, to apply different mechanisms for extracting and representing these three basic features of a 3D object; (2) apply a process, that allows to establish a normalized saliency for each feature; and (3) calculate individual saliency of each object in the virtual environment by giving a weight to the obtained saliency by feature.

In the contextual approach we also followed three steps: (1) calculating the average of the color, size and shape within the context; (2) measuring the distance of color, size and shape of

each object with respect to the average obtained above; and (3) applying a criterion of normalization to calculate the contextual saliency of each object in the virtual environment.

Some limitations of our computational model of saliency, which more than limitations we see them as future lines of research are: the inclusion of objects' texture, and the improvement of the variability of appearances that might have the shape of an object as: changing perspectives, different views and occlusions.

An advantage that presents this computational model of perceptual saliency is that all features considered such as the color, size and shape can help to determine the perceptual saliency of the objects to be located and of all the potential candidates to be selected as reference objects. Also, this model allows making an intelligent selection of the reference object to be used in the generation of directions for the location of objects. This computational model is very significant during the generation of indications for the localization of objects, as it will be seen in Chapter 6.

Both options were successfully implemented (individual and context). For the implementation and evaluation of this computational model of perceptual saliency we selected the Unity 3D platform; this platform is recognized for its potency in the area of game development. Unity 3D offers a work environment divided into two areas. The first area is a graphical environment where the author can create all the objects in the virtual environment and can easily manipulate its features. The second work area is a code editor, allowing the creation and modification of scripts, to implement the behaviours of objects. The scripts created for this model of saliency were developed with the programming language C#.

On the other hand, three experiments were conducted to validate the values provided by the computational model of saliency. The first experiment allowed to verify the measure of saliency for the shape of 3D objects. In the second experiment the color and size features of the object were included. Results demonstrated that the metric of the computational model of individual saliency fits, to an extent, with the perception of saliency of the participants. In the third experiment we verified that to greater saliency of an object in the context, minor is the time that takes them to the participants to locate it. The results of the evaluation of these three

experiments gave favourable enough results. Therefore, it was considered that the computational model of saliency proposed in this thesis turns out to be convenient and effective to measure the saliency in 3D objects. The algorithms were easy to implement, requires minimal space and computational time.

Chapter 5: User modeling and virtual world semantics

5 User modeling and virtual world semantics

One of the fundamental pillars for the generation of directions that are adapted to the user's needs and characteristics and to the current state of the virtual world is the capability to build explicit models of the user, of the virtual world and of the knowledge and experience that the user has of the different objects in the virtual world. These models will be represented in two data ontologies. The present chapter is divided into two parts:

In the first part is presented the semantic representation of the virtual environment that is necessary for the generation of directions, in which properties and relations between objects in the virtual environment are captured, through a data ontology.

In the second part is described in detail a proposal for user modeling, also through an ontology. This proposal allows representing cognitive and perceptual attributes (also called characteristics) of the users that interact with the virtual environment in which the process of object location is simulated. The selected cognitive attribute is the general capability to remember the location of objects (spatial memory), which will be used to infer the probability of remembering the location of a specific object. Perceptual attributes represent characteristics of the visual system of the user such as visual acuity, or color-blindness type. Likewise, information about the previous knowledge of the users of specific domains, and the knowledge they have of the objects in the virtual environment is also represented in the ontology.

To facilitate the semantic definition of virtual environments and users, a semantic editor-exporter was designed and implemented which is meant to be used in a pre-processing stage. The structure and functionalities of this software component is also briefly described in this chapter.

5.1 Semantics of virtual worlds

According with (Pittarello and De Faveri, 2006), an important part of the description of the semantics of a virtual environment is based on the concept of object. Two kinds of objects can be defined: the geometric objects (GO) and the semantic objects (SO). One or more geometric

objects can be associated to a semantic object, an entity that has a specific meaning shared by the author and by the final user of the environment. Semantic objects can be organized in a hierarchical structure, originating a multi-level description.

Under this theoretical approach, and using an ontology, a semantic description of virtual worlds is developed. The world ontology is undoubtedly of great relevance for the system. This ontology stores data of objects of the scene (also called virtual environment), in which the activity of locating objects is carried out.

The design of the “world ontology” was realized following the scheme proposed by (González, 2014), which includes three stages:

1. Identification of the characteristics of an object within a virtual environment.
2. Creation of the conceptual model of the world ontology.
3. Implementation of the world ontology with the Protégé tool.

The following describes each of these stages.

5.1.1 Identification of the characteristics of an object within a virtual environment

- a) *Common characteristics*: These attributes contain static information about an object, such as name, volume, absolutePosition_X, absolutePosition_Y, absolutePosition_Z, material, colorRGB_R, colorRGB_G, colorRGB_B
- b) *Saliency metrics*: these are the attributes that will represent the computed saliency of the object in different variables, including colorSaliency, sizeSaliency, shapeSaliency, individualSaliency.

5.1.2 Creation of the conceptual model of the world ontology

In agreement with the characteristics of every object, the conceptual model of the world ontology was developed. Figure 5.1 presents the different classes of this ontology.

Figure 5. 1 Hierarchy of classes of the world ontology

Figure 5.2 presents a conceptual map of the world ontology.

Figure 5. 2 Conceptual map of the world ontology

The central class of the world ontology is “**Object**”. This class represents every object of each scene.

The class “**KnowledgeArea**” refers to the knowledge area to which an object can belong. For instance, a mouse would belong to the area of Computing.

The objects inside a virtual environment will be categorized in two classes: StructuralObjects and ContentObjects. While a “**StructuralObject**” delimits the virtual space such as a wall, a floor, or windows; a “**ContentObject**” may be furniture, decoration of the

virtual environment or other objects, such as tables, chairs, computers or vases. All objects in the virtual environment are labelled as of content or structural type.

5.2 User modeling

The computational model for the generation of directions for the location of the objects that is proposed in this thesis, tries to be as adaptive as possible to the model of the user and to the environment. The user model that is presented in this section seeks to represent the user in a computational manner with an ontology called “user ontology”.

Using the same scheme proposed by (González, 2014), the “user ontology” was developed:

1. Identification of the relevant characteristics of the user.
2. Creation of the conceptual model of the user ontology.
3. Implementation of the user ontology with the Protégé tool.

The following describes these stages.

5.2.1 Identification of the characteristics of the user

Different types of characteristics were analysed and selected:

- a) *Basic characteristics*: essential attributes that identify each user. These attributes can also be called personal data, and contain static information about the user, such as: name, age, gender, email and previous training, among others. This set of characteristics is used for administrative purposes, except for the previous training data that will allow adapting the system to the previous knowledge of the user of specific domains such as Chemistry, Computing, etc.
- b) *Cognitive characteristics*: the general capability of remembering the spatial location of objects was included, as it will be useful to determine the type of directions that are more adequate for the user.

- c) *Perceptual characteristics*: including characteristics that allow the system to infer the way in which the user visually perceives a scene, such as visual acuity, and blindness of the color.
- d) *Knowledge and experience characteristics*: representing the knowledge the user has of the specific objects and object types in the environment, as well as the area or areas of knowledge of the user.

5.2.2 Creation of the conceptual model of the user ontology

Based on the different types of characteristics cited previously, the following conceptual model of the user ontology was created (see Figure 5.3).

Figure 5.3 Hierarchy of classes of the user ontology

Figure 5.4 presents a conceptual map of the user ontology.

La figura # presenta un mapa conceptual de la ontología mundo.

Figure 5. 4 Conceptual map of the user ontology

The principal class of the user ontology is "**User**" and it is related to three classes: UserCharacteristic, KnowledgeOfField, and Profile.

The class "**Profile**" contains all the personal information or basic characteristics of the user.

The class "**UserCharacteristic**" has two sub-classes that represent: (1) the cognitive characteristics, (2) the perceptual characteristics.

Within the subclass "**CognitiveCharacteristics**" there is a subclass called MemoryOfLocation. The subclass "**PerceptualCharacteristics**" has two subclasses called: ColorBlindness, and VisualAcuity. The following describes in detail each of these classes and their sub-classes.

- Cognitive characteristics

Cognitive processes include memory, attention, language, problem solving, and planning (Pessoa, 2008). Cognitive characteristics represent how individuals perform these internal processes within the brain, that is, how they store and manage information along the time. Only one cognitive characteristic is considered in this user model: memory of the spatial location of objects.

- (1) *Memory of location (Mol)*: This characteristic allows representing the ability that a user has for memorising the location of a series of objects in a scene and the relative position between them. Memory is widely studied by science; it represents the ability to store information along time in the brain (Aguado-Aguilar, 2001; Iachini et al., 2005). According to (Gutiérrez, 1991) visual memory is the skill to remember the visual characteristics and position that a set of objects had at a time, which were at sight but which are not in sight now or which have been changed of position.

There are two major categories of memory:

- a) *Short term memory (STM)* (also called primary memory). This type of memory contains information for a limited time, and tends to be used immediately. The information that is retained in this time is degraded or is replaced by new information in a matter of seconds or minutes.
- b) *Long term memory (LTM)*. This type of memory stores information derived from perceptual processes of people and stays there longer, even for a lifetime (Ballesteros, 1999; Gutiérrez Garralda and Fernández Ruíz, 2011).

To measure the user's memory of location, we propose to use the Rey–Osterrieth complex figure test (ROCF) which is a widely used neuropsychological test for visual perception and long term visual memory. The ROCF was proposed by Rey in 1941, and standardized by

(Osterrieth, 1944). This is a neuropsychological evaluation, which measures ability in different functions such as: visuospatial abilities, memory, attention, planning, and working memory. The test consists of a series of steps that are next described:

- a) Examinees receive the ROCF stimulus card (see Figure 5.5), and are then asked to copy the figure.

Figure 5.5 Rey-Osterrieth's complex figure

- b) Then, they are instructed to draw what they remember.
- c) After 30 minutes, they are required to draw the same figure once again.

The anticipated results vary according to the used scoring system, but commonly it includes scores related to location, accuracy and organization (Shin et al., 2006).

For the numerical evaluation of the results obtained in the ROCF test 18 constitutive elements of the drawing are considered, assigning a value to each of them according to the criteria in Table 5.1 (Osterrieth, 1944).

Table 5. 1 Criteria for evaluating the ROCF

Criteria	Score
Correct and well placed	2 points
Correct and wrong placed	1 point
Misplaced, but recognizable	½ point
Deformed or incomplete element	0 point
The total maximum of points	36

The scale proposed for evaluating Memory of Location, from the results of the ROCF test is the follows:

- '0.00' equal to 'Low', for punctuation lower than 12 points.
- '0.50' equal to 'Medium', for punctuation bigger than 13 and lower than or equal to 24 points.
- '1.00' equal to 'High', for punctuation bigger than 25 up to 36 points.

This characteristic will be very important in our model to measure the probability of the user to remember the location of an object, which is described in Chapter 6 (see section 6.1 Algorithm for the selection of the best reference object (SBRO)).

- **Perceptual characteristics**

Perceptual characteristics allude to how the user interprets the information that comes to him/her through the visual route. Sight, also known as visual perception, gives users the ability of identifying and locating objects in their environment.

In visual processing there may exist difficulties or disorders that affect the way in which the users interpret the information that they perceive. The most habitual difficulties of the vision are eye refractive errors such as myopia, hypermetropia or astigmatism, and strabismus.

Other causes that can provoke bad vision are the congenital anomalies such as cataracts or blindness, among others. Visual problems can be caused also by different affections like presbyopia, glaucoma, diabetic retinopathy, night blindness, stroke, or brain tumour.

In our user model, the clarity or sharpness with which a user can perceive or identify objects in space, that is, visual acuity, is modeled, as well as a congenital anomaly that causes difficulty in distinguishing colors, known as color blindness. Next a more detailed description of these two disorders of vision is presented.

- (1) *Visual acuity (Va)*: Westheimer (1975) defined this concept like the performance limit in distinguishing spatial details of a visual object in ideal conditions of luminosity. For Goto (2002) visual acuity is the size of the smallest retinal image that can be seen by the user and is expressed as the minimum visual angle (minutes of arc) projected onto the sensory retina. Likewise, Howarth et al. (1997) cited by Herrero et al. (2005), specified that the visual acuity is a measure of the eye's ability to resolve fine detail and is dependent upon the person itself, the accommodative state of the eye, the illumination level and the contrast between target and background.

The goal of modeling this characteristic is to allow the system to know the distance at which a user sees an object blurry. For measuring the visual acuity of users the Snellen test is commonly used. The Snellen test was introduced on the 19th century. This test consists in identifying correctly the letters of a graph known as Snellen's graph or Snellen's table (see Figure 5.6).

Figure 5. 6 Snellen's table

The letters have a diminishing size depending on the level or line in which they are. When the test is applied to a user, and for example it is observed that the user sees correctly line four (20/50 in the drawing), but the line five (20/40) is seen blurry, then visual acuity will be assigned the value resulting from the division 20/50, which is equivalent to the reciprocal of the minimum visual angle, a visual acuity of 1.0 being 1 minute of arc, and 0.1 being 10 minutes of arc (Goto et al., 2002).

The goal of modeling this characteristic is to obtain the maximum distance (Md) to which the user can see with clearness an object, that is, his clarity of the vision, which is described in Chapter 6 (see section 6.1 Algorithm for the selection of the best reference object (SBRO)).

- (2) *Color-Blindness (CB)*: Color blindness is a visual deficiency that affects the perception of color, causing a series of changes in the seen colors. This deficiency presents in addition the difficulty to distinguish soft tones, with a total limitation to detect the color (Kim et al., 2007).

In color blindness different types and degrees exist, among the most common are: acromacy, monochromacy, dichromacy in three types: protanopia, deuteranopia, and tritanopia, trichomacy in two types: protanomaly and deuteranomaly, which were described with more detail in Chapter 2 (see section 2.4.2 - LMS Color Model).

In our user model we decided to model the three types of dichromacy because they are the most common ones: protanopia, a problem to recognize the red color; deuteranopia, a problem to recognize the green color; and tritanopia, a problem to recognize the colors blue and yellow.

The general transformation for people who look good RGB – CIE-XYZ - CIELab is amplified with an intermediate step to use LMS. Figure 5.7 shows the transformation sequence that will be followed for the conversion of the color of a 3D object in the color space RGB into their dichromatic versions. The proposed procedure includes three steps: (1) the 3D object color is obtained in the RGB system; (2) then RGB color values are converted to the LMS

color system; (3) after LMS color values are converted to the CIE-XYZ color system; and (4) the values of CIE-XYZ are used to convert the color to the CIE-Lab system.

Figure 5. 7 Sequence of transformations for the conversion of the color of a 3D object in the "normal" RGB color space into its dichromatic versions

The transformation from RGB to LMS color is obtained using the matrix T_1 , defined as follows:

$$T_1 = \begin{bmatrix} 17.8824 & 43.5161 & 4.1193 \\ 3.4557 & 27.1554 & 3.8671 \\ -0.02996 & 0.18431 & 1.4670 \end{bmatrix}$$

With the matrix T_1 , it is possible the transformation from RGB to LMS:

$$[L \ M \ S]^t = T_1[RGB]^t,$$

People with protanope visual impairment perceive colors differently due to the lack of red cone photopigment. The different color perception of protanopes can be modeled as a linear transformation, mapping normal cone responses LMS to protanope cone responses $L_p M_p S_p$. This can be represented with a linear mapping matrix on T_2 , which is defined as:

(protanopia)

$$T_2 = \begin{bmatrix} 0 & 2.02344 & -2.52581 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Using T_2 , $[L_p M_p S_p]$ is obtained as:

$$[L_p M_p S_p]^t = T_2[LMS]^t,$$

For cases of deuteranopia and tritanopia, these can also be modeled with a linear transformation. This can be represented with linear mapping matrices T_3 and T_4 .

(deuteranopia)

$$T_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0.494207 & 0 & 1.24847 \\ 0 & 0 & 1 \end{bmatrix}$$

Using T_3 , $L_d M_d S_d$ is obtained as:

$$[L_d M_d S_d]^t = T_3 [LMS]^t$$

(tritanopia)

$$T_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.395913 & 0.801109 & 0 \end{bmatrix}$$

Using T_4 , $L_t M_t S_t$ is obtained as:

$$[L_t M_t S_t]^t = T_4 [LMS]^t$$

This characteristic of the user is saved in data ontology with the type of color blindness.

- **Knowledge and experience characteristics**

With knowledge characteristics we want to reflect a user's familiarity with some knowledge domains or areas of knowledge, which will be used to infer the user's probability to recognize the objects associated to these domains within the virtual environment. Example areas of knowledge are Computer science, Chemistry, Geometry or Nature.

The class "**KnowledgeOfField**" in the user's ontology stores the degree of familiarity the user has with a knowledge domain relevant to the virtual environment. This degree of knowledge can be high, medium or low. This value should be established by the user, and is saved with a string value.

In this same class there is a link, to class “*KnowledgeAreas*” of the world ontology, to which every object belongs. With this link connects the user's ontology to the world ontology of the virtual environment.

With experience characteristics we want to represent the history of interaction of the user with specific objects in the virtual environment. These include the time of central vision, time of peripheral vision, and seen surface, and they will be computed according to the perceptual characteristics of the user such as visual acuity. The procedure for the computation of this set of characteristics will be described in detail in Chapter 6 (see section 6.2.1 - Criteria chosen for the selection of the best object of reference).

5.3 Implementation of the ontologies with the Protégé tool

To implement the world and the user ontologies, the tool Protégé, version 5.0.0-beta-17, was used. This is a free and open code tool which is used as ontology editor and framework for building intelligent systems based on models of ontological knowledge, and it has become the most widely used ontology editor in the world for research (Knublauch et al., 2004)

In order to allow the introduction and computation of relevant information about the objects in the scenes and about the users of the system, and its storage in the corresponding data ontologies, we developed a semantic editor-exporter in the EditorWindow of Unity 3D™. The editor-exporter consists of several scripts programmed in C#. We used the Jena framework to handle the ontologies. Jena is a Java application programming interface that is available as open-source (McBride, 2002).

The principal functions of the editor-exporter are:

- To create or modify a specific ontology.
- To create and delete classes representing the object types in the scene (Content Objects or Structural Objects).
- To add new properties to the objects.
- To create or delete already existing knowledge areas.
- To allocate areas of knowledge to the objects.

- To allow inputting values for some properties of the objects, such as name, description.
- To automatically determine physical parameters for each object: color, material, height, width, volume, orientation, and position.
- To calculate the individual saliency metrics for each object

The editor-exporter bases all its calculations on the model of perceptual saliency, which was described in Chapter 4.

A snapshot of the developed editor-exporter is shown in Figure 5.8 with an example of calculated data. This software represents a powerful tool to extract knowledge from 3D objects.

Figure 5. 8 Image of the Editor - Exporter in Unity

5.4 Conclusion

At present many of the computer systems that offer help to users are characterized by personalizing the system's behaviour to the characteristics of the users. Likewise, we selected the user characteristics that are considered more relevant to help the user during the process of object location, which were classified as cognitive and perceptual characteristics.

On the other hand, the principal aim of the development of the world ontology was to make a semantic modeling of the virtual environments that allows representing all the relevant information related to the objects in the scenes.

Both the world model and the user model are highly reusable and easily extensible.

Chapter 6: Automatic generation of directions to locate objects in VEs

6 Automatic generation of directions to locate objects in VEs

The main goal of this thesis work is to design a software component that in an intelligent way generates directions in natural language for a user to be able to locate objects in a virtual environment.

For the design of this software component three situations were analysed:

- (a) The object to be located (OL) is in the same room that the user, but not in the user's field of view.
- (b) The OL is in another room, and therefore not at the sight of the viewer.
- (c) The OL is in the user's field of view.

A series of sub-cases that stem from the third case were then identified, see Figure 6.1.

This chapter presents the proposed approach to deal with the three analysed cases, and the algorithm for the generation of natural language directions to be implemented in the software component.

It is necessary to indicate that only the third case was considered in detail, as the first and second cases can be easily transformed into the third one.

Figure 6. 1 Cases considered for automatic directions generation

(a) The OL is in the same room that the user, but not in the user's field of view

In this case, it was considered that the object to be located is in the same room that the user, but out of his field of vision (see Figure 6.2).

Figure 6. 2 (a) The OL is in the same room that the user, but not in the user's field of view

For treating this case, it was proposed as a strategy to give an extra direction to the user, with the goal that the object enters in the field of vision of the user. For example: "Turn around to your right ", "Turn around a little to the left", "Look at the door", etc. With this type of directions, the user is placed in the case (c).

(b) The OL is in another room, and therefore not at the sight of the viewer

In this case the OL is in another room (see Figure 6.3).

Figure 6. 3 The OL is in another room

Under this circumstance it is also proposed to give the user an additional direction that allows him to go towards the room where the object is located.

If the user has a high degree of familiarity with the environment he can be assumed to have already created a representation or mental map of the site, therefore it can be simple for him to move to the target room. The type of direction to be given to the user might be: "Go to your brother's room", "Go to the bacteriology laboratory", "Go to the infirmary". These directions guide the user towards the room or place where the object is.

However, if the environment is unknown to the user, it would be necessary to give more elaborate directions with more specific references to the trajectory to be followed, such as "get out of this room, turn right, follow the corridor, and enter through the fourth door on your left". The generation of this type of directions is out of the scope of this thesis.

(c) The object to be located (OL) is in the user's field of view

In this case it is considered that the object to be located is in the user's field of view, that is, within his field of vision (also called visual field) (see Figure 6.4). The concept of field of vision is described in detail later in the paragraph of "*Time of vision*" of the section 6.1.

Figure 6. 4 The OL is in the user's field of view

This case was subdivided into the following four cases:

- (c.0) The object to be located (OL) is inside another object or occluded, so it cannot be seen by the user.
- (c.1) The OL is visible and salient in itself.
- (c.2) The RO has a high probability of being remembered by the user, because he has a high previous knowledge of the environment.
- (c.3) The RO does not have a high probability of being remembered by the user, because he has a low prior knowledge of the environment.

All these cases are described in detail in the section 6.3.

The designed solution relies on three auxiliary algorithms:

- An algorithm for the selection of the best reference object (SBRO), which is based on the previous knowledge of the user about the object types in the scene, the probability of the user remembering the location of specific objects, and their perceptual saliency.
- An algorithm for the selection of a relevant spatial relation (SSR) between the object to be located (OL) and the reference object (RO).
- An algorithm for the generation of directions (GD) that considers the cases or situations that are treated.

The algorithm SBRO comprises a series of mathematical representations considered for the implementation of some perceptive and cognitive characteristics of the user, which are related with the knowledge about objects of virtual environment. These characteristics were described in the Chapter 5 (see section 5.1.2 Creation of the conceptual model of the user ontology). These characteristics were very important to determine the used criteria in the implementation of the cases of situation analysed.

Given that to locate an object, is specified with another object whose location is usually assumed by the speaker to be known by the hearer. This supports a relativistic notion of space is the concept of proximity. The notion of proximity is an important concept for the human spatial knowledge. To refer to the space a spatial relation can help. The spatial relation is described as a spatial preposition (Costello and Keller, 2006).

For selecting the set of spatial relations used, for the generation of sentences within of the component of software, in this chapter are presented the results of a study realized with several participants, across a set of virtual scenes. This study allowed identifying the spatial relations most used by the participants, in the development from directions in natural language, for the location of objects.

Based in the obtained information in this study, our algorithm of SSR included two types of spatial relations: the projective relations and the topological. In addition to the spatial

relation "between" as a special case. The algorithm SSR includes the computational criteria for the selection of the 12 spatial relations used in the generation of directions: on / under, close, left, right, above, below, in front of, behind, inside, between, first, second, or third, which are described in the section 6.2.2 of this chapter.

The algorithm GD generates expressions in the form of descriptive sentences for locating of objects. Each sentence includes information necessary for a correct comprehension as: object to be located, spatial relation, reference object, and properties of object, according to each case. The structure of a sentence generated by our system is based on the linguistic structure that Kelleher presented (2003), see Figure 6.5:

Figure 6. 5 Scheme of a final sentence

where

OL is the object to be located

RO is the reference object

The representation of the syntactic structure of the sentence appears in the Figure 6.6.

Figure 6. 6 Syntactic structure of a sentence

- The nominal syntagm refers to the object to be located, its structure is:

“The + feature of object (optional) + Name of the object to be located”

Example: The yellow pencil, the red pen, the blue ball.

- The verbal syntagm considers the nature of the problem, that is, the need to locate an object. This verbal syntagm will always be the character string "is", that to indicate the position situation.
- The adverbial syntagm is the one that contains both the spatial relation and the reference object, its concrete structure is:

"Spatial relation + "position of the reference object (in case there are several) + feature of the object (optional) + Name of the reference object + and + Name of the second reference object (if we have a case of spatial relation "between")

Example: In front of the desk, between the red ball and the bicycle, to the right of the second printer.

The implementation of this algorithm was based in the different cases of situation, aforementioned, which will be discussed in the section 6.3 of this chapter. There are the sentences together with their syntaxes allow seeing, how each situation case is treated.

The sentence generated by the system can be displayed on the screen and heard. The voice as a written sentence is expressed in English.

Next are presented the situations considered for the generation of natural language directions for the localization of objects.

On the other hand, for locating an object that is in the view of a user, several aspects of this one can be evaluated. The aspects selected for the evaluation of the object inside this work were: perceptual saliency, prior knowledge, and probability of remember.

These aspects are orientated especially towards the selection of a reference object, as support in the location of an object that is in the user's view. The reference object is used in the

expression generated by the component of software, according to the case. There is the possibility of not using a reference object, by the fact that the object to be located already is salient by itself, and so it could be easily recognized by the user.

The combination of these aspects is proposed as a direct alternative for selecting the best reference object. These and other aspects derived from them are described in the next section, along with the criteria and the logic of the algorithm for the selection of the best reference object.

6.1 Algorithm for the selection of the best reference object (SBRO)

Given that reference objects are a key resource for locating objects in a physical or virtual environment, the selection of the best reference object is an important factor in the generation of directions. Different criteria can be used to choose a reference object. Gapp (1996b) described nine of them, which are:

- (1) *Distance*: The distance treats itself about the nearness or distance that exists between objects. In this case the distance between the reference and the located object respective to the point of view.
- (2) *Mobility*: Is the property of an object that represents the displacement which occurs within the physical environment. Nevertheless, Talmy (1983) mentioned that in most cases, the preferred reference objects are objects with a stationary setting within a certain reference frame.
- (3) *Previously mentioned objects*: This is a factor that considers the reference objects that have helped in previous locations. In a situation like this, might be better to use reference objects "previously mentioned" than unknown new objects.
- (4) *Functional dependencies*: This property is related to objects that are related, and can be more likely to be remembered together. Also, it is important the semantic relation between them and their spatial nearness.

- (5) *Prior knowledge*: This characteristic helps to the selection of an object of reference, only if the object is known by the person that asks for its location.
- (6) *Visual saliency*: The visual salience is a property of an object, its perception is in charge of people. Likewise, the visual salience occurs in the interaction of the basic features of an object to be located, in correlated to the corresponding attributes of the surrounding objects.
- (7) *Intervening objects*: This criterion refers to how complicated is the identification of an object when the object is not in its intrinsic position or its parts are occluded by other objects.
- (8) *Referentiality*: Referentiality is an important aspect in the process of object location. For be referred an object is necessary to have a semantic interpretation of him, for identifying it of unequivocally way, from a perspective.
- (9) *Frame of reference*: In the process of object location it is necessary to use a set of criteria by an observer or listener to identify the position and other physical magnitudes of the environment where the localization process is performed.

This order is not strict, it is only a reference that facilitates the choice of the criteria for selecting the best reference object. All criteria can be determined and scaled quantitatively by its relevance, and under a certain context (Gapp, 1995, 1996b).

Based in some criteria, the following procedure and the lineal function were proposed by Gapp (1995) for deciding which object should be selected as the optimal reference object:

Each object receives a graded evaluation in the interval $\tau = [0,1] \subseteq \mathbb{R}$, for each ordering feature, e.g., 0.0 for a moving object, 0.5 for a movable (but not moving) object, and 1.0 for a permanently located object. Distance can be scaled by intervals and the complete

range from 0 to 1 could be used for visual salience. The same is performed for each context dependent factor. If a reference object candidate was previously mentioned, the graduated factor “previously mentioned” for this object depends on the number of objects, located objects as well as reference objects, mentioned since its last occurrence. Functional dependencies are rated in relation to their intensity and prior knowledge is evaluated by the probability that a particular object is known to a listener.

Assume that n factors are used for the evaluation of an object’s quality as a reference object. This results in a n -digit feature vector $f = (f_1 \dots f_n)$ in the interval space $\tau^n = [0,1]^n$. Depending on a particular situation, context factors can change the importance of one or more dimensions, e.g., if the user has only a mental image of the scene and no direct visual access. The influence of these situational aspects is accounted for by the possibility of weighting the affected dimensions using a context dependent n -dimensional scaling function SC_{ext} :

$$SC_{ext}: \tau^n \rightarrow R^n; (f_1, \dots, f_n) \rightarrow (SC_1(f_1), \dots, SC_n(f_n))$$

with

$$SC_i(f_i): \tau \rightarrow R, 1 \leq i \leq n$$

The optimal reference object can now be defined as the $n-x$ dimensional subspace $S^{n-x} \subseteq \tau^n$, $1 \leq x \leq n$, in which x is the sum of factors defined by an interval. The means, that if an optional reference object is defined by only one single value for each factor f_i , then hence $x = n$ and $S^{n-x} = S^0$ denotes a point in τ^n .

We quote textually this procedure, for the reason of having been one of the referents most nearby to our solution.

To design the algorithm SBRO, three criteria were considered. From the list proposed by Gapp (1995, 1996b) we selected: (5) prior knowledge and (6) visual saliency. A third criterion used in our algorithm is the probability of remembering the position of the object within the environment. These three criteria were chosen with the aim of considering both perceptual and

cognitive characteristics of users as well as physical aspects of the objects in the virtual environment.

The following describes each of these criteria and how they are operationalized to obtain a normalized value between 0 and 1 for each of these criteria.

- ***Perceptual saliency (also called General Saliency (GS))***

This aspect allows measuring the most prominent feature or features of an object. In accordance with the computational model of perceptual saliency that was proposed in Chapter 4, this criterion can be measured from an individual point of view and relative to a context. In the procedure three specific characteristics of objects are taken into account, which are: color, size and shape. A measure of perceptual saliency or general saliency (GS) will be obtained for all the visible objects within a virtual environment from the current user's location as a normalized value between 0 and 1, or as a standard score.

Logically, objects with a higher perceptual saliency will be better candidates to be selected as reference objects (RO).

- ***Probability of remembering (PR)***

Considering that the probability of remembering something is always subjective and imprecise, we attempted to estimate the ability of a user to remember the position of an object previously seen in a virtual environment. To measure it the following equation (21) is proposed:

$$PR = \frac{GS + Mol + Hv}{3} \quad (21)$$

where:

- *GS* represents the value of the perceptual saliency of the object (see equation (20b) within Chapter 4). We consider that salient objects are more probable to be remembered than low salient objects.

- *Mol* is the value of the cognitive characteristic that represents the general user's ability to remember object locations, which is modelled in the user ontology. The value of *Mol* can be obtained from the application of the proposed scale (the Rey–Osterrieth complex figure test, (see section 5.1.2 Creation of the conceptual model of the user ontology)). The higher the general ability of the user to remember object locations, the higher the probability that he/she remembers the location of a specific object.
- *Hv* represents the history of vision of a user with regard to a specific object seen inside a scene. The next section explains the process proposed for the calculation of this feature.

- ***History of Vision (Hv)***

The proposed process for obtaining the history of vision *Hv* involves several concepts, which are represented in the following equation (22):

$$Hv = [vso * Cv * Tv] - To \quad (22)$$

where:

- *vso* represents the maximum viewed surface of the object inside the scene from any of the past user's points of view. For obtaining the seen surface of an object we propose the use of the same rendering technique to project 3D graphics onto a 2D image that was applied to obtain the objects that are visible to the user (see Chapter 4, section 4.3 Modelling the Contextual Saliency). The bigger the object's surface that has been visible to the user, the higher the value for its History of Vision, and therefore the higher the probability of remembering it.
- *Cv* represents the clarity of vision or clarity of perception. This property allows determining how clearly a user has been able to perceive an object in a scene. The clearer the perception achieved of an object, the higher the value for its History of

Vision, and therefore the higher the probability of remembering it. Next section will detail the procedure proposed for the computation of this feature.

- T_v is the time of vision of an object by a user. The longer an object has been observed, the higher the value for its History of Vision, and therefore the higher the probability of remembering it. Below we will detail the procedure proposed for the computation of this feature.
- T_o is the time of oblivion, that is, the time in which is involved the involuntary action of stopping remembering or keeping in memory the information of the objects observed in a scene.

The following conceptually describes the three last variables. Likewise, a procedure to obtain a value between 0 and 1 for each of the variables is presented.

- ***Clarity of Vision (C_v)***

From a sensorial point of view, C_v is the ability to distinguish what object is being perceived, which depends on the object's image projected on the retina. The process of human visual perception is continuous and the size of an object's image on the retina will continuously depend on the distance between the eye and the object to be perceived.

The process to estimate the clarity of the vision of an object for a user turns out to be rather complex (Herrero and De Antonio, 2005). Here we will constrain our model to consider the effect of visual acuity. The greater the user's visual acuity the better the clarity of vision at a given distance.

Visual acuity is equivalent to the minimal vision size at 6 meters, that is to say, the distance at which the letters poster used for the Snellen test is placed. To clarify this definition, figure 6.7 presents a user that observes an object at the aforementioned distance.

Figure 6. 7 Clarity of vision

The object has a height H , while the user is at a D distance of the object. In this case it is necessary to find out the angle (α) indicated in the Figure 6.7, for which the trigonometric function *sine* is used (see Equation (23)):

$$\sin \alpha = \frac{H}{C} \quad (23)$$

The values of H and D are known, but not the value of C . To obtain the value of C the Pythagorean Theorem is applied, using the Equation (24).

$$C^2 = H^2 + D^2 \approx C = \sqrt[2]{H^2 + D^2} \quad (24)$$

The value of C is substituted in the Equation (23), resulting in:

$$\sin \alpha = \frac{H}{\sqrt[2]{H^2 + D^2}} \quad (25)$$

To obtain the angle, the *arcsine* function is applied using the Equation (26):

$$\alpha = \arcsin \frac{H}{\sqrt[2]{H^2 + D^2}} \quad (26)$$

Given a specific object, the value of H should be replaced by the value V_a that represents the visual acuity of the user and D by the distance between the user and the object (see Equation (27)).

$$\alpha = \arcsin \frac{Va}{\sqrt[2]{Va^2 + D^2}} \quad (27)$$

Va is expressed as a fraction where the higher number refers to the distance to which a person stands in front of the table, which is typically 20 feet (6 m), and the lower number indicates the distance that a person with normal eyesight could read the same line as another person reads correctly (see section 5.1.2 Creation of the conceptual model of the user ontology).

After obtaining the angle α , the following step is to obtain the maximum distance at which the user can see with clearness an object. To calculate this value, a reverse process to the previous one is used, in the Equation (28):

$$Md = \frac{H}{\sin \alpha} \quad (28)$$

Where

H is the real height of the object that the user perceives.

After obtaining the value of Md, if the object the user is observing is closer than Md, this means that he sees clearly the observed object, and therefore the value '1' is assigned to the clarity of the vision. Otherwise, a greater distance between the object and the user will decrease the clarity of vision. It is necessary to compute the maximum distance at which the user can appreciate some detail of the object. For this, it is necessary to calculate the so called infinite distance of vision D^∞ , using a minimal angle of 5 minutes. Equation (29) reflects this operation.

$$D^\infty = \frac{H}{\sin 0,01^\circ} \quad (29)$$

Once obtained the infinite distance of vision D^∞ , this one divides with the distance to the object for finding the normalized value of Cv.

- *Time of vision (Tv)*

Tv represents the time that an object has been seen by a user. To obtain the value of Tv we will analyse the time of central and of peripheral vision.

The field of vision is a key concept that takes special relevancy here. This concept was originally described by (Sanders, 1970) who used the term “functional visual field” to define the visual field area about which information can be acquired in a brief glance, without movements of the eyes or the head. (Harrington and Drake, 1993) also defined field of view (FoV) as the segment of space in which the objects can be perceived simultaneously to look at a fixed and immobile object. This concept is a determining factor in the visual quality of the user, during an instant of time.

The FoV is measured in degrees. For this, the horizontal and vertical axes are taken into account (see Figure 6.8). In the case of the human beings it spans 180° horizontal (semicircle) and 130° vertically (Lienhard, 2010). This field of vision is divided into two parts: the central vision and the peripheral vision (the peripheral vision is also called partial or lateral view).

Figure 6. 8 Field of vision of a user

The central vision is supported on the fovea, that is, on the centre of the retina, which only contains cones joined with intermediate cells to the fibres of the optic nerve. The fovea is characterized by being very discriminant. The luminous rays are focused on the fovea, provoking that its threshold of luminous sensibility is very high. The visual acuity that the fovea has is the maximum. The fovea works especially in the light of day (photopic vision).

The peripheral vision is supported on the peripheral retina, which is richer in rods than in cones, and its threshold of luminosity is lower than that of the fovea. This type of vision adapts more to the night vision (scotopic vision), but the visual acuity in exchange is lower (Doron and Parot, 2004). Figure 6.9 shows a perspective of these two types of vision.

Figure 6. 9 Central and peripheral vision

These two types of vision are important for object recognition. When a person observes an object directly, he can recognize all its details, whereas if he sees it peripherally he cannot see it completely. This is due to the distribution of the receptor elements of visual stimuli in the human eye.

In this context, to calculate the value of T_v , the following criteria are taken into account:

- If the time of central vision (t_{cv}) of an object by a user exceeds 20 seconds, we will consider that it has been stored in long term memory (Bataller and Moral, 2006; Craik and Lockhart, 1980) and T_v will take a value of 1.

- Otherwise the time of peripheral vision (tpv) is also considered. For it, we proposed the following equation which allows to calculate the combined time of vision:

$$T_v = t_{cv} + \frac{tpv}{5} \quad (30)$$

Where

t_{cv} is the time of central vision, and

tpv is the time of peripheral vision

tpv is divided by 5 because it is considered that the peripheral vision has less weight at the moment of being able to remember an object, since it does not allow to perceive correctly all its characteristics.

Finally, if T_v is bigger than or equal to 30 seconds the value of 1 is also assigned to it because we consider that the user had enough time to store the object's position in long term memory. Otherwise, it is divided by 30 and a value between 0 and 1 is obtained.

- Time of oblivion (T_o):

This time is associated with retention or time of permanency of the information in short-term memory (STM). Peterson and Paterson (1959) determined that the time of storage of the STM is of 15 and 30 seconds. Likewise, the oblivion can be explained depending on the decadence or the simple passage of time. The oblivion is the result of the competition between the new and the old associations.

To obtain the value of T_o the time of central vision t_{cv} is taken into account. If an object has been observed in detail for a long time it is more difficult to forget. For this calculation is proposed the Equation (31).

$$T_o = (T_v * 60 * 60 * 24) \quad (31)$$

The result of this equation is given in seconds like the aforementioned times. The principal reason of multiplying T_v by the number of seconds of one day is because it was considered that for every second that the user has seen an object, it might take one day to forget this object.

- ***Prior knowledge (PK)***

This aspect is evaluated considering the areas of knowledge to which every object belongs, and the general familiarity that the user has with an area of knowledge, mainly associated to the user's previous training or experience. A user might have knowledge in several of these areas and possess a different level of knowledge of each of them. For example, a computer engineer should have a high knowledge of input/output devices. On the other hand, it is probable that an architect has a high knowledge of graduated rules and other objects related to building design, and a medium knowledge of the field of computer science, and consequently of input/output devices.

Previous knowledge represents the level of familiarity that a user possesses with a certain area of knowledge and the objects associated with this area. For the valuation of this criterion the following scale was established: if the user has a high knowledge of the area a value '1' is assigned, if it is medium '0,5' and if it is low '0'. This aspect is valued by the own user and stored in the user ontology, which was described in Chapter 5 in detail (see section 5.1.2 Creation of the conceptual model of the user ontology).

A threshold of acceptable previous knowledge was established at 0,5. This would imply that the user can at least recognize the object and knows how it is called. An object that belongs to a knowledge area for which the user's previous knowledge is lower than 0,5 should not be used as reference object because it would not be meaningful for the user.

A concept associated with the probability of remembering the position of an object and the prior knowledge is the knowledge of the environment (K_e). When a user is very familiar with the environment and its content there will be a very high probability of remembering the positions of the objects. Therefore, it would not have too much sense to select a reference object based on its saliency.

For calculating the global user's knowledge of the environment the following equation (32) is proposed:

$$Ke = \frac{\sum Pr}{No} \quad (32)$$

In order to select a reference object, it is important to be sure that it is close enough to the object to be located. We have established a distance threshold that may not exceed 10 % of the size of the room, that is, only the objects that are located inside a sphere with its centre in the object to be located and a radius of 10% the size of the environment, will be considered. The threshold was established relative to the size of the environment considering that the concept of two objects being close is a quite fuzzy one, and it is relative to the dimension of the surrounding space.

The results of these operations are stored in the model. For it, all data is treated in an independent way and are executed before generating a direction. Likewise, there are obtained the instances of knowledge to be edited (formed by the user and the object), and the property related to each data. Subsequently, it is checked whether the property has any value, if not it is created with a 0 or an empty string. The existing value is overwritten and stored in the model. Each time the model is modified the set of data within the system is updated.

- *Selection of the Best Reference Object (BRO)*

In the first place, a list of potential reference objects is computed by applying the distance criterion, that is, those objects whose distance to the object to be located is lower than the 10% threshold.

Then three values have to be calculated for each of these objects (prior user's knowledge, perceptual saliency, and probability of remembering), and they are added to obtain a final value that combines these three aspects. The object with the maximum value will be considered as the best reference object. This is represented by the following equation (33):

$$Bro = Max\ of\ (GS + PR + PK) \quad (33)$$

The obtained results should be then analysed for determining the criteria of the cases (c.0), (c.1), (c.2) and (c.3) cited in this Chapter. These cases are described in detail in the section 6.3.

Once a reference object has been selected, it is necessary to determine the spatial relations that hold between the OL and the RO.

6.2 Algorithm for the Selection of Spatial Relations (SSR)

In the generation of directions in this computational model the concepts of space and spatial relation are crucial.

According to Piaget's theory about spatial knowledge, three types of spatial relations are established: topological, projective and Euclidian, which are described below:

- The topological relations consider the space inside an object or figure; likewise they comprise the relations of proximity, separation, order, and continuity.
- The projective and Euclidean relations consider the objects and their representations, taking into account the relations between those objects according to projective systems (projective space), or according to coordinate axes (or Euclidean space).

Based in this theory (Gapp, 1994, 1995a) also divided the spatial relations into two classes: topological (“at”, “near”, etc.) and projective, or directional (“in front of”, “to the right of”, etc.). He also considered the relation “between”, which takes an exceptional position in the group of spatial relations.

Topological relations

Two topological relations are: “at” and “near”, both refer to a region proximal to an object. Their range of applicability therefore often overlaps. Differences appear mainly on the pragmatic level. “At” localizes an object in the proximal exterior of a RO (reference object).

Contact is not necessary. For the applicability of the relation “near”, the contact between objects is explicitly prohibited (Gapp, 1994).

Projective relations

Some projective relations are: “in front of”, “behind”, “right”, “above”, “below”, and as a special case the relation “beside”, which can be viewed as the disjunction of the relations “right” and “left”.

The relation “beside” is the only projective relation with an orientation that involves two opposite directions; these correspond to “right” and “left” relations. When the relation “beside” is used, it always implies an alignment with the RO (Gapp, 1994).

The relation “between”

The relation “between” occupies an exceptional position among the spatial relations aforementioned, because it refers to two objects. The basic meaning of “between” is defined by the structure of its region of applicability. The location with the highest degree of applicability is exactly midway between the two ROs. Increasing the distance to this location decreases the applicability, depending on the extension of the ROs. This dependence on the RO is expressed by a displacement of the region of applicability towards the ROs with the smaller extension (Gapp, 1994).

Starting with these theoretical foundations on spatial relations, it was decided to realize an exploratory study with the goal to obtain evidences of what directions users might give to locate an object. This is described in the next section. Furthermore, to check whether all the relations described above would be used by the users for this task, and to select the important relations to be implemented in the component generating directions in this model.

On the other hand, in section 6.2.2 we describe the computational criteria used for the implementation of spatial relationships recovered from our exploratory study.

6.2.1 Exploratory study

The following exploratory study was realized with the same undergraduate students and professors of the Escuela Técnica Superior de Ingenieros Informáticos of the Universidad Politécnica de Madrid that participated in the experiment of contextual saliency of the Chapter 4 (see section 4.5.3 Experimental evaluation of the computational model of contextual saliency). The study consisted in showing fourteen different objects (see Figure 4.44) in seven different scenes.

Every scene was presented twice changing the objects to be located and their positions. Once the user located each object, we asked him/her to write down the way in which he/she would explain another person where the object was located.

The generated sentences by the 28 users with respect to the objects to be located within each scene were analysed regarding the use of spatial relations, obtaining the diagram in Figure 6.10, which reflects the spatial relations most used by the users.

Figure 6. 10 Percentage of spatial relations used by the users

We can emphasize that the users use more frequently spatial relations where an object is above, on, or in another object, also considering when an object is placed on the floor. In both, allusion is made to the object that is under the object to be located.

Another spatial relation most commonly used was the relation when an object is placed near or next to another object, which was used around 21% of the times, much more frequently than the more specific relations behind, below, left, right or in front of another object, which were used 13% of the times. This seems to indicate that users prefer using imprecise spatial relations like near, instead of describing the exact relation that holds between the OL and the RO.

The spatial relation *between* was quite often used, since it turns out to be very effective if we manage to identify two objects to the sides of the target object.

One of the relations that we have observed that the users employed is when a user refers to an object that is placed in a furniture with different shelves. A sentence used by the participants to locate an object, using this relation was:

“The object is located on the third shelf, starting from the bottom up”

Nevertheless, a second spatial relation was used some times as support, to locate the object with major precision. This is observed in the following sentence:

“The object is located on the third shelf starting from the bottom up, between the red sphere and the yellow cube”

Finally, we found a spatial relation that does allusion to when an object is in touch with both the floor and the wall, in this case the user identifies the object as being supported in the wall. This spatial relation only was used in 2 % of the cases, due to the scarce situations in which we could find an object like that. Nevertheless we have considered this spatial relation important and therefore we have included it inside the system.

6.2.2 Criteria for the computation of the spatial relations

Based on the above, an algorithm was designed with the established criteria to determine the position of a reference object, with regard to the object to be located.

The absolute coordinates of the objects within the virtual environment turn out to be not very effective for this purpose, because for the spatial relations to be valid they should consider the point of view of the user. Therefore, the following describes criteria considered for the computational procedure of the spatial relations used.

On / Under

For the evaluation of this spatial relation it is checked which objects are in contact with the object to be located; for this, the collisions of the objects that are in contact with the object to be located are detected. If any of these collisions is the reference object it would be the case of a spatial relation On / Under. Only the absolute coordinates are compared in the axis 'Y' of both objects, for knowing which object is above or below.

Colliders are components that enclose the shape of an object for the purpose of evaluating physical collisions. A mesh is a set of triangles arranged in 3D space that create the impression of a solid object.

Close to

For determining if an object is near another, the distances between the camera (representing the user's position) and the object to be located, along with the distance between the two objects, are considered. Figure 6.11 illustrates better this spatial relation.

Figure 6. 11 Representation of the spatial relation “close to”

where

- d_g represents the distance of the camera to the object to be located.
- d represents the distance between the reference object and the object to be located.

In this spatial relation a distance threshold called d_t , is defined. This distance is established under the criterion of the designer, and acts as a range around the object to be located. If the user is inside the established range, the "close to" this relation will not be considered, because it could be used a binary spatial relation. Therefore, an object is near another as long as the double of the distance d is minor that d_g , and d_g is greater that a distance defined by the designer d_t . This can be represented by the following logical expression:

$$\text{Close to if } 2*d < d_g \ \&\& \ d_g > d_t$$

Left / Right / Above / Below / In front of / Behind

These six cases are mutually exclusive, that is, these are evaluated simultaneously and if one holds the others are discarded. For this, the edges of the object are determined using its bounding box, for evaluating the coordinates of the object. Likewise, it allows evaluating the nearest points of the user to those of the bounding box of the object.

On the other hand, an operation is applied to obtain the coordinates of a point used as roominess, for the evaluation of the spatial relation. The roominess is a distance that allows us to evaluate the points that fall down inside or out of the above mentioned distance, in such a way that it might give us one or another spatial relation coinciding with the subjective perception of the user. Bearing in mind that the system of units by default in the mesh of an object is in centimetres, the roominess used is defined by the designer with a size of 10 centimetres. This roominess is based on the criteria of that there must be a range that emulates the impressions that the user has about the position of various objects, instead of using an absolute value, an example of this can be seen in Figure 6.12.

Figure 6. 12 Representation of the spatial relations “left / right / above / below / in front of / behind”

Figure 6.12 presents a scene where a sphere and a cube are on a box. If the object to be located is the sphere, the cube might be used as reference object. In this case, the algorithm will apply any roominess, and the coordinates of the nearest points of the objects to camera might be used. Clearly it is observed that the sphere is in front of the cube, if the roominess is not applied, the result of this spatial relation in terms of the natural language would be:

“The sphere is to the left of the cube”

Although the sphere is in front of the cube, the logic of the algorithm considers that the sphere is to the left of the cube, since there is not a roominess, applying the above mentioned criterion. The algorithm skips the conditions to determine if is to the left side, to the right, above or below, and then it evaluates if the objects are in front or behind. Finally, the algorithm takes into account the coordinates in the Z axis, giving as result the following sentence in natural language:

“The sphere is in front of the cube”

For this, Cartesian coordinates are defined and the distance to the origin of the orthogonal projections of a given point on each of the axes as is shown in the Figure 6.13.

Figure 6. 13 Spatial Cartesian coordinate system

Inside

The case of the spatial relation "inside" is illustrated in Figure 6.14. In this figure a sphere is presented inside a box. The sphere is the object to be located and the box could serve as the reference object.

Figure 6. 14 Representation of the spatial relation “inside”

To determine if the sphere is or not inside the box, it was decided to use the technique known as ray casting. The ray structure allows us to create a line with an origin and a direction. The function that implements ray casting provides information on the collider with which the ray hit, and the distance of the collision.

In accordance with the above considerations eight rays are defined from the centre of the object to be located. Six rays are directed towards the directions of the Cartesian axes (up, down, right, left, front and back), and other two rays are directed at an angle of 45° with respect to the axes Y and Z, and it is determined if the rays hit the collider of a reference object. For this, if five of eight rays collide with the collider of the object of reference is determined that an object is inside other. This allows determining if an object is in the box with other objects. Figure 6.15 shows that seven of the eight rays collide with the collider of the box, therefore it is possible to affirm that the sphere is inside the box.

Figure 6.15 Representation of the ray collisions with a box

Between

For this spatial relation, several criteria were considered: First, it is examined if there exist several candidates to be objects of reference. Then the two closest objects of reference (RO_1 and RO_2) to the object to be located are identified. Finally, the distances of OL to RO_1 and RO_2 are calculated. Here, if the distance to the most distant object is lower than the distance to the nearest object in more than 20 % of its distance, then it is understood that the object to be located is between both objects of reference (see Figure 6.16).

Figure 6.16 Distances of two OR to an OL

This is, if $d1 \leq 1.2$ of $d2$, then we use the spatial relation "between", in contrary case we use a binary spatial relation relative to OR2.

First, second, third, etc.

This spatial relation is an auxiliary case which is taken into account once the best reference object is chosen, if there are in the view of the user several objects of the same class, for example: several printers, chairs, tables, spheres, columns, etc. In this case it is necessary to help the user discern which of the objects among them is the reference object. An example that exposes this situation is presented in Figure 6.17, this figure shows a scene with 6 columns and a sphere.

Figure 6. 17 Representation of the spatial relation "first, second, third, etc."

The sphere is the object to be located, whereas one of the columns will be the object of reference, obviously the nearest column to the object to be located would be the best candidate to be the reference object. In this case the selected column is the second from the right.

6.3 Algorithm for the generation of directions when the OL is in the user's field of view

In this section the criteria considered for the cases (c.0), (c.1), (c.2), (c.3) cited before are explained.

(c.0) The object to be located (OL) is inside another object or occluded

In this case, we considered that if the object to be located is inside another object or is occluded by another object, then the container object, that is the object that contains the object to be located, or the object that occluded it is transformed into a secondary object to be located OL2.

Next two directions of this case are presented along with the syntax used by the software component.

1. “The trash can is behind the brown desk that is to the right of the chair”

OL SR RO/OL2 SR OR2

2. “The blue ball is inside the white box that is on the brown desk”

OL SR RO/OL2 SR OR2

Where

OL	is the object to be located
SR	is a spatial relation
RO/OL2	is the reference object transformed into the secondary object to be located
OR2	is the reference object for the secondary object to be located

Figure 6.18 Illustrates the second aforementioned example

Figure 6. 18 Direction where the object to be located (OL) is inside another one

Figure 6.19 presents the flow diagram of the Algorithm to determine the generation of directions of the cases (c.1), (c.2) and (c.3).

Figure 6. 19 Flow diagram of the algorithm to generate directions

(c.1) The OL is salient in itself

For the implementation of this case, it was established that if the object to be located has a high perceptual saliency, that is, it is salient in itself with a value in $GS_{ss} > 4.8$; a direction will be generated with a spatial relation of the intrinsic type (listener centred). This type of spatial relation takes the listener as spatial reference system. Figure 6.20 presents the flow diagram of this case.

Figure 6. 20 Flow diagram of the case c.1

Next a pair of sample directions where two cases occur, along with their syntax.

1. "The box is located in front of you"
 OL SR RO

2. "The brown box is located in front of you"

OL SR RO

Figure 6.21 shows a view of a direction given by the system when the OL is salient in itself, within the environment.

Figure 6. 21 Listener centred intrinsic direction where the OL is salient in itself

(c.2) The RO has a high probability of being remembered by the user, because he has a high prior knowledge of the environment

For this case (c.2) we first analyse the value of K_e , the general knowledge of the environment, without caring about the value of perceptual saliency of the objects or the value of the previous knowledge that the user has on them (the concept of K_e was described in the section 6.1 of this chapter).

Here we considered that if the user has a $K_e > 0,8$ (value proposed as an initial approach, to be refined with experience) then, the object with the highest probability of being remembered (Pr) inside the radius of nearness of the OL will be our reference object. Then a spatial relation is determined between the object to be located and the reference object and a direction is generated in terms of the spatial relation.

Figure 6.22 presents the flow diagram of this case.

Figure 6. 22 Flow diagram of the case c.2

Next some sample directions of this case are presented along with their syntax.

1. "The pen is located on the desk"
OL SR RO
2. "The red pen is located on the desk"
OL SR RO
3. "The pen is located on the brown desk"
OL SR RO
4. "The red pen is located on the brown desk"
OL SR RO

Figure 6. 23 Direction where the user is very familiar with the environment and the RO has a high probability of being remembered

(c.3) No RO has a high probability of being remembered by the user, because he has a low prior knowledge of the environment

When $Ke \leq 0,8$, two more cases can occur (c.3.1 and c.3.2), which are described below.

(c.3.1) In this case we rely on another criterion, the maximum value of the saliency as a standard score GS_{ss} , obtained by transforming into standard score the value of saliency of the color, the size and the shape of an object with respect to the context (see section 4.3 Modelling the Contextual Saliency). Each of these characteristics in standard score has a very low probability of taking values beyond 3 or -3. Therefore, the value of saliency with respect to the context of one object, as a standard score, would generally take values between -9 and 9. The higher the value of the standard score for the saliency in any of the object's characteristics, or for the global object's saliency with respect to the context, the more atypical the object will be with respect to its context. A value of 9 would denote that the object's color, size and shape are three times the standard deviation separated from the average in the context. Therefore, we assume that this object would be a perfect candidate to be selected as RO because it clearly stands out from the context from the perceptual point of view.

As an initial approach we established a threshold of 4.8 for GS_{ss} to determine when an object is highly salient with respect to the context. This parameter, as well as: ScC_{ss} , $SsiC_{ss}$, and $SshC_{ss}$, were proposed by us, to establish the criteria of these cases of application.

Taking into consideration this criterion, all the objects that are candidates for being reference objects, which are also inside the defined closeness threshold around the OL, are ordered with regard to the value of GS_{ss} . The object with the maximum value is chosen.

Now two subcases can occur, which are described next:

- (a) If the selected reference object has $ScC_{ss} > 1.8$ or $SsiC_{ss} > 1.8$, and also the user has a $PK \geq 0.5$ of this object, we consider that the user can recognize and name the object, so the object will be indeed considered as the best reference object.

Figure 6.24 presents the flow diagram of this case.

Figure 6. 24 Flow diagram of the case c.3.1 (a)

Next a sample direction of this case is presented in the figure 6.25.

Figure 6. 25 Direction where the OR has not probability of been remembered

- (b) If the value of $PK = 0$, that means that the user does not have any previous knowledge of this type of objects, so even if the object perceptually stands out, it would not be wise to generate a direction that names the object's type. However, if its $ScCss > 1.8$ (its saliency by color with respect to the context in standard score is quite high), and do not exist anymore objects of the same color, or its $SsiCss > 1.8$ (its saliency by size with respect to the context in standard score is quite high), and do not exist anymore objects of the same size, the object will still be considered as the best reference object, and the direction will refer to the RO by its color or size, respectively.

Figure 6.26 presents the flow diagram of this case.

Figure 6. 26 Flow diagram of the case c.3.1 (b)

Next two sample directions of this case are presented.

Figure 6.27 illustrate a situation where the color of the OR can be highlighted in the direction to help to locate the OL.

Figure 6. 27 Direction where the color of OR can help to locate the object

Figure 6.28 illustrate a situation where the size of the OR can help to locate the object.

Figure 6. 28 Direction where the size of OR can help to locate the object

(c.3.2) In this case we are in a scenario where none of the objects is very salient, that is, there is no object whose $GS_{ss} > 4.8$. There are in turn three subcases:

(a) In this first sub-case:

- The selection of the RO will be based in the maximum value of the sum between GS, PR, and PK (see the Equation 33). For this, an ordered list of all the objects candidates to be RO is obtained.
- The first object in the list (the one with the highest value) that also verifies the condition $PK \geq 0,5$ (the user at least can recognize and name the type of object), will be selected as RO.

Figure 6.29 presents the flow diagram of this case.

Next a sample direction of this case is presented in the figure 6.30.

Figure 6. 30 Direction where the OR is selected by GS, PR and KE

(b) In this second sub-case:

- There are no other objects in the environment that are useful as RO, that is, if $K_e = 0$ then the first object is discarded and the second one is analysed. This way up to coming to a minimum, where none of the objects could be selected as RO.
- Then we consider if the size of OL with respect to the context (v_{so}), that is the object's visible surface, is greater than 0.05 in a scale of 0 to 1. In this case, we consider that the OL can at least be visually identified so we can generate an intrinsic (listener centred) direction that guides the sight of the user towards the object.
- Also is verified that $PK > 0,5$.

Figure 6.31 presents the flow diagram of this case.

Figure 6. 31 Flow diagram of the case c.3.2 (b)

Next a sample direction of this case is presented in the figure 6.32.

Figure 6. 32 Direction based in the seen surface of OL

(c) For this third sub-case:

- When the seen surface of the OL is lower than 0,05, then we consider the object with the biggest visible surface in the radius of nearness of the OL. We consider a big surface of the object if its size is > 0.60 in a scale of 0 to 1. With this parameter we assure to give a direction with respect to a big enough object from the point of view of the user.
- Also is verified that $PK > 0,5$. If these two conditions are fulfilled, then is generated a direction.
- This sub-case will combine two types of directions: allocentric and intrinsic (listener centred).

This case, as well as the case (c.0) selects a reference object, for the object to be located, but when this one is not sufficient to locate the object, then a second reference object is selected with regard to the reference object. Therefore, it would be necessary to include in the direction each of the corresponding spatial relations between the object to be located and the first reference object, and between the first reference object and the second reference object.

Figure 6.33 presents the flow diagram of this case.

Figure 6. 33 Flow diagram of the case c.3.2 (c)

Next some sample directions of this case are presented along with their syntax.

1. "The pen is located on the desk, which is located next to the chair"

OL SR RO/OL2 SR OR2

2. "The red pen is located on the desk, which is located in front of you"

OL SR RO/OL2 SR OR2

3. "The red pen is located on the big desk, which is located next to the chair"

OL

SR RO/OL2

SR

OR2

4. "The red pen is located on the big desk, which is located next to the black chair"

OL

SR RO/OL2

SR

OR2

6.5 Conclusion

In this chapter were presented several processes that integrate the software component that generates directions for the location of objects inside virtual environments, considering spatial and perceptual aspects. Next the most significant contributions of this chapter are mentioned:

- An algorithm to mathematically calculate several criteria considered in selecting the best reference object is presented.
- It is also worth mentioning the exploratory study realized for selecting the spatial relationships to be implemented within the component for the generation of indications. This study supports the theoretical bases mentioned by (Piaget and Inhelder, 1947) and (Gapp, 1994, 1995).
- Likewise, were described the computational criteria considered for the implementation of twelve spatial relations.

It was described a broad set of cases of application treated for their implementation in the software component which generates directions in an intelligent way. These directions depend on the context of every case.

Chapter 7: Conclusion and Future Work

7 Conclusion and feature work

This chapter concludes the investigation with four sections devoted: to discuss on topics learned and dealt with during the development of the thesis; to specify the contributions reached; to propose future research lines derived from the results obtained in the thesis; and to identify dissemination products generated during the project, which were put to review by the international scientific community.

7.1 Discussion

In essence, the thesis was directed to propose a computational model for the generation of directions in natural language for helping in the location of objects, considering spatial and perceptual aspects.

We analyzed topics ranging from virtual reality (VR), through spatial language, spatial reference systems, perceptual saliency, color representation models, methods for object shape representation, to the field of user modeling with data ontologies.

Furthermore, we analyzed several computational models of saliency and the purpose for which they were developed. Similarly, computer systems for the generation of natural language directions for object location were analyzed.

These two analyses allowed us to identify gaps and research challenges and led us to propose a computational model, based on aspects of perceptual saliency of 3D objects as well as on modeling relevant user's perceptual and cognitive abilities, and the definition of a set of algorithms for the generation of directions in natural language for the localization of objects. The analysis of the state of the art allowed us to check that there was no model covering all those aspects within the framework of a virtual environment.

Our model is characterized by incorporating the representation of the salient physical characteristics of a 3D object (color, size and shape) towards the human visual system. Also by representing the characteristics of users which are related to the process of locating objects (location memory, visual acuity, history of view, among others).

7.2 Contributions

According to the research objectives proposed in Section 1.1, the following results were achieved, listed in the order in which they were presented in previous chapters:

- A metric that measures the saliency of the shape of 3D objects in a simple, fast and effective way. This metric can be applied to all kinds of regular and irregular objects. It is also adaptable to all scales of 3D models. The efficiency of this metrics was validated with the design of an experiment, the compilation of empirical information, the analysis of the collected information and the discussion of the results.
- A computational model of perceptual saliency based on color, size and shape of 3D objects.

The definition of the model comprises a set of stages:

- (1) Characterization of the color, size and shape of each object.
- (2) Definition of a process that allows establishing a value of individual saliency for each object.
- (3) Definition of a process that allows calculating the saliency, with respect to the context, of each object in the virtual environment.
- (4) Definition of a metric for the general saliency of all the objects in the environment.

Likewise, the efficiency of our model was validated with the design of two experiments, which allowed us the compilation of empirical information, the analysis of the collected information and the discussion of the results.

- A semantic representation of virtual environments that allows explicitly representing the properties and the relations between objects in a virtual environment using data ontologies.

- A user model which can represent perceptual and cognitive characteristics of a user, which are relevant in the process of locating objects.
- A computational model for the automatic generation of directions in the location of objects within VEs. The design of model was based on three auxiliary algorithms:

(1) An algorithm for the selection of the best reference objects (SBRO).

(2) An algorithm for the selection of a relevant spatial relation (SSR) between the object to be located (OL) and the reference object (RO).

(3) An algorithm for the generation of directions (GD), that distinguishes between several possible cases regarding the position, visibility and perceptual properties of the object to be located with respect to the observing user.

Our model is characterized for being intelligent and adapt to the profile and previous experience of each user, as well as to the current context, in the process of location of objects within the virtual environment.

7.3 Future research lines

From the results derived from the research project and the fruits achieved, it was possible to identify several courses of action to extend the scope of research and application of the computational model for the generation of directions for location of objects. The following lines of future work were identified:

- Considering the textures of the objects at the moment of determining their saliency by color.
- Treating in our metric the variability of appearances that the shape of an object could have due to changes of perspective and occlusions of the shape.

- Performing a systematic comparison of the metric of saliency of the shape for 3D objects with respect to other related metrics found in the literature.
- Extending the computational model of saliency, to include additional factors such as the utility of objects, their emotional meaning, mobility, etc.
- Validating the generation of indications experimentally in terms of their usefulness and naturalness for users.
- Integrating this model to an intelligent tutoring system that requires the exchange of information between users.

7.4 Scientific dissemination

The scientific production derived from the project of investigation and doctoral thesis, has been up to date translated into three publications, of which two appear in journals indexed in ISI's Journal Citation Reports:

Publications in international archival journals

- Lara G., Peña A., De Antonio A., Ramírez J., Imbert R. *Comparative analysis of shape descriptors for 3D objects*. Multimedia Tools and Applications. Indexed journal with Impact Factor (1.346), Q2, ISSN: 1573-7721 (online). See <http://link.springer.com/article/10.1007/s11042-016-3330-5>. 1 – 48. (Accepted and published online).
- Lara G., De Antonio A., Peña A., *Computerized spatial language generation for object location*. Virtual Reality. Indexed journal with Impact Factor (0,667), Q4, ISSN 1434-9957 (online). DOI: 10.1007/s10055-016-0289-5. (Accepted, in edition).

Publications in international conferences with referees

- Lara G., De Antonio A., Peña A. *A computational measure of saliency of the shape of 3D objects*. In Proceedings 4th International Conference on Software Process Improvement. Springer. Advances in Intelligent Systems and Computing. Campus of the Faculty of Computer Mazatlan (FIMAZ) of the Autonomous University of Sinaloa (UAS) at Mazatlan, Sinaloa, Mexico. ISBN 978-3-319-26285-7. See <http://www.springer.com/br/book/9783319262833>. 235 – 245. (Accepted and published online)

References

References

- Abella, A., and J. R. Kender, 1999, From Images to Sentences via Spatial Relations: In Proc. of the W. on Integration of Speech and Image Understanding, p. 117 - 146.
- Achanta, R., S. Hemami, F. Estrada, and S. Süsstrunk, 2009, Frequency-tuned Salient Region Detection: *In IEEE CVPR*, p. 1597 - 1604.
- Aguado-Aguilar, L., 2001, Aprendizaje y memoria: *Revista de neurología*, v. 32(4), p. 373 - 381.
- Akgül, C. B., 2007, Density-Based Shape Descriptors and Similarity Learning for 3D Object Retrieval, Boğaziçi University, Istanbul. PhD. Thesis, Paris, 1 -153 p.
- Akgül, C. B., B. Sankur, Y. Yemez, and F. Schmitt, 2009, 3D Model Retrieval Using Probability Density-Based Shape Descriptors: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 31(6), p. 1117 - 1133.
- Anagnostopoulos, C.-N., G. Tsekouras, I. Anagnostopoulos, and C. Kalloniatis, 2007, Intelligent Modification for the Daltonization Process of Digitized Paintings: Proceedings of the 5th International Conference on Computer Version Systems. Applied Computer Science Group.
- André, E., G. Bosch, G. Herzog, and T. Rist, 1986, Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions, Brighton, UK, Project VITRA, Universität des Saarlandes, p. 1 - 8.
- André, E., G. Bosch, G. Herzog, and T. Rist, 1987, Coping with the Intrinsic and Deictic Uses of Spatial Prepositions, *Artificial Intelligence II: Methodology, Systems, Applications (AIMSA)*, Amsterdam: North-Holland, p. 375 - 382.
- André, E., G. Herzog, and T. Rist, 1988, On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: The System SOCCER: In Proceeding of 8th European Conference on Artificial Intelligence (ECAI-88), p. 449 - 454.
- André, E., G. Herzong, and T. Rist, 1989, Natural Language Access to Visual Data: Dealing with Space and Movement: In: Proceedings of the 1st Workshop on Logical Semantics of Time, Space and Movement in Natural Language, p. 1 - 21.
- Ankerst, M., G. Kastenmüller, H.-P. Kriegel, and T. Seidl, 1999, 3D Shape Histograms for Similarity Search and Classification in Spatial Databases Proceedings of the 6th International Symposium on Spatial Database.
- Ballesteros, S., 1999, Memoria humana: investigación y teoría: *Psicothema*, v. 11(4), p. 705 - 723.
- Barclay, M., 2010, Reference Object Choice in Spatial Language: Machine and Human Models, University of Exeter. PhD. Thesis, 274 p.
- Barla, M., 2010, Towards Social-based User Modeling and Personalization Slovak University of Technology in Bratislava, Slovak, 1 - 162 p.
- Barrios, J. M., and B. Bustos, 2011, Automatic Weight Selection for Multi-Metric Distances: ACM. In Proceeding of the 4th International Conference on Similarity Search and Applications (SISAP'11), p. 61 - 68.
- Bataller, S. B., and J. Moral, Meléndez, 2006, Cambios en la memoria asociados al envejecimiento: *Geriátrika*, v. 22(5), p. 179 - 185.
- Bates, J., 1991, Virtual Reality, Art, and Entertainment: *The Journal of Teleoperators and Virtual Environments*, MIT Press, Carnegie Mellon School of Computer Science technical report CMU-CS-91-133, v. Carnegie Mellon School of Computer Science technical report CMU-CS-91-133, p. 1 - 9.
- Behley, J., V. Steinhage, and A. B. Cremers, 2012, Performance of Histogram Descriptors for the Classification of 3D Laser Range Data in Urban Environments *IEEE International Conference on Robotics and Automation (ICRA)*, Saint Paul, MN, IEEE, p. 4391 - 4398.
- Bespalov, D., W. C. Regli, and A. Shokoufandeh, 2006, Local Feature Extraction and Matching Partial Objects: Elsevier. *Computer-Aided Design*, v. 38(9), p. 1020 - 1037.
- Botella, C., R. M. Baños, C. Perpiña, H. Villa, M. Alcañiz, and A. Rey, 1998, Virtual Reality Treatment of Claustrophobia: a Case Report: Elsevier. *Science Ltd. Behaviour Research and Therapy.*, v. 36, p. 239 - 246.
- Bowman, D. A., and R. P. McMahan, 2007, Virtual reality: how much immersion is enough?: *IEEE Computer*, v. 40(7), p. 36 - 46.

- Brooks, F. P., 1999, What's Real About Virtual Reality?. *Special Report: IEEE Computer Graphics and Applications*, v. 19, p. 16 - 27.
- Bryson, S., 1996, Virtual Reality in Scientific Visualization: *Communications of the ACM*, v. 39, p. 62 - 71.
- Bustos, B., D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranić, 2005, Feature-based similarity search in 3D object databases: *ACM. Computing Surveys (CSUR)*, v. 37, p. 345-387.
- Caduff, D., and T. Sabine, 2008, On the assessment of landmark salience for human navigation: *Springer. Cognitive Processing.*, v. 9, p. 249 - 267.
- Carlson-Radvansky, L. A., and G. D. Logan, 1997, The Influence of Reference Frame Selection on Spatial Template Construction: *Journal of Memory and Language*, v. 37, p. 411 - 437.
- Choungourian, A., 1968, Color Preferences and Cultural Variation: AmSci, Southern Universities Press, p. 1203 - 1206.
- Chung, F. R. K., 1997, Spectral Graph Theory. Regional Conference Series in Mathematics American Mathematical Society, v. 92: Washington, DC., Published for the Conference Board of the Mathematical Sciences.
- Connolly, C., and T. Fliess, 1997, A Study of Efficiency and Accuracy in the Transformation from RGB to CIELAB Color Space *IEEE. Transactions on Image Proccessing*, v. 6, p. 1046 - 1048.
- Costello, F. J., and J. D. Keller, 2006, Spatial Prepositions in Context: The Semantic of near in the Presence of Objects, in *A. f. C. Linguistic*, ed., *Third ACL-SIGSEM Workshop on Prepositions*, Trento, Italy.
- Coventry, K. R., and S. C. Garrod, 2004, Saying, Seeing and Acting: The Pyschological Semantics of Spatial Prepositions, Psychology Press; Edición: New.
- Craik, F. I., and R. S. Lockhart, 1980, Niveles de procesamiento: Un marco para la investigación sobre la memoria: *Estudios de psicología*. Taylor & Francis, v. 1(2), p. 93 - 109.
- Cruz, M. I. A., 2007, Rotaciones multidimensionales generales, Universidad de las Américas Puebla. Thesis of Master, 173 p.
- Doliotis, P., G. Tsekouras, C. N. Anagnostopoulos, and V. Athitsos, 2009, Intelligent Modification of Colors in Digitized Paintings for Enhancing the Visual Perception of Color-blind Viewers: *IFIP International Federation for Information Processing*, p. 293 - 301.
- Domínguez, S., 2012, Medida de similitud basada en saliencia: *Revista Iberoamericana de Automática en Informática Industrial*, v. 9, p. 359 - 370.
- Doron, R., and F. Parot, 2004, Diccionario Akal de Psicología, AKAL, p. 09 - 10.
- Dos Santos, F. J., 2007, Retrieval of 3D Models using Partial Matching Universidade Tecnica de Lisboa. PhD. Thesis, Lisboa.
- Dutağaci, H., B. Sankur, and Y. Yemez, 2005, Transform-based Methods for Indexing and Retrieval of 3D Objects: *IEEE. Proceedings of the Fifth International Conference on 3D - Digital Imaging and Modeling*. Computer Society. (3DIM'05), p. 188 - 195.
- Egeth, H. E., and S. Yantis, 1997, Visual attention: control, representation, and time course: *Annual Review of Psycology*, v. 48, p. 269 - 297.
- EINaghy, H., S. Hamad, and E. Khalifa, 2013, Taxonomy for 3D Content-Based Object Retrieval Methods: *IJARRAS*, v. 14(2), p. 412 - 446.
- Elias, B., and C. Brenner, 2005, Automatic Generation and Application of Landmarks in Navigation Data Sets: *11th International Symposium on Spatial Data Handling*, p. 469 - 480.
- Esquivel-Rodriguez, J., and D. Kihara, 2012, Fitting multimeric protein complex into electron microscopy maps using 3D Zernike descriptors: *The Journal of Physical Chemisty B*, v. 116(23), p. 6854 - 6861.
- Ferreira-Satler, M., V. H. Menéndez, F. P. Romero, A. Zapata, and M. E. Prieto, 2010, Ontologías borrosas para respresentar perfiles de usuario en una herramienta de gestión de objetos de aprendizah: In *Actas del XV Congreso español sobre Tecnologías y Lógica Fuzzy*. ESTYLF, p. 537 - 542.
- Fischer, G., 2001, User modeling in human-computer interaction: User modeling and user-adapted interaction, v. 11(1-2).
- Frias-Martinez, E., G. Magoulas, S. Chen, and R. Macredie, 2005, Modeling human behavior in user-adaptive systems: Recent advances using soft computing techniques: *Expert Systems with Applications*, v. 29(2), p. 320 - 329.
- Frintrop, S., and E. Rome, 2010, Computational Visual Attention Systems and their Cognitive Foundations: A Survey: *ACM. Journal Name*, v. 7, p. 1 - 46.
- Gapp, K.-P., 1994, From vision to language: A cognitive approach to the computation of spatial relations in 3D space, Fed. Rep. of Germany, Universität des Saarlandes, p. 1 - 19.

- Gapp, K.-P., 1995, Object Localization: Selection of Optimal Reference Objects, *Fed. Rep. of Germany, Universität des Saarlandes* p. 1-18.
- Gapp, K.-P., 1995a, Angle, distance, shape, and their relationship to projective relations.: In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, p. 112 - 117.
- Gapp, K.-P., 1996a, Processing Spatial Relations in Object Localization Tasks, *Fed. Rep. of Germany, Universität des Saarlandes*, p. 1 - 7.
- Gapp, K.-P., 1996b, Selection of Best Reference Objects in Objects Localizations: In *Proceedings of the AAAI Spring Symposium on Cognitive and Computational Models of Spatial Representations*, p. 23 - 34.
- Gascuña, J. M., A. Fernández-Caballero, and P. González, 2005, Ontologías del modelo del alumno y del modelo del dominio en sistemas de aprendizaje adaptativos y colaborativos, In *VI Congreso Interacción Persona Ordenador*, Universidad de Granada.
- Gevers, T., and A. W. M. Smeulders, 1999, Color-based object recognition: Elsevier. *The Journal of the Pattern Recognition Society.*, v. 32, p. 453 - 464.
- Gomes, D. S. A., and G. Zachmann, 1999, Virtual Reality as a Tool for Verification os Assemble and Maintenance Processes: Elsevier. *Computers & Graphics*, v. 23.
- González, G., 2014, Ontología del perfil de usuario para personalización de sistemas de u-learning universitarios, In *XLIII Jornadas Argentinas de Informática e Investigación Operativa (43JAIIO)-XVII. Concurso de Trabajos Estudiantiles Buenos Aires*, p. 90 - 114.
- Gossett, N., and B. Chen, 2004, Paint Inspired Color Mixing and Compositing for Visualization, in *IEEE, ed., Symposium on Information Visualization*, Austin, Texas, USA, p. 113 - 117.
- Goto, E., Y. Yagi, Y. Matsumoto, and K. Tsubota, 2002, Impaired functional visual acuity of dry eye patients: Elsevier. *American journal of ophthalmology*, v. 133(2), p. 181 - 186.
- Gruber, T. R., 1993, A translation approach to portable ontology specifications: *Knowledge acquisition*, v. 5(2), p. 199 - 220.
- Guo, C., Q. Ma, and L. Zhang, 2008, Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform, *Conference on Computer Vision and Pattern Recognition, IEEE*, p. 1 - 8.
- Guo, Y., M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, 2015, A Comprehensive Performance Evaluation of 3D Local Feature Descriptors *International Journal of Computer Vision*, p. 1 - 24.
- Gutierrez, M., F. Vexo, and D. Thalmann, 2005, Semantics-based representation of virtual environments: *International journal of computer applications in technology*, v. 23(2-4), p. 229 - 238.
- Gutiérrez, A., 1991, Procesos y habilidades en visualización espacial, In *Memorias del 3er Congreso Internacional sobre investigación en Educ. Mat., Valencia, España*, p. 44 - 59.
- Gutiérrez Garralda, J. M., and J. Fernández Ruíz, 2011, Sustrato neuronal de la memoria de trabajo espacial, *Revista eNeurobiología*, p. 1 - 12.
- Hall, D., B. Leibe, and B. Schile, 2002, Saliency of Interest Points under Scale Changes, *British Machine Vision Conference (BMVC'02)*, Cardiff, UK, p. 646 - 655.
- Harrington, D. O., and M. V. Drake, 1993, Los campos visuales: texto y atlas de perimetría clínica., Ediciones Científicas y Técnicas.
- Herrero, P., and A. De Antonio, 2005, Intelligent virtual agents keeping watch in the battlefield: Springer. *Virtual Reality*, v. 8(3), p. 185 - 193.
- Herzog, G., 1992, Visualization Methods for the VITRA Workbench, Saarbrücken, Germany, *Universität des Saarlandes*, p. 1 - 16.
- Herzog, G., 1995, From Visual Input to Verbal Output in the Visual Translator In: *Proc. of the AAAI Fall Symposium on Computational Models for Integrating Language and Vision*, p. 1 - 15.
- Hill, B., T. Roger, and F. W. Vorhagen, 1997, Comparative Analysis of the Quantization of Color Spaces on the Basis of the CIELAB Color-Difference Formula: *ACM. Transactions on Graphics*, v. 16, p. 109 - 154.
- Hoffman, D. D., and M. Singh, 1997, Saliency of visual parts: Elsevier. *Cognition*, v. 68, p. 29 - 78.
- Hou, X., and L. Zhang, 2007, Saliency Detection: A Spectral Residual Approach, *Computer Vision and Pattern Recognition*, Minneapolis, MN, IEEE, p. 1 - 8.
- Howarth, P. A., and P. J. Costello, 1997, Contemporary ergonomics, in R. e. T. a. Francis, ed., London, Ed. SA, p. 109 - 116.
- Huang, L., and H. Pashler, 2005, Quantifying object salience by equating distractor effects: Elsevier. *Vision Research*, v. 45, p. 1909 - 1920.
- Huang, P., J. Starck, and A. Hilton, 2007, A Study of Shape Similarity for Temporal Surface Sequences of People, *International Conference on 3-D Imaging and Modeling - 3DIM* p. 408 - 418.

- Iachini, T., I. Sergi, G. Ruggiero, and A. Gnisci, 2005, Gender differences in object location memory in real three-dimensional environments: *Brain and Cognition*, v. 59(1), p. 52 - 59.
- Itti, L., 2006, Quantitative Modelling of Perceptual Saliency at Human Eye Position: Taylor & Francis Group. Psychology Press. Visual Cognition, v. 14, p. 959 - 984.
- Itti, L., C. Koch, and E. Niebur, 1998, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis: *IEEE Transactions on Pattern Analysis and Machine Intelligence* v. 20(11), p. 1254 - 1259.
- Jefferson, L., and R. Harvey, 2007, An Interface to Support Color Blind Computer Users, CHI 2007 Proceedings • Color/Blind, San Jose, California, USA., ACM, p. 1535 - 1538.
- Kadir, T., and M. Brady, 2001, Saliency, Scale and Image Description: *International Journal of Computer Vision*, v. 45, p. 83 - 105.
- Kapur, S., 2003, Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia: *Am J Psychiatry*, v. 160, p. 13 - 23.
- Katramados, I., and T. Breckon, 2011, Real-Time Visual Saliency by Division of Gaussians: Proceedings of the 18th IEEE International Conference on Image Processing, p. 1701 - 1704.
- Kazhdan, M., T. Funkhouser, and S. Rusinkiewicz, 2003b, Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors Eurographics Symposium on Geometry Processing. The Eurographics Association, p. 156 - 164.
- Keim, D. A., 1999, Efficient Geometry-based Similarity Search of 3D Spatial Databases ACM. In Proceedings of the International Conference on Management of Data (SIGMOD'99), p. 419 - 430.
- Kelleher, J. D., 2003, A Perceptually Based Computational Framework for the Interpretation of Spatial Language Dublin City University. PhD. Thesis, Dublin, 1 - 463 p.
- Kelleher, J. D., and F. J. Costello, 2009, Applying Computational Models of Spatial Prepositions to Visually Situated Dialog: *Computational Linguistics*. Association for Computational Linguistics, v. 35(2), p. 271 - 306.
- Kihara, D., L. Sael, R. Chikhi, and J. Esquivel-Rodriguez, 2011, Molecular surface representation using 3D Zernike descriptors for protein shape comparison and docking: *Current Protein and Peptide Science*, v. 12(6), p. 520 - 530.
- Kim, Y. K., K. W. Kim, and X. Yang, 2007, Real time traffic light recognition system for color vision deficiencies: *IEEE International Conference on Mechatronics and Automation*, p. 76 - 81.
- Klatzky, R. L., 1998, Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections: *Spatial Cognition*. Lecture Notes in Computer Science, v. 1404, p. 1 - 17.
- Klippel, A., and S. Winter, 2005, Structural Saliency of Landmarks for Route Directions: Springer. *Spatial Information Theory*. International Conference, p. 347 - 362.
- Knublauch, H., R. W. Ferguson, N. F. Noy, and M. Musen, 2004, The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications: In *The Semantic Web-ISWC*, p. 229 - 243.
- Kobsa, A., 2001, *Generic User Modeling Systems*: Springer. *User Modeling and User-Adapted Interaction*, v. II, p. 49 - 63.
- Kröser, B. J., and B. Julesz 1989, The control and speed of shifts of attention.: *Vision Research*, v. 29(11).
- Ku, M., and K. Aoyama, 2006, Color Universal Design Handbook, in E. N. CORPORATION, ed., Japan, p. 1 - 13.
- Laga, H., H. Takahashi, and M. Nakajima, 2006, Spherical Wavelet Descriptors for Content Based 3D Model Retrieval: *IEEE Proceedings of the International Conference on Shape Modeling and Applications*, p. 15 - 25.
- Lahera, G., N. Freund, and J. Sáin-Ruiz, 2013, Asignación de relevancia (saliency) y desregulación del sistema dopaminérgico: Elsevier Doyma. *Revista de Psiquiatría y Salud Mental*, v. 6, p. 45 - 51.
- Landau, B., and R. Jackendoff, 1993, "What" and "Where" in Spatial Language and Spatial Cognition: *Behavioral and Brain Sciences*, v. 16.
- Lara, G., A. De Antonio, and A. Peña, 2015, A computational measure of saliency of the shape of 3D objects: 4th International Conference on Software Process Improvement, p. 235 - 245.
- Lara, G., A. De Antonio, and A. Peña, 2016b, (in press) Computerized spatial language generation for object location: Springer. *Virtual Reality*.
- Lara, L. G., P. N. A. Peña, J. A. De Antonio, R. J. Ramírez, and P. R. Imbert, 2016a, (in press) Comparative analysis of shape descriptors for 3D objects: *Multimedia Tools Applications*, p. 1 - 48.
- Latecki, L. J., R. Lakämper, and U. Eckhardt, 2000, Shape Descriptors for Non-rigid Shapes with a Single Closed Contour: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* p. 424 - 429.

- Lazebnik, S., Cordelia, and J. Ponce, 2005, A Sparse Texture Representation Using Affine Invariant Regions: IEEE. Proceeding Computer Society. Computer Vision and Recognition Pattern, p. 1265 - 1278.
- Levinson, S. C., 1996, Language and Space. Chapter Frames of Reference and Molyneux's Question: Crosslinguistic Evidence: Cambridge, MIT Press.
- Levinson, S. C., 2003, Space in Language and Cognition: Explorations in Cognitive Diversity: Cambridge. University Press, LCC5. Language, culture & cognition, 415 p.
- Li, C., and B. Hamza, 2013, A multiresolution descriptor for deformable 3D shape retrieval: *The Visual Computer*, v. 29(6-8), p. 513 - 514.
- Li, J., M. Levine, X. An, and H. He, 2011, Saliency Detection Based on Frequency and Spatial Domain Analyses: In Jesse Hoey, Stephen McKenna and Emanuele Trucco, In *Proceedings of the British Machine Vision Conference*, p. (86)1 - 11.
- Li, J., M. D. Levine, X. An, X. Xu, and H. He, 2013, Visual Saliency Based on Scale-Space Analysis in the Frequency Domain: IEEE. Transactions on Pattern Analysis and Machine Intelligence, v. 35, p. 996 - 1010.
- Lienhard, J. H., 2010, Binocular Vision, 2581, The Engines of Our Ingenuity.
- Ling, H., and D. W. Jacobs, 2007, Shape Classification Using the Inner-Distance: IEEE Transactions. Pattern Analysis and Machine Intelligence, v. 29(2), p. 286 - 299.
- Liu, Y.-J., Y.-F. Zheng, L. Lu, and Y.-M. Xuan, 2012, 3D model retrieval based on color + geometry signatures: Springer-Verlag. Vis Comput, v. 28, p. 75 - 86.
- Lockwood, K., K. Forbus, D. T. Halstead, and J. Usher, 2006, Automatic Categorization of Spatial Prepositions: In Proceedings of the 28th Annual Conference of the Cognitive Science Society.
- Lockwood, K., K. Forbus, and J. Usher, 2005, SpaceCase: A Model of Spatial Preposition Use: In Proceedings of the 27th Annual Conference of the Cognitive Science Society.
- Lowe, D. G., 2004, Distinctive Image Feature from Scale-Invariant Keypoints: Kluwer Academic Publishers. Journal of Computer Vision, v. 60(2), p. 91 - 110.
- López, F., J. M. Valiente, R. Baldrich, and M. Vanrell, 2005, Fast Surface Grading Using Color Statistics in the CIE Lab Space *Springer-Verlag Berlin Heidelberg. Pattern Recognition and Image Analysis*, v. 3523, p. 666 - 673.
- Ma, Y., X. Hu, and F. A. Wilson, 2012, The egocentric spatial reference frame used in dorsal-lateral prefrontal working memory in primates: Neuroscience & Biobehavioral Reviews, v. 36(1), p. 26 - 33.
- Machover, C., and S. R. Tice, 1994, Virtual Reality: IEEE. Computer Graphics & Applications, p. 15 - 16.
- Majid, A., M. Bowerman, S. Kita, and D. B. M. Haun, 2004, Can Language Restructure Cognition? The Case for Space: Elsevier. TRENDS in Cognitive Sciences, v. 8, p. 108 - 114.
- Mangold, R., 1986, Sensorische Faktoren Beim Verstehen Überspezifizierter Objektbenennungen. Sensory factors in identification of overspecified objects: Frankfurt am Main: Lang, v. 185.
- Mark, D. M., and A. U. Frank, 1989, Concepts of space and spatial language: In Proceedings, Ninth International Symposium on Computer-Assisted Cartography (Auto-Carto 9), p. 538 - 556.
- Mateus, L. D. C., 2010, Spectral Tools for Unsupervised Modeling of Articulated Objects from Multiple-view Videos, Institut National Polytechnique de Grenoble. PhD. Thesis, French, 171 p.
- Mazuryk, T., and M. Gervautz, 1996, Virtual Reality History, Applications, Technology and Future Vienna, Austria, Vienna University of Technology, p. 1 - 72.
- McBride, B., 2002, Jena: A semantic web toolkit: IEEE Internet computing, v. 6(6), p. 55 - 59.
- McNamara, T. P., 2003, How are the Locations of Objects in the Environment Represented in Memory? : Springer. Spatial Cognition III Lecture Notes in Computer Science, v. 2685, p. 174 - 191.
- Mikolajczyk, K., and C. Schmid, 2006, A performance Evaluation of Local Descriptors: IEEE. Transactions on Pattern Analysis and Machine Intelligence, v. 27(10), p. 1615 - 1630.
- Mitsumoto, H., S. Tamura, K. Okazaki, N. Kajimi, and Y. Fukui, 1992, Reconstruction Using Mirror Images Based on a Plane Symmetry Recovering Method: IEEE. Transactions on Pattern Analysis and Machine Intelligence, v. 14(9), p. 941 - 946.
- Moratz, R., and T. Tenbrink, 2006, Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations Spatial Cognition and Computation, v. 6(1), p. 63-106.
- Moratz, R., T. Tenbrink, J. Bateman, and K. Fischer, 2003, Spatial Knowledge Representation for Human-Robot Interaction Springer-Verlag Berlin Heidelberg. Spatial Cognition III. Lecture Notes in Computer Science, v. 2685, p. 263 - 283.

- Mortensen, E. N., H. Deng, and L. Shapiro, 2005, A SIFT Descriptor with Global Context, IEEE. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), p. 184 - 190.
- Mou, W., and T. P. McNamara, 2002, Intrinsic Frames of Reference in Spatial Memory: The American Psychological Association. *Journal of Experimental Psychology: Learning, Memory, and Cognition* v. 28, p. 162 - 170.
- Moulin, B., and D. Kettani, 1999, Route generation and description using a logical and an analogical framework *Annals of Mathematics and Artificial Intelligence*, v. 24, p. 155 - 179.
- Mujber, T. S., T. Szecsi, and M. S. J. Hashmi, 2003, Virtual Reality Applications in Manufacturing Process Simulation: Elsevier. *Journal of Materials Processing Technology*, p. 1834 - 1838.
- Mukerjee, A., K. Gupta, S. Nautiyal, M. P. Singh, and N. Mishra, 2000, Conceptual description of visual scenes from linguistic models: Elsevier. *Image and Vision Computing*, v. 18, p. 173 - 187.
- Neitz, M., and J. Neitz, 2005, Color Vision Defects, p. 1 - 4.
- Novotni, M., and R. Klein, 2001a, A Geometric Approach to 3D Object Comparison, IEEE. International Conference on Shape Modeling and Applications, IEEE, p. 166 - 175.
- Noy, N. F., and D. McGuinness, 2005, Desarrollo de Ontologías-101: Guía para crear tu primera ontología. Traducido del inglés por: E. Antezana.
- O'Meara, C., and B. G. Pérez, 2011, Spatial frames of reference in Mesoamerican languages: Elsevier. *Language Sciences*, v. 33, p. 837 - 852.
- Oliva, A., A. Torralba, M. S. Castelhano, and J. m. Henderson, 2003, Top-Dow Control of Visual Attention in Object Detection: IEEE. In *Proceedings of the International Conference on Image Processing*, p. 253 - 256.
- Osterrieth, P. A., 1944, Le test de copie d'une figure complexe: *Arch Psychol*, v. 30, p. 206 - 356.
- Paquet, E., M. Rioux, A. Murching, T. Naveen, and A. Tabatabai, 2000, Description of Shape Information for 2-D and 3D Objects: Elsevier. *Signal Processing: Image Communication*, v. 16, p. 103 - 122.
- Passalis, G., I. A. Kakadiaris, and T. Theoharis, 2004, Efficient Hardware Voxelization: IEEE. *Proceedings of the Computer Graphics International*, p. 374 - 377.
- Pederson, E., E. Danziger, D. Wilkins, S. Levinson, S. Kita, and G. Senft, 1998, Semantic Typology and Spatial Conceptualization Linguistic Society of America. JSTOR. *Language*, v. 74, p. 557 - 589.
- Peleg, S., M. Werman, and H. Rom, 1989, A Unified Approach to the Change of Resolution: Space and Gray-Level: IEEE. *Transactions on Pattern Analysis and Machine Intelligence*, v. 11(7), p. 739 - 742.
- Pessoa, L., 2008, On the relationship between emotion: *Nature reviews neuroscience*, v. 9(2), p. 148 - 158.
- Peterson, L., and M. J. Paterson, 1959, Short-term retention of individual verbal items.: *Journal of experimental psychology*, v. 58(3), p. 193 - 198.
- Peña, A. A., 2007, Un modelo del estudiante basado en mapas cognitivos, Instituto Politécnico Nacional. PhD. Thesis, 217 p.
- Piaget, J., and B. Inhelder, 1947, La représentation de l'espace chez l'enfant. París, P.U.F., Trad. inglesa: The child's conception of space, London, Routledge & Kegan Apul, 1956, 1971.
- Pinna, B., 2010, New Gestalt principles of perceptual organization: An extension from grouping to shape and meaning., *Gestalt Theory*, p. 11 - 78.
- Pittarello, F., and A. De Faveri, 2006, Semantic description of 3D environments: a proposal based on web standards.: ACM. In *Proceedings of the eleventh international conference on 3D web technology* p. 85 - 95.
- Proulx, M. J., 2007, Bottom-Up Guidance in Visual Search for Conjunctions: *Journal of Experimental Psychology*, v. 33(1), p. 48 - 56.
- Pértegas, D. S., and F. S. Pita, 2001, La distribución normal, Unidad de Epidemiología Clínica y Bioestadística., Complejo Hospitalario. Universidad de A Coruña, CAD ATEN PRIMARIA, p. 268 - 274.
- Raubal, M., and S. Winter, 2002, Enriching Wayfinding Instructions with Local Landmarks In *Proceedings Second International Conference, GIScience*, p. 243 - 259.
- Regier, T., 1996, The Human Semantic Potential: Spatial Language and Constrained Connectionism *Computational Linguistics*, v. 23(3), p. 483 - 486.
- Regier, T., and L. A. Carlson, 2001, Grounding Spatial Language in Perception: An Empirical and Computational Investigation: *Journal of Experimental Psychology: General*, v. 130, p. 273 - 298.
- Riekheit, G., and I. Wachsmuth, 2006, Situated Communication: Gruyter, v. 7 - 30.
- Rock, I., 1990, The frame of reference, in I. I. Rock, ed., *The Legacy of Solomon Asch: Essays in Cognition and Social Psychology*: NJ: Lawrence Erlbaum Associates, Psychology Press, p. 243 - 268.

- Rothbaum, B. O., L. Hodges, S. Smith, J. H. Lee, and L. Prince, 2000, A controlled study of virtual reality exposure therapy for the fear of flying: *Journal of Consulting and Clinical Psychology*, v. 68(6), p. 1020 - 1026.
- Roy, D. K., 2002, Learning Visually-Grounded Words and Syntax for a Scene Description Task: *Computer Speech and Language*, v. 16, p. 1 - 39.
- Röser, F., K. Hamburger, and M. Knauff, 2011, The Giessen virtual environment laboratory: human wayfinding and landmark salience Springer-Verlag. *Cognitive Processing* v. 12, p. 209 - 214.
- Röser, F., K. Hamburger, A. Krumnack, and M. Knauff, 2012, The structural salience of landmarks: results from an on-line study and a virtual environment experiment: Taylor & Francis Group. *Journal of Spatial Science*, v. 57, p. 37 - 50.
- Röser, F., A. Krumnack, and K. Hamburger, 2013, The influence of perceptual and structural salience: Cooperative Minds: Social Interaction and Group Dynamics. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, p. 3315 - 3320.
- Sael, L., D. La, B. Li, R. Rustamov, and D. Kihara, 2008, Rapid comparison of properties on protein surface.: *Proteins: Structure, function, and bioinformatics*, v. 73(1), p. 1 - 10.
- Saleh, A., M. Khalil, and A.-M. Wahdan, 2003, An Adaptive Image Watermarking Scheme for Color Images Using S-CIELAB, *IEEE 46th Midwest Symposium on Circuits and Systems*, IEEE, p. 1575 - 1578.
- Sanders, A. F., 1970, Some aspects of the selective process in the functional visual field: *Ergonomics*, v. 13(1), p. 101 - 117.
- Schmitt, W., J. L. Sotomayor, A. Telea, C. T. Silva, and J. L. D. Comba, 2015, A 3D Shape Descriptor based on Depth Complexity and Thickness Histograms In *Graphics, Patterns and Images (SIBGRAPI)*, 2015 28th SIBGRAPI Conference on, p. 226 - 233.
- Schultheis, M. T., and A. A. Rizzo, 2001, The Application of Virtual Reality Tecnology in Rehabilitation: *Rehabilitation Psychology*, v. 46(3), p. 296 - 311.
- Shelton, A. L., and T. P. McNamara, 2001, *Systems of Spatial Reference in Human Memory*: Elsevier Science. *Cognitive Psychology* v. 43, p. 274 - 310.
- Shilane, P., P. Min, M. Kazhdan, and T. Furkhouser, 2004, The Princeton Shape Benchmark: IEEE. In *Proceedings of the Shape Modeling International*, p. 167 - 168.
- Shin, M. S., S. Y. Park, S. R. Park, S. H. Seol, and J. S. Kwon, 2006, Clinical and empirical applications of the Rey-Osterrieth complex figure test: *Nature protocols*, v. 1(2), p. 892 - 899.
- Simuladordevuelo.net, 2014, Simuladordevuelo.net.
- Skubic, M., D. Perzanowski, S. Blisard, A. Schultz, and W. Adams, 2002, Spatial Language for Human-Robot Dialogs: *IEEE Transactions on SMC, Part C, Special Issue on Human-Robot Interaction*, p. 1 - 39
- Spotorno, S., B. W. Tatler, and S. Faure, 2013, Semantic consistency versus perceptual salience in visual scenes: Findings from change detection Elsevier. *Acta Psychologica*, v. 142, p. 168 - 176.
- Stoia, L., 2007, Noun phrase generation for situated dialogs, Ohio State University. PhD. Thesis, Ohio, 1 - 179 p.
- Sutherland, I. E., 1965, The ultimate display: *Proceeding of IFIP Congress*, p. 506 - 509.
- Suzuki, M. T., T. Kato, and N. Otsu, 2000, A similarity Retrieval of 3D Polygonal Model Using Rotation Invariant Shape Descriptors, *IEEE*, p. 2946 - 2952.
- Taimouri, V., and J. Hua, 2014, Deformation Similarity Measurement in Quasi-Conformal Shape Space: Elsevier. *Graphical Models*, v. 76, p. 57 - 69.
- Talmy, L., 1983, How language structures space, in Springer, ed., US, p. 225 - 282.
- Tangelder, J. W. H., and R. C. Velkamp, 2008, A survey of content based 3D shape retrieval methods: Springer Science. *Multimedia Tools Applications*, v. 39, p. 441 - 471.
- Tenbrink, T., 2005, Identifying Objects on the Basis of Spatial Contrast: An Empirical Study Springer. *Spatial Cognition IV. Reasoning, Action, Interaction. Lecture Notes in Computer Science* v. 3343, p. 124 - 146.
- Tenbrink, T., and M. Ragni, 2012, Relevance in Spatial Navigation and Communication: Springer. *Spatial Cognition VIII. Lecture Notes in Computer Science* v. 7463, p. 279 - 298.
- Titchener, E. B., 1908, *Lectures on the Elementary Psychology of Feeling and Attention*, New York, The MacMillan Company.
- Tkalčič, M., and J. F. Tasič, 2003, Colour spaces - perceptual, historical and applicational background EUROCON 2003. *Computer as a Tool. The IEEE Region 8 (Volume:1)*, IEEE, p. 304 - 308.
- Trinh, T.-H., 2013, A Constraint-based Approach to Modelling Spatial Semantics of Virtual Environments, Université de Bretagne Occidentale. PhD. Thesis.

- Undurraga, C., and D. Mery, 2011, Improving Tracking Algorithms Using Saliency: Springer-Verlag Berlin Heidelberg. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. In Proceedings 16th Iberoamerican Congress (CIARP), p. 141 - 148.
- Undurraga, C., D. Mery, and S. L. E. Sucar, 2010, Modelo de Saliencia utilizando el descriptor de covarianza.
- Vargas, M. L., and G. Lahera, 2011, "Asignación de relevancia": Una propuesta para el término inglés "saliency", *Actas Esp Psiquiatría, España*, p. 271 - 272.
- Venkatraman, V., S. Lee, and D. Kihara, 2009, Potential for Protein Surface Shape Analysis Using Spherical Harmonics and 3D Zernike Descriptors, <http://link.springer.com/article/10.1007/s12013-009-9051-x/fulltext.html>, Springer, p. 23 - 32.
- Vezhnevets, V., V. Sazonov, and A. Andreeva, 2003, A Survey on Pixel-Based Skin Color Detection Techniques: In Proceeding Graphicon, p. 85 - 92.
- Viscarra, R. R. A., B. Minasny, P. Roudier, and A. B. McBratney, 2006, Colour space models for soil science: Elsevier. *Geoderma*, v. 133, p. 320 - 337.
- Vorwerg, C., 2003, Use of Reference Directions in Spatial Encoding: Spatial Cognition III. Lecture Notes in Computer Science v. 2685, p. 321 - 347.
- Vranić, D. V., and D. Saupe, 2002, Description of 3D-Shape using a Complex Function on the Sphere: In Proceedings of: IEEE International Conference on Multimedia and Expo, p. 177 - 180.
- Westheimer, G., 1975, Visual acuity and hyperacuity: Investigative Ophthalmology & Visual Science, v. 14(8), p. 570 - 572.
- Winograd, T., 1971, Procedures as a representation for data in a computer program for understanding natural language, Massachusetts Institute of Technology. PhD. Thesis, Massachusetts, 1 - 461 p.
- Wraga, M., S. H. Creem, and D. R. Proffitt, 1998, The influence of spatial reference frames on imagined object and viewer rotations Elsevier. *Acta Psychologica*, v. 102, p. 247 - 264.
- Wyszecki, G., and W. S. Stiles, 1982, Color Science: Concepts and Methods, Quantitative Data and Formulae: New York: Wiley, 968 p.
- Yam, K. L., and S. E. Papadakis, 2004, A simple digital imaging method for measuring and analyzing color of food surfaces Elsevier. *Journal of Food Engineering*, v. 61, p. 137 - 142.
- Yuan, J. C.-C., J. D. Brewer, E. A. Monaco, and E. L. Davis, 2007, Defining A Natural Tooth Color Space Based on a 3-Dimensional Shade System: The Journal of Prosthetic Dentistry, v. 98, p. 110 - 119.
- Zhang, D., and G. Lu, 2004, Review of shape representation and description techniques: Elsevier. *Pattern Recognition*, v. 37, p. 1 - 19.
- Zhang, L., Y. Gao, Y. Xia, Q. Dai, and X. Li, 2015, A fine-grained image categorization system by cell-encoded spatial pyramid modeling.: *Industrial Electronics, IEEE Transactions on*, v. 62(1), p. 564 - 571.
- Zhang, L., Y. Han, Y. Yang, M. Song, and S. Yan, 2013, Discovering discriminative graphlets for aerial image categories recognition.: *Image Processing, IEEE Transactions on*, v. 22(12), p. 5071 - 5084.
- Zhang, L., D. S. Joao, Manuel, and A. Ferreira, 2004, Survey on 3D Shape Descriptors, in D. R. POSC/EIA/59938, ed., Republica Portuguesa, p. 1 -28.
- Zhang, L., Y. Yang, Y. Gao, Y. Yu, C. Wang, and X. Li, 2014, A probabilistic associative model for segmenting weakly supervised images.: *Image Processing, IEEE Transactions on*, v. 23(9).
- Zhao, X., and M. Lu, 2013, 3D Object Retrieval Based on PSO-K-Modes Method Academy Publisher. *Journal of Software*, v. 8(4), p. 963 - 970.

Resumen amplio en español

Resumen Amplio en Español

Modelo Computacional para la Generación de Indicaciones en la Localización de Objetos en Entornos Virtuales: Aspectos espaciales y perceptivos

Introducción

Las simulaciones basadas en la realidad virtual están cada vez más presentes en las diferentes áreas de investigación, tales como: el modelado, diseño y planificación, la formación y la educación, teleoperación, la medicina, los tratamientos psicológicos, la rehabilitación y el entrenamiento.

Hoy en día la Tecnología Informática ha permitido recrear, con la ayuda de la Realidad Virtual, más o menos cualquier escena y actividad de la vida real, algunas veces con el objetivo de brindar apoyo a los seres humanos en tareas peligrosas o procesos que son demasiado costosos. La simulación y la realidad virtual han ayudado a los seres humanos a solucionar muchos desafíos del mundo real.

Se reconoce que el ser humano es por naturaleza una entidad activa que lleva a cabo actividades diarias en respuesta a sus necesidades y deseos, tales como: la higiene personal, la alimentación, la movilidad funcional –es decir, moverse de un lugar a otro- localizar un objetos en espacio físico, etc.-.

Asimismo, en el mundo real los seres humanos interactúan con objetos que le pueden ser familiares o no. Las personas a menudo pasan mucho tiempo buscando objetos, los cuales pueden estar perdidos, olvidados, o se encuentran en otra habitación u ocultos con otros objetos. Algunos factores que podrían complicar la ejecución de esta tarea son: el desorden, el que lo objetos hayan sido cambiados de lugar, el que los objetos son desconocidos o poco familiares para las personas, la baja visibilidad de usuario hacia el objetos o la poca probabilidad de recordar la localización de éste.

La tarea de localizar objetos en el mundo real es muy compleja y variable. Esta tarea depende de la posición de los elementos en el momento de la búsqueda, así como del conocimiento previo del medio ambiente y la habilidad de las personas que realizan la tarea.

Una técnica que se utiliza comúnmente en el mundo real, para ayudar en la localización de objetos es usar elementos presentes del entorno donde se lleva a cabo la tarea de localización. Estos elementos se identifican como objetos de referencia, y se pueden utilizar como apoyo para dar indicaciones en la localización de objeto. Por ejemplo, un libro podría ser localizado, usando el estante sobre el cual se coloca como objeto de referencia.

Los objetos de referencia son un recurso efectivo, pero ¿qué hace un objeto un candidato bueno, para usar como objeto de referencia? La selección de un objeto de referencia puede depender de varias variables tales como:

- Las características del objeto (color, tamaño y forma).
- La localización espacial del objeto con respecto al punto de observación (a la derecha, a la izquierda, detrás, etc.)
- El conocimiento previo de los objetos o del entorno donde se encuentran.
- Las características perceptivas y cognitivas del usuario.

En este sentido, los entornos de aprendizaje, así como el entrenamiento virtual, para la industria y las visitas virtuales a museos son algunas de las actividades que requieren la localización de objetos y el intercambio de información mediante la transmisión de mensajes que pueden mejorar el aprendizaje y la construcción del conocimiento.

Para esto, los tutores virtuales son un recurso potencialmente útil, para proporcionar a los usuarios la guía y el apoyo para realizar actividades: estudio de contenidos, operación de equipo industrial, viajes de museos, etc. El desarrollo y uso de estos entornos virtuales abren una gama infinita de posibilidades para el aprendizaje de los usuarios.

En respuesta a esta necesidad, esta tesis propone un mecanismo que permite a un sistema (entorno virtual inteligente), responder a preguntas acerca de dónde un objeto en particular es de cierta manera comprensible para los usuarios. Este mecanismo combina diferentes

elementos como objetos de referencia, saliencia perceptiva, sistemas de referencia espacial, etc.

Objetivos de la investigación

El objetivo de esta tesis es proponer un modelo computacional para la generación automática de indicaciones, para ayudar los usuarios en el desarrollo de tareas que envuelvan la localización de objetos en entornos virtuales. Este modelo considera los diferentes factores que son parte de esta tarea, específicamente aspectos espaciales y perceptivos.

El modelo propuesto proporciona a los usuarios el apoyo, de indicaciones en lenguaje natural, para simular lo que un usuario haría durante la tarea de localizar objetos. La forma de estas indicaciones es compleja, debido a que son altamente variables. La selección de un objeto de referencia se propone como un mecanismo eficaz y rápido en muchas situaciones. Algunas características cognitivas y perceptivas del usuario son también representadas y consideradas con el fin de generar indicaciones que sean adaptables a las necesidades específicas del usuario.

En la búsqueda de este objetivo general, la tesis se dirige a los objetivos prácticos siguientes:

- La definición y el diseño de un modelo computacional de saliencia perceptual de objetos 3D, aplicable a entornos virtuales.
- La definición y el diseño de mecanismos para la extracción de las características de un objeto 3D a fin de medir la saliencia.
- La definición y el diseño de mecanismos, para el modelado de usuario y la adaptación del sistema, de acuerdo al perfil del usuario.
- La definición y el diseño de mecanismos, para seleccionar objetos de referencia, considerando intrínsecamente sus características dependiendo del contexto y del usuario.
- La definición y el diseño de mecanismos para la generación automática de indicaciones en lenguaje natural, para la localización de objetos.

- El desarrollo de un conjunto de herramientas, que reúna e implementa todos los elementos establecidos en el modelo computacional propuesto en esta tesis, para su aplicación a cualquier entorno virtual.

Modelo Computacional de Saliencia Perceptual para objetos 3D en Entornos Virtuales

Este modelo computacional de saliencia perceptual surge de la necesidad de seleccionar el mejor objeto de referencia dentro de un entorno virtual 3D. El objeto se usará dentro de un sistema informático que permite la generación de indicaciones en lenguaje natural, para ayudar a un usuario a localizar objetos específicos dentro del mismo entorno virtual. El modelo computacional de saliencia perceptual se centra en las características visuales de un objeto, como el color, el tamaño y la forma. Las características básicas de un objeto 3D en un entorno virtual se extraen individualmente con diferentes métodos propuesto en esta tesis. Después, la información obtenida se normaliza de acuerdo con los criterios, y los valores se combinan usando una función lineal que ayuda a determinar una medida de la saliencia general para cada objeto 3D en un entorno virtual.

Un objeto puede atraer la atención del observador por sus características físicas o distintivas (Vargas and Lahera, 2011), pero también depende de la interacción de sus características básicas con respecto a otros objetos (Gapp, 1995; Hoffman and Singh, 1997; Spotorno et al., 2013; Stoia, 2007). El modelo de saliencia debe ser capaz de medir la saliencia intrínseca y contextual de cada objetos, al mismo tiempo que permite la identificación del objeto más sobresaliente en un escenario. A continuación se explica la representación de las características del modelo. Después se explica el proceso de implementación del modelo. Por último la evaluación de la aplicación del modelo computacional de saliencia se presenta en la sección 4.5.

Habiendo realizado una revisión de la literatura, nos llevó a la conclusión de que las características con el mayor estímulo para el sistema visual humano son: color, tamaño y forma; esta es la razón de seleccionar estas características, para ser utilizadas en este modelo computacional de saliencia. El orden de la influencia de estas características se estableció de acuerdo con (Gapp, 1995). Además Mangold (1986) demostró en sus estudios experimentales

que el color es más dominante que el tamaño y la forma en las tareas de identificación de objeto. Del mismo modo, el tamaño es más fácilmente reconocido que la forma.

El modelo computacional de saliencia perceptual propuesto se basa principalmente en tres partes: en el primer cálculo de la saliencia individual del objeto, independientemente de su contexto (medio ambiente); segundo el cálculo de la saliencia del objeto con respecto a su contexto, y tercero, el cálculo de una medida de saliencia general (también llamada “saliencia global”), ver Figura 1.

Figura 1 Esquema del modelo computacional de saliencia perceptiva

Modelado de la saliencia individual

En esta sección se explican los procesos de caracterización (también llamados "procesos de extracción de características") de los tres factores (o características básicas del objeto) elegidos para el modelo computacional de saliencia perceptiva.

La caracterización se considera como el proceso de abstracción y cuantificación de las características (color, tamaño y forma) de un objeto 3D en un entorno virtual. La información obtenida de esta caracterización se normaliza en el caso de tamaño, usando un conjunto de funciones con varios criterios. Para el color y la forma de los objetos, otros criterios se tienen en cuenta. Los valores obtenidos en el proceso de normalización o durante el proceso de extracción de características se combinan utilizando una función lineal que nos permite obtener una medida de la saliencia individual en el intervalo $[0,1]$ para cada objeto en 3D en un entorno virtual.

Caracterización del color

Debido a que el modelo RGB se orienta hacia el hardware, no es intuitivo, y no está relacionado con los conceptos de tono, saturación y brillo. Por lo tanto, para que un color en formato RGB sea adecuado para nuestro modelo de saliencia, es necesario recurrir a la transformación de sus valores en un espacio de color uniforme, donde la distancia entre los puntos es directamente proporcional a la diferencia de los colores percibidos. Nuestro espacio de color de destino se conoce como CIE-Lab, que fue diseñado para ser perceptualmente uniforme (Connolly y Fliess, 1997; López et al., 2005). El espacio de color CIE-Lab (estrictamente conocido como CIE 1976 $L^* a^* b^*$) se orienta hacia el ojo humano; que permite cuantificar las diferencias de color que se pueden expresar en términos de la percepción visual humana.

La diferencia de color (ΔE^*) se utiliza a menudo para la evaluación de la calidad de reproducción de color en un sistema de procesamiento de imágenes (Saleh et al., 2003). Esta caracterización se representa en un vector de tres componentes. Para representar el color individual de un objeto dentro de este modelo computacional de saliencia perceptiva se utiliza la variable (C).

Saliencia por color

Con el fin de identificar los colores más destacados a la vista de los usuarios, y medir la saliencia por el color, llevamos a cabo una búsqueda profunda. Sin embargo, no fue posible encontrar en la literatura una forma de cuantificar el grado de saliencia de color, de modo que pudiera dar lugar a una función matemática. En esta búsqueda, lo que encontramos fue una lista de colores que estimulan el sistema visual humano propuesto por (Choungourian, 1968), que decidimos seguir para determinar la saliencia por el color. En este sentido Kröser y Julesz (1989) citaron que la saliencia se puede relacionar, con la detectabilidad del estímulo.

Choungourian (1968) declaró que los colores con los principales estímulos para el sistema visual humano son: rojo, naranja, amarillo, verde-amarillo, verde, azul-verde, azul, y púrpura. Esto fue demostrado en una serie de experimentos y estudios llevados a cabo con estudiantes universitarios de diferentes países, con igual número de hombres y mujeres. Los resultados mostraron que no hubo diferencias en el sexo o la cultura en la preferencia de los colores antes mencionados. La Tabla 1 presenta la lista de estos colores con su representación numérica en los sistemas RGB y CIE-Lab.

Tabla 1 Lista de colores saliente de acuerdo con Choungourian

Color	R	G	B	CIE		
				L*	a*	b*
Red	255	0	0	53.23	80.10	67.22
Orange	255	112	40	64.16	51.02	62.62
Yellow	255	255	0	97.13	-22.55	94.48
Yellow-green	154	205	50	76.53	-37.99	66.58
Green	0	255	0	87.73	-86.18	83.18
Blue-green	13	152	186	58.11	-20.81	-26.94
Blue	0	0	255	32.30	79.19	-107.86
Purple	102	2	153	27.20	59.25	-56.42

Para el modelo computacional el cálculo de la saliencia por color de un objeto, un proceso de tres pasos es seguido:

(1) Las distancias entre el color del objeto con respecto al grupo de los 8 colores más salientes se calcula utilizando las distancias euclidianas.

(2) A continuación, se selecciona el color saliente con menos distancia para el color del objeto, y se le asigna un valor de saliencia a cada uno de sus 3 coordenadas (L, a, b) de acuerdo con los siguientes rangos:

- '1.00' para distancias de menos de 5 unidades
- '0.75' para distancias mayores de 5 y menos de 10 unidades
- '0.50' para distancias mayores de 10 y menos de 15 unidades
- 0.00 " para distancias más grandes que 15 unidades

(3) Finalmente, los tres valores de saliencia se promedian para obtener la saliencia por el color del objeto.

El valor de saliencia por el color del objeto se representa en el modelo computacional por la variable ($Sc_{[0-1]}$).

Caracterización del tamaño

En la medición del volumen de un objeto se puede identificar su tamaño o extensión física en tres dimensiones (es decir, la altura, longitud y anchura), utilizando, por ejemplo, centímetros cúbicos (cm³) o metros cúbicos (m³) como unidad de medida.

Para el cálculo del volumen de un objeto 3D en un entorno virtual, por lo general es fácil de obtener la longitud, anchura y altura del objeto 3D en el sistema de coordenadas. El producto de longitud * anchura * altura nos da una aproximación rápida al volumen; Sin embargo, esta expresión matemática sólo puede calcular con precisión el espacio ocupado por geometrías tales como cubos y prismas rectangulares. Para el cálculo del volumen de objetos geométricos tales como pirámides, cilindros, conos y esferas, entre otros, la fórmula debe variar de acuerdo con las características específicas de cada objeto. Sin embargo, los objetos en el mundo real o en un entorno virtual con frecuencia no son figuras geométricas regulares,

por lo tanto las expresiones matemáticas que se utilizan para calcular el volumen de figuras geométricas regulares no son siempre apropiadas para todos los objetos.

Buscamos una medida para el tamaño que se ajuste a la orientación y la posición del objeto en el espacio en todas sus posibles formas colocado, sin cambiar su punto de referencia fijo. Por lo tanto, proponemos una medida del volumen a través de un proceso de pre-voxelización del objeto, lo que implica tener en cuenta los siguientes tres pasos previos:

- (1) El objeto 3D se voxeliza. El proceso de voxelización es la segmentación de un objeto en pequeñas porciones cúbicas, una unidad llamada voxel, que se ajusta y representa el objeto tridimensional.
- (2) Luego, se obtiene el número total de voxeles (nv). El número de voxeles (nv) de un objeto se obtiene mediante el algoritmo voxelización.
- (3) A continuación, el número de voxeles (nv) se multiplica por el tamaño del voxel (vsi). Este paso obtiene el volumen del objeto (vo) en centímetros cúbicos.

Adicionalmente al volumen, se consideran tres parámetros para caracterizar completamente el tamaño de un objeto 3D: longitud, anchura y altura de su bounding box del objeto.

Saliencia por tamaño

La saliencia por tamaño no se puede medir de forma totalmente independiente para el resto de los objetos que existen en el medio ambiente. Para esto, es necesario determinar, el tamaño mayor de los objeto con el fin de asignar la máxima saliencia (1).

De la lista de todos los objetos en el entorno virtual, el objeto más grande en cada uno de sus parámetros de tamaño (es decir, longitud, anchura, altura y volumen) se utiliza para establecer el valor máximo de saliencia para cada uno de ellos. Los cuatro valores se promedian para obtener la saliencia por el tamaño del objeto 3D, representado en el modelo computacional con la variable ($Ssi_{[0-1]}$).

Caracterización de la forma

Con el fin de encontrar una manera óptima para extraer y representar las propiedades de la forma de un objeto 3D en un entorno virtual, es decir, para obtener la caracterización de la forma, se proponen las siguientes dos formas de representación:

- (1) Una representación vectorial del objeto, usando el descriptor de Zernike. Se utiliza el descriptor 3D Zernike, también conocido como descriptor de momentos geométricos con bases ortogonales.
- (2) Una representación del objeto voxelizado en su cuadro delimitador. Esta representación nos permite determinar la saliencia individual de cada objeto dentro del entorno virtual.

La caracterización de la forma está representada en el modelo computacional con la variable (Sh).

Saliencia por forma

Suponiendo que la superficie plana de un objeto es menos saliente, e inversamente los objetos puntiagudos tienden a ser perceptivamente más relevantes, se propone calcular la proporción de espacio vacío y el espacio lleno en voxeles del bounding box del objeto 3D.

Para esto, se obtiene primero el volumen del bounding box en voxeles (siBBox), y después el volumen en voxeles del objeto (v_o) se resta de este número. El resultado de esta operación es el "espacio vacío" (E_s), que representa el número de voxeles libres dentro del cuadro delimitador que no forman parte del objeto; por el contrario, el volumen del objeto en voxeles corresponde con el "espacio lleno" (F_s) en el bounding box, ver Figura 2.

Figura 2 Espacio vacío y lleno en el bounding box de un objeto 3D

Por último, el espacio vacío (Es) dividido por el total de voxels del bounding box (BBox). Es la proporción de espacio vacío que se utiliza como medida de la saliencia por forma para el modelo.

Calculo de la saliencia individual de objetos 3D

El cálculo de la saliencia individual para cada objeto en el escenario se hace a través de una suma ponderada.

Desafortunadamente, la literatura no proporciona una medida de la importancia relativa de cada factor. Como solución tentativa un peso fue asignado a cada factor de la siguiente manera: 0.5 al color, 0,3 a el tamaño, y 0,2 a la forma.

Modelando de la Saliencia Contextual

La estrategia general de esta etapa consiste en calcular los promedios de color, tamaño y forma en su contexto. Después de eso, se calcula la distancia del color, el tamaño y la forma de cada objeto con respecto a la media del color, tamaño y forma en el contexto. Por último, se describe la manera de medir la saliencia por el color, tamaño y forma de los objetos, usando ciertos criterios determinados.

Promedio del color en el contexto

El enfoque consiste en la obtención de la cantidad de cada color, que el usuario ve en un determinado contexto. Si un color es abundante en la escena, los objetos con ese color puede ser poco saliente, o por el contrario, un objeto con un color que no aparece demasiado pueden ser saliente con respecto al color.

Para simplificar los cálculos, y no tener que tratar con todos los colores por separado, se calcula el color medio del contexto, lo cual es posible porque tenemos una representación del color (CIELab) que permite operar aritméticamente con los colores.

Para obtener el color promedio en un contexto, es necesario calcular la cantidad de cada color en el entorno virtual, desde una perspectiva 2D. Es importante darse cuenta de que las imágenes 2D que las personas ven en un ordenador se pueden representar como una matriz de

píxeles, teniendo cada píxel un color determinado. Con el porcentaje total de cada color individual y la representación matemática de los colores en tres coordenadas bajo el sistema CIELab, el color promedio en el contexto se calcula.

Los resultados de los tres promedios ponderados (L_{avg} , a_{avg} and b_{avg}) representa el color promedio en el contexto. Para representar el color promedio en el contexto utilizamos la variable ($cctx$).

Distancia en color con respecto al contexto

Después, el cálculo de la distancia del color (dc), entre el color promedio del contexto y el color individual de todos los objetos de contenido en nuestro entorno virtual se calcula. Una vez que hemos obtenido las coordenadas del color promedio del contexto, la saliencia del color de un objeto en el entorno virtual con respecto al contexto se calcula con la distancia euclidiana entre los dos colores.

Saliencia del color con respecto al contexto

Para determinar matemáticamente el valor de la saliencia por color, tamaño y forma de todos los objetos del entorno virtual con respecto al contexto, se propone un proceso que se basa en las siguientes operaciones:

- Distribución estandar: Suponiendo que las distancias en el color (tamaño y forma) con respecto al contexto tienen una distribución normal, vamos a transformar estas distancias en sus valores correspondientes en una distribución normal estándar.
- El uso de una distribución proporcional para determinar el valor de saliencia normalizado. El objeto con el valor de la distancia estándar más alto tendrá el valor de saliencia '1', el objeto con el valor de la distancia estándar más bajo tendrá el valor de saliencia '0', y el resto será ponderado en proporción a sus valores.

El valor de saliencia en color con respecto al contexto se representa con las variables ($ScC_{ss[-3 - 3]}$) cuando la saliencia se considera de la puntuación típica, y ($SCC_{[0-1]}$) cuando la saliencia se considera desde la distribución proporcional.

Este procedimiento se utiliza para determinar la saliencia de un objeto con respecto al contexto, por sus tres características físicas (color, tamaño, forma).

Promedio del tamaño en el contexto

La distancia en tamaño (*dsi*), entre el tamaño promedio del contexto y el tamaño individual (en porcentaje) de todos los objetos de contenido en nuestro entorno virtual se calcula.

Distancia en tamaño con respecto al contexto

Una vez que hemos obtenido el tamaño promedio del contexto, la saliencia por tamaño de un objeto con respecto al contexto se calcula como la distancia euclidiana entre dos tamaños. La variable que se utiliza para representar este valor es (*dsi*).

Saliencia del tamaño con respecto al contexto

El valor de saliencia del tamaño de un objeto con respecto al contexto se calcula con la transformación en puntuaciones típicas.

Promedio de la forma en el contexto

En este modelo de saliencia contextual decidimos tratar la forma de un objeto teniendo en cuenta sus tres dimensiones, a diferencia de las características de color y tamaño, que fueron tratados desde una perspectiva de dos dimensiones. En este sentido, la teoría de la Gestalt menciona que la mente humana transforma y organiza los elementos percibidos en algo para ser integrados en un todo. Esta teoría también se extiende a la organización de las formas y sus significados (Pinna, 2010).

La forma de un objeto es una característica básica que permite identificar y distinguir un objeto de otro, desde una perspectiva individual y en un contexto. La forma de un objeto nos puede dar información sobre su función y uso. Hay objetos como una caja o un coche cuyas propiedades funcionales están claramente relacionadas con su forma y estructura perceptible.

Por otro lado, la forma de los objetos naturales como árboles, flores, rocas, etc., cambian ligeramente dentro de sus entornos, por lo menos aquellos que tienden a crecer, cambiar de

posición o desgastarse físicamente. En resumen, la forma de un objeto proporciona información importante para reconocer objetos en futuras ocasiones o para determinar si un objeto es similar a otros dentro de un entorno real o virtual.

Para calcular el promedio de la forma en el contexto, el vector de la forma del contexto se obtiene con el siguiente método que incluye tres pasos:

- (1) Los objetos del entorno virtual se voxelizan.
- (2) El descriptor 3D Zernike se aplica a todos los objetos voxelizados con un orden 8, para generar un vector de 25 valores numéricos.
- (3) A continuación, se añaden los valores de todos los vectores y se promediaron en cada uno de sus componentes. El resultado final de este proceso es el vector del contexto (V_{cont}) con 25 valores.

El promedio de la forma del contexto se caracteriza con el vector del contexto y está representada en este modelo con la variable ($shctx$).

Distancia en forma con respecto al contexto

Para obtener la distancia en forma utilizamos la distancia euclídea. Esta distancia se calcula entre los componentes de los vectores del descriptor de Zernike obtenidos de cada objeto 3D, y los componentes del vector del contexto. La variable que se utiliza para representar este valor es (dsh).

Saliencia de la forma con respecto al contexto

El valor de saliencia de la forma de un objeto con respecto al contexto se calcula con la transformación en puntuaciones típicas.

Cálculo de la saliencia contextual de un objeto 3D

El cálculo de la saliencia con respecto al contexto de cada objeto en el escenario se hace también a través de una suma ponderada. Se decidió establecer un peso de 0,5 para el color, el tamaño de 0,3 y 0,2 para la forma.

Saliencia General

La saliencia general (también llamada saliencia global) de un objeto, en este modelo computacional de saliencia, se calcula como la suma de un tercio del valor obtenido en la saliencia individual más dos tercios del valor en la saliencia con respecto al contexto.

Modelado de usuario y semántica de mundos virtuales

Uno de los pilares fundamentales para la generación de indicaciones es que se adapte a las necesidades y las características del usuario. Así como al estado actual del mundo virtual, la capacidad de construir modelos explícitos del usuario en el mundo virtual, y de los conocimientos y la experiencia que el usuario tiene de los diferentes objetos en el mundo virtual. Estos modelos estarán representados en dos ontologías de datos.

Semántica de mundos virtuales

De acuerdo con (Pittarello y De Faveri, 2006), una parte importante de la descripción de la semántica de un entorno virtual se basa en el concepto de objeto. Hay dos tipos de objetos pueden ser definidos: los objetos geométricos (GO) y los objetos semánticos (SO). Uno o más objetos geométricos se pueden asociar a un objeto semántico, una entidad que tiene un significado específico compartida por el autor y el usuario final del medio ambiente. Objetos semánticos pueden ser organizados en una estructura jerárquica de varios niveles.

Bajo este enfoque teórico, y el uso de una ontología, se desarrolla una descripción semántica de los mundos virtuales. La ontología mundo es, sin duda, de gran relevancia para el sistema. Esta ontología almacena los datos de los objetos de la escena (también llamado entorno virtual), en el que la actividad de localización de objetos se lleva a cabo.

El diseño de las "ontologías de mundo y usuario" se realizó siguiendo el esquema propuesto por (González, 2014), que incluye tres etapas:

1. Identificación de las características de un objeto / usuario dentro de un entorno virtual.
2. Creación del modelo conceptual de la ontología mundo / usuario.
3. Aplicación de la ontología mundo / usuario con la herramienta Protégé.

En esta ontología se representa la semántica del entorno virtual que es necesario para la generación de indicaciones, en la que las propiedades y las relaciones entre los objetos en el entorno virtual son almacenadas.

La figura 3 presenta un mapa conceptual de la ontología mundo.

Figure 3 Mapa conceptual de la ontología de mundo

Modelado de usuario

El modelo computacional para la generación de indicaciones para la localización de los objetos que se propone en esta tesis, intenta ser tan adaptable como sea posible al modelo del usuario y al entorno virtual. El modelo de usuario que se presenta en esta tesis pretende representar al usuario de una manera computacional con una ontología denominada "la ontología de usuario".

En la ontología de usuario se describe en detalle una propuesta de modelado de usuario. Esta propuesta permite que represente atributos cognitivos y perceptivos (también llamados características) de los usuarios que interactúan con el entorno virtual en el que se simula el proceso de localización de objetos. El atributo cognitivo seleccionado es la capacidad de recordar la localización de objetos (la memoria espacial), que se utiliza para deducir la

probabilidad de recordar la localización de un objeto específico. Los atributos perceptuales representan características del sistema visual del usuario, tales como la agudeza visual, o el tipo de daltonismo. Del mismo modo, la información sobre los conocimientos previos de los usuarios en ciertos dominios específicos, y el conocimiento que tienen de los objetos en el entorno virtual. Esta información también se representa en la ontología. La figura 4 presenta un mapa conceptual de la ontología de usuario.

Figure 4 Mapa conceptual de la ontología de usuario

Para facilitar la definición semántica de los entornos virtuales y de los usuarios, un editor-exportador semántico fue diseñado e implementado, para ser utilizado en una etapa de pre-procesamiento.

Generación automática de indicaciones para localizar objeto en EVs

El objetivo principal de esta tesis es diseñar un componente de software que de una manera inteligente genere instrucciones en lenguaje natural para que un usuario pueda localizar objetos en un entorno virtual.

Para el diseño de este componente de software fueron analizadas tres situaciones:

- (A) El objeto a ser localizado (OL) está en la misma habitación que el usuario, pero no en el campo de visión del usuario.
- (B) El OL está en otra habitación, y por lo tanto no a la vista del usuario.
- (C) El OL se encuentra en el campo de visión del usuario.

Una serie de sub-casos que se originan a partir del tercer caso fueron identificados, véase la Figura 5.

Figura 5 Casos considerados para la generación automática de indicaciones

La solución diseñada se basa en tres algoritmos auxiliares:

- Un algoritmo para la selección del mejor objeto de referencia (SBRO), el cual se basa en la saliencia perceptual del objeto, el conocimiento previo del usuario acerca de los tipos de objetos en la escena, y la probabilidad que el usuario tiene de recordar la posición de los

objetos. El algoritmo SBRO comprende una serie de representaciones matemáticas consideradas para la implementación de algunas características perceptivas y cognitivas del usuario, que se relacionan con el conocimiento acerca de los objetos del entorno virtual.

- Un algoritmo para la selección de una relación espacial relevante (SSR) entre el objeto a ser localizado (OL) y el objeto de referencia (RO). El algoritmo de la SSR incluye los criterios computacionales para la selección de las 12 relaciones espaciales utilizadas en la generación de indicaciones: sobre / bajo, cerca, izquierda, derecha, arriba, abajo, delante, detrás, dentro, entre, primero, segundo, o tercero.

- Un algoritmo para la generación de las instrucciones (GD) que tiene en cuenta los casos o situaciones que se tratan. El algoritmo GD genera expresiones en forma de frases descriptivas para la localización de objetos. Cada frase incluye la información necesaria para una correcta comprensión como: objeto a localizar, relación espacial, objeto de referencia, y las propiedades del objeto, de acuerdo a cada caso. La estructura de una frase generada por nuestro sistema se basa en la estructura lingüística que presenta Kelleher (2003), ver Figura 6:

Figura 6 Esquema de una sentencia final

Donde

OL	es el objeto a ser localizado
RO	es el objeto de referencia

La representación de la estructura sintáctica de la frase aparece en la Figura 7.

Figura 7 Estructura sintáctica de una sentencia

- El sintagma nominal se refiere al objeto a ser localizado.
- El sintagma verbal considera la naturaleza del problema, es decir, la necesidad de localizar un objeto. Este sintagma verbal siempre será la cadena de caracteres "es", que indica la situación de posición.
- El sintagma adverbial contiene tanto la relación espacial y el objeto de referencia.

Conclusiones

- ✓ Se propone una métrica que mide la saliencia de la forma de objetos 3D, de manera simple, rápida y efectiva. Esta métrica se puede aplicar a todo tipo de objetos regulares e irregulares. Además es adaptable a todas las escalas de los modelos 3D.
- ✓ Se propone un modelo computacional de saliencia perceptiva basado en el color, tamaño y forma de los objetos 3D.
- ✓ El modelo se basa en un conjunto de etapas:
 - (1) Caracterizar el color, tamaño y forma de cada objeto.
 - (2) Aplicar un proceso, que permite establecer un valor de saliencia individual.
 - (3) Calcular la saliencia respecto al contexto de cada objeto en el entorno virtual.
 - (4) Medir la saliencia general de todos los objetos del entorno.
- ✓ Una representación semántica de los entornos virtuales que permite representar de forma explícita las propiedades y las relaciones entre objetos en un entorno virtual utilizando ontologías de datos.
- ✓ Un modelo de usuario que puede representar características perceptivas y cognitivas de un usuario, la cuales son relevantes en el proceso de localización de objetos.
- ✓ Un método para la generación de indicaciones compuesto por:
 - (1) Un algoritmo para la selección de los mejores objetos de referencia (SBRO).

- (2) Un algoritmo para la selección de una relación espacial relevante (SSR) entre el objeto a localizar (OL) y el objeto de referencia (RO).
- (3) Un algoritmo para la generación de indicaciones (GD), que se distingue entre varios casos posibles en cuanto a la posición, visibilidad y propiedades perceptuales del objeto a localizar con respecto al usuario.
- ✓ Nuestro modelo para la generación de indicaciones se caracteriza por representar un Sistema Inteligente por ajustarse al perfil y la experiencia previa de cada uno de los usuarios en el proceso de localización de objetos.

Trabajo Futuro

- ✓ Considerar las texturas de los objetos al momento de determinar la saliencia por color.
- ✓ Tratar en nuestra métrica la variabilidad de apariencias que la forma que un objeto podría tener debido a los cambios de perspectiva y la influencia del significado emocional.
- ✓ Hacer una comparación sistemática de la métrica de saliencia por forma de los objetos 3D, con respecto a otros indicadores relacionados que se encuentran en la literatura.
- ✓ Extender el modelo de saliencia, incluyendo factores adicionales como: la utilidad de los objetos, su significado emocional, su movilidad, etc.
- ✓ Validar la generación de indicaciones experimentalmente en términos de su utilidad y naturalidad para los usuarios.
- ✓ Integrar este modelo en un Agente Pedagógico que requiera guiar al usuario en un entorno virtual.