



# **Object Categorization using Biological Models**

Ana Milene Mestre Vieira

Nr. 30367

Dissertation to obtain the

Master Degree in Electrical and Electronics Engineering

Specialization Area: Information and Telecommunication Technologies

Supervisor

João Miguel Fernandes Rodrigues

2013

# **Object Categorization using Biological Models**

Ana Milene Mestre Vieira

Nr. 30367

Dissertation to obtain the

Master Degree in Electrical and Electronics Engineering

Specialization Area: Information and Telecommunication Technologies

Supervisor

João Miguel Fernandes Rodrigues

2013

## ***Object Categorization using Biological Models***

Categorização de Objetos usando Modelos de Inspiração Biológica

---

### **Declaração de autoria de trabalho**

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Ana Milene Mestre Vieira

### **Copyright by Ana Vieira**

A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

*À minha família e amigos.*

# Abstract

Humans are natural at categorizing objects, i.e., at dividing them into groups depending on their features and surroundings. We do it easily and in real-time. Additionally, our Human Visual System (HVS) is the only one reliable for object detection, categorization and recognition; the latter events take place in the visual cortex, being object recognition achieved around 150-200ms, and occurring also a categorization-specific activation in prefrontal cortex before or around 100ms. This provides one of the evidences which substantiate that categorization is a more bottom-up process than recognition. Visual cortical area V1 is composed - among others - by simple and complex cells which are adjusted to different spatial frequencies (scales), orientations and disparity. These cell's responses were used to build a model for events detection in V1; these events are classified by type - lines and edges - and polarity - positive and negative. Being the goal of this thesis to develop a cortical model for object categorization - inspired in the HVS and based on 2D object views -, the V1 multi-scale events generated by the former model were used to accomplish that goal.

In the developed categorization model the final category attributed to an object is the convergence of three similarity concepts which define in different ways the resemblance degree between an object and a certain category; the resemblance degree is therefore accomplished by comparing the V1 events between templates and objects. The resemblance degree or similarity percentage was calculated (a) on the first concept as the quotient between the number of common events between object and category templates (considering type and polarity) in all scales, and the number of object's events in all scales; (b) on the second concept the similarity percentage was calculated as the quotient between the number of common events between object and category templates (not considering type nor polarity) in all scales, and the number of object's events in all scales; (c) finally, on the third concept this ratio was calculated as the quotient between the number of common events between object and category templates (considering type and polarity) in all scales, and the category's "events number" in all scales. The final category assigned to an object is then (1<sup>st</sup>) a category on which the three concepts agree on and (2<sup>nd</sup>) the best scored one.

For the proof of concept a database composed by 8 different categories and 10 objects per category was used; left and right profile views were chosen to represent each object. Regarding the 80 results obtained by categorizing 40 objects in both views, an average categorization success rate of 93.75% was accomplished, being 92.50% the success rate achieved for left profile, and 95.00% the one achieved for right profile; even each of the miscategorized images was attributed a category which is similar to its true one. In order to conclude the proof of concept, the model was also tested in terms of small invariance to rotation, scale and noise, having been then achieved high categorization success rates (above 82%).

**Keywords:** Categorization, Lines and Edges, Multi-scale, Visual Cortex.

# Resumo

É inata a nossa capacidade de categorizar objetos, i.e., de dividi-los em grupos dependendo das suas características e contexto em que estes surgem. Fazêmo-lo facilmente, e em tempo real. O nosso Sistema Visual Humano (SVH) é o único considerado fiável para deteção, categorização, e reconhecimento de objetos, e, embora o seu funcionamento não seja inteiramente conhecido, já há muita informação acerca do mesmo.

Os nossos olhos captam a informação exterior sob a forma de luz; esta, sendo projetada na retina que é fotossensível, é convertida em sinais elétricos que são encaminhados para o cérebro através do nervo ótico; daí os sinais seguem para o *Lateral Geniculate Nucleus* (LGN) que por sua vez os encaminha para o córtex visual onde a informação é processada. É no córtex visual que ocorre a deteção, categorização e reconhecimento de objetos; estes processos envolvem dois fluxos de dados - “*bottom-up*” e “*top-down*” (subsistemas “*what*” e “*where*”) – incluindo também a integração de ambos. Há evidência de que reconhecemos um objeto após uma exposição de apenas 150-200ms, e também, logo após 100ms (ou menos) há uma ativação específica da categorização que ocorre no córtex pré-frontal; isto é uma das evidências que fundamentam que a categorização é um processo mais “*bottom-up*” do que o reconhecimento. Basta um olhar muito breve sobre uma imagem para sabermos os seus principais conteúdos (*gist*); isto sugere que alguma informação é transmitida muito rapidamente ao córtex pré-frontal, para que os mais prováveis *templates* de objetos e vistas possam ser selecionados e encaminhados pelos subsistemas “*what*” e “*where*”; o fenómeno descrito anteriormente corresponde à categorização de objetos – o tema em estudo nesta dissertação.

A área cortical V1, uma das áreas do córtex visual, é composta (entre outras) por células simples e complexas, as quais estão ajustadas a diferentes frequências espaciais (escalas), orientações e disparidades. As respostas destas células foram usadas para construir um modelo de deteção de eventos em V1, sendo que estes podem ser separados por tipo – linhas e arestas -, e polaridade - positiva e negativa; estes atributos corticais caracterizam cada objeto, pelo que cada categoria de objetos tem uma distribuição espacial muito própria destes eventos, mais oticamente visível nas escalas mais baixas (escalas finas). Sendo o objetivo desta tese desenvolver um modelo cortical para categorização de objectos – inspirado no SVH e baseado em vistas 2D – as linhas e arestas em multi-escala geradas pelo modelo de deteção de eventos em V1 foram usadas nesse sentido. Existem já alguns modelos que propõem arquiteturas corticais para categorização de objetos, no sentido de se aproximarem progressivamente da forma de funcionamento do SVH; porém, este por sua vez é complexo e não inteiramente conhecido, havendo muitas hipóteses a considerar e simular.

No modelo de categorização desenvolvido, a categoria final atribuída a um objeto consiste na convergência (multi-escala) de 3 conceitos de semelhança que definem de diferentes maneiras o grau de parença entre um objecto e uma dada categoria; o grau de parença é então obtido através da comparação entre os eventos em V1 dos objetos/categorias padrão (*templates*). O grau de parença ou percentagem de semelhança entre um objecto e uma categoria foi calculado (a) no primeiro conceito como o quociente entre o número de eventos comuns entre o objeto e os *templates* da categoria (considerando o tipo e a polaridade) em todas as escalas, e o número de eventos do objeto em todas as escalas; (b) no segundo conceito a percentagem de semelhança é calculada como o quociente entre o número de eventos comuns entre o objeto e os *templates* da categoria (não considerando o tipo nem a polaridade) em todas as escalas, e o número de eventos do objeto em todas as escalas; (c) finalmente, no terceiro conceito a percentagem de semelhança é calculada como o quociente entre o número de eventos comuns entre o objeto e os *templates* da categoria (considerando o tipo e a polaridade) em todas as escalas, e o ‘número de eventos’ de uma categoria em todas as escalas.

Para a prova de conceito utilizou-se uma base de dados constituída por 8 categorias diferentes (maçãs, carros, vacas, chávenas, cães, cavalos, pêras e tomates), cada uma com 10 objetos, sendo cada objeto disponibilizado em 41 vistas diferentes; para representar um objecto, foram seleccionadas duas vistas 2D – perfil esquerdo e direito. As comparações efectuadas entre uma imagem de um objeto e os *templates* das categorias (na mesma vista) geram, considerando os três conceitos, um total de  $3 \times 8 = 24$  rácios de semelhança; mediante a definição de uma margem em torno da percentagem máxima obtida em cada conceito, cada um dos 3 fornecerá a(s) sua(s) categoria(s)-resposta. Intersectando as 3 respostas, ‘a’ ou ‘as’ categorias comuns são seleccionadas; se houver uma só categoria comum, essa é a resposta final; se houver mais do que uma, é escolhida a categoria mais pontuada.

Da base de dados usada que tem um total de 80 objectos, foram seleccionados 40 com o propósito de criar os *templates*, sendo os restantes objetos reservados para realizar os testes e completar a prova de conceito – objetos teste. Assim, para cada vista, há 40 imagens disponíveis dos objectos *template* (80 nas duas vistas); as imagens dos objectos *template* de uma mesma categoria são usadas para preparar os *templates* da categoria. Os *templates* de comparação de uma categoria para uma dada vista e escala dividem-se em 4 imagens correspondentes aos eventos disponíveis: uma imagem representante das linhas positivas, outra representante das linhas negativas, outra das arestas negativas, e outra das arestas positivas.

Baseando-se o modelo na concordância entre conceitos distintos, este modelo apresentou resultados ‘primários’ muito promissores: considerando os 80 *templates*, i.e., 40 objetos *template* em ambas as vistas, 100% dos resultados de categorização estavam corretos (como esperado). No teste seguinte, usou-se os restantes objetos disponíveis – os 40 objetos teste – com o objetivo de observar se o modelo

era bem-sucedido na categorização de objetos ‘desconhecidos’ (pois no teste anterior os objetos testados eram ‘conhecidos’ uma vez que os *templates* das categorias são construídos com base nas imagens dos objetos *template*). Assim, considerando os 80 resultados obtidos (ambas as vistas) obteve-se uma percentagem média de sucesso de 93.75% (75 imagens), sendo que das imagens correspondentes aos objetos em perfil esquerdo 92.50% foram corretamente categorizadas, e das correspondentes ao perfil direito esta percentagem correspondeu a 95.00%. A grande maioria das imagens de ambas as vistas foi assim bem categorizada, e, até a cada uma das restantes 5 imagens (3 objectos) foi atribuída uma categoria similar à sua categoria original.

Finalmente, para terminar a prova de conceito, o modelo foi testado em termos de ‘pequena’ invariância à rotação, escala e ruído; estes testes foram realizados tendo como base as imagens dos objetos teste em perfil direito, e, foram então obtidas elevadas percentagens de categorização (acima de 82%).

**Palavras-chave:** Categorização, Linhas e Arestas, Multi-escala, Córtex Visual.



# Contents

Abstract .....	V
Resumo .....	VI
Contents.....	IX
1. Introduction .....	1
1.1 - Thesis goals, contributions and context .....	5
1.2 – Thesis overview.....	5
2. General Concepts and State-of-the-Art .....	6
2.1 - Categorization approaches gist and non-biological approaches description .....	8
2.2 - Choosing the HVS as an inspiration .....	12
2.3 - General concepts about the Human Visual System .....	15
2.4 - Cortical attributes: LEs detection and classification.....	19
2.5 - Biologically inspired categorization models.....	23
3. Categorization Model.....	27
3.1 – The categorization model: an introduction.....	27
3.2 - Obtaining each category’s template set .....	33
3.3 - Calculating the similarity percentages between an object and a category .....	34
3.4 - Final categorization model.....	41
4. Tests and Results.....	50
4.1 – Comparing the categorization model’s results to other models’ results .....	50
4.2 - Evaluating the categorization model in terms of invariance.....	53
5. Conclusions and Future Work.....	58
6. Appendices .....	62
Appendix 6.1 – ETH-80 database: some samples, and ‘segregated’ template objects .....	62
Appendix 6.2 - Examples of the lines and edges from some template objects .....	65
Appendix 6.3 - Example of the animals’ templates sets .....	68

Appendix 6.4 - Detailed categorization results .....	71
Bibliography.....	84

# 1. Introduction

---

## ABSTRACT

**Objects biological categorization - the main subject of the thesis - is introduced in this chapter. Here the thesis' goals, contributions and context are presented, and a brief glance over the thesis' topics is also provided.**

---

Computer Vision (CV) is a challenging and useful technology area. Its origin dates back to the late 1950s and early 1960s (Bebis et al., 2003) – just about 60 years ago; nevertheless, it is quickly spreading throughout the world, being a powerful tools source for several areas, such as medicine, robotics, manufacturing, remote sensing, industry and multimedia (Bebis et al., 2003). CV is “the study of enabling computers to understand and interpret visual information from static images and video sequences” (Bebis et al., 2003), being a study subject which is becoming very important in computer science and engineering. Bebis et al., 2003 referred that CV research aims to understand the processes underlying vision, in order to computationally implement them. The authors refer two motivations to back up this statement: 1) “to develop computer vision systems as a method of testing and evaluating models of human or other biological vision systems” and 2) to solve concrete problems by means of computer systems conceived through engineering approaches. Furthermore, some of the CV’s sub-domains comprise object detection, retrieval, categorization and recognition, face detection and recognition, scene categorization and rebuilding, events detection, video tracking, optical flow, learning, indexing, image restoring and movement prediction.

Among the above mentioned CV’s sub-domains, one of them is especially important for this thesis - object categorization - as the focus was to build a biologically based categorization model. Visual object categorization – also termed ‘generic object recognition’ – can be defined “as the process of assigning a specific object to a certain category” (Pinz, 2006). Savarese & Fei-Fei, 2010 and Kinnunen et al., 2010 mention in different words that object categorization has been a very significant CV research sub-area in recent years. There are several applications for categorization. For instance, Kim & Medioni, 2011 mention that

applications for shape-based object classification “include autonomous robotic navigation and manipulation, and urban scene understanding,” and Pinz, 2006 provided some examples of categorization application fields, such as “image database annotation, image retrieval and video annotation”. Moreover, the latter author states that “reliable categorization in real-time will open up applications in surveillance, driver assistance, autonomous robots, interactive games, virtual and augmented reality and telecommunications.” Considering a broader viewpoint, these applications might include ‘cognitive personal assistance’ systems “with many potential aspects, ranging from user support in complex environments to very basic support capabilities for elderly or disabled people” (Pinz, 2006). As could be seen, there are many applications for artificial categorization systems; however, several obstacles need to be surpassed so that these systems function properly.

An artificial categorization system needs information about an object to recognize its class. Galleguillos & Belongie, 2010 mentioned that object categorization traditional approaches use appearance features as the principal data to categorize objects in real world images, naming as well some appearance features: color, edge responses, texture and shape. Pinz, 2006 referred that some ‘recent’ approaches try to (a) “model appearance more locally”, (b) “group simple geometric primitives” and (c) “use learning algorithms to find common patterns that can be shared over many individuals of a category”. Besides the already mentioned categorization-related publications, several others enriched the CV state-of-the-art; as an example, we can name the one by Petre & Zaharia, 2011 who presented a scheme for categorization of 2D objects present in still images, which resorted to categorized 3D models; also, Leibe & Schiele, 2003 studied the efficiency of some appearance- and contour-based recognition methods for multi-class object categorization, and Han et al., 2011 assessed a canonical correlation- and view-based approach for object categorization and recognition. Further categorization methods will be mentioned in chapter 2 (General Concepts and State-of-the-Art), being some of them described in detail.

Not only object categorization, but also object detection and recognition are important CV topics, once there's a wide range of commercial applications in surveillance and robotics. Rodrigues & du Buf, 2009b refer that “categorization is the last step before recognition in which attention shifts to finer scales that reflect minute differences”. Object recognition is a classical problem which is addressed in any CV, image processing and machine vision book (Rodrigues, 2007). Among the existent object recognition algorithms, there are those more

object generalized and others more object specific; the latter - named “dedicated” algorithms - are tuned to a specific set of objects (Saleiro, 2011); the more object generalized algorithms – which work in a broader case range – have, usually a higher computational cost (Evans, 2009; Lowe, 1999). As referred by Rodrigues, 2007, each author applies his/her own definition of ‘recognition’; “for instance, we can refer to recognition as to recognize one or several pre-specified or learned objects or object classes (e.g. this is a coffee mug), it may be the identification of a specific object (this is Paul’s coffee mug), or even detection (there are two coffee mugs in this image).” In this thesis we mostly refer to ‘object recognition’ as the identification of a particular object, and to ‘object categorization’ as the recognition of an object’s class; however, as with every word, the usage of the term ‘recognition’ will depend on the surrounding context (for a small state-of-the-art about recognition see chapter 2).

Rodrigues, 2007 referred to categorization as being a tougher task than recognition, and described an elucidating example that distinguishes both tasks. “Object recognition is a clearly defined task: a certain cat, like the neighbors’ red tabby called Toby, is recognized or not” (Rodrigues, 2007); the same author refers that for categorization there are different levels to consider “before deciding between our own red tabby called Tom and his brother Toby living next door”, for instance: “(a) an animal, (b) one with four legs, (c) a cat, and (d) a red tabby” (Rodrigues, 2007). Regarding recognition, categorization can be perceived as a more challenging CV topic, or at least very challenging in a different way: an object may be studied in different views, lightening conditions, occlusions, and contexts in order to be posteriorly recognized; yet, in order to study a category, several different objects from the latter may be studied in all the previously mentioned variations, in the pursuit of a common pattern that represents them all; plus, considering that a certain number of categories is considered, each category’s pattern has to be sufficiently descriptive to contrast from the remaining categories’: a particular middle term has to be achieved in order to build a categorization model. Accordingly, Pinz, 2006 refers that “major problems are related to the concept of a ‘visual category’, where a successful recognition algorithm has to manage large intra-class variabilities versus sometimes marginal inter-class differences.” The latter author mentioned some issues an artificial categorization system has to deal with, such as representation, recognition and learning. Pinz, 2006 also pointed out an object location issue, mentioning that many systems are capable of categorizing images, but are unable to localize and delineate an object in an image; the latter author referred as well to a database issue - “it turns out that often background (context) is learned rather than object-specific information” – a problem

that we will allude to later in chapter 3 (Categorization Model). Furthermore, Pinz, 2006 referred to an evaluation issue, i.e., ‘how should a categorization system’s output be evaluated?’, and a system integration issue: “*system integration* can get quite complex when many components are required to interact smoothly” (Pinz, 2006).

Looking at categorization from a biological perspective, it appears that humans perform categorization far better than artificial systems. However, recognition “can often be handled more efficiently, reliably or simply faster by an artificial vision system” (Pinz, 2006). The last statement is consistent with the fact that more research had taken place in the recognition area than in the categorization field “in the past” (Pinz, 2006). According to VanRullen & Thorpe, 2001 humans are capable of performing ultra-rapid categorization tasks; they can decide if a quickly flashed image corresponds to a certain category in less than 150ms (Pinz, 2006). Moreover, cognitive psychology evidence reveals that humans handle with approximately 30.000 different categories – says Pinz, 2006 while referring to Biederman, 1995; Pinz, 2006 also adds: “this would require solving currently intractable computational complexity”. The “abstraction level of object classes” is not clearly defined (Leibe & Schiele, 2003); referring to Brown, 1958, Leibe & Schiele, 2003 indicate that “the question of how humans organize knowledge at different levels has received much attention in Cognitive Psychology”; “it is important to note that categories do not exist per se in the world” (Leibe & Schiele, 2003); the latter authors, referring to Rosch et al., 1976, state that categories are a “learned representation” which depends on education and experience (Leibe & Schiele, 2003).

Many have been the motives that sort categorization as a hard task; we also denoted the superiority of the Human Visual System (HVS) when compared to a categorization artificial system. Naturally and effortlessly, we distinguish various categories in real-time; however, the HVS that we use on the daily basis is not entirely known. Being a unique system and the most reliable for object categorization, the HVS is used as reference in some studies, as for example the one by Mutch & Lowe, 2008 where “the role of sparsity and localized features in a biologically-inspired model of visual object classification” was studied; other examples are the features clustering method by Mundhenk et al., 2004, and the V1 lines and edges detection model by Rodrigues & du Buf, 2009b.

## **1.1 - Thesis goals, contributions and context**

As referred before, this thesis work is focused on object categorization – a challenging CV field; also, since the HVS is the best at categorizing objects, doing it easily in real-time, this thesis' goal is to develop a biological categorization model that allows (a) creating a biologically based descriptor that represents an object, (b) separating objects in categories and sub-categories, and (c) categorizing objects using their 2D views.

The proposed categorization model uses the V1 lines and edges detector from Rodrigues & du Buf, 2009b. The main contribution to highlight in this thesis was the principle created for the model, as the category assigned to an unknown object is achieved through a consensus among different categorization concepts; each concept makes class suggestions for a certain object, being the final assigned category obtained by intersecting all concepts class propositions and, if the intersection is composed by several categories, the best scored one is then chosen through a proper process. This model is thus based on a mutual agreement principle for the purpose of increasing the certainty degree of the given final answer.

This thesis took place in Vision Laboratory –UALG, being also integrated in a project named “A neuro-dynamic framework for cognitive robotics: scene representations, behavioural sequences, and learning”. This project was funded by EC, CINTAL/UALg (Vision Laboratory), FP7-ICT-2009-6 PN: 270247 (2011-2015).

## **1.2 – Thesis overview**

As a starting point, this chapter presented an introduction to the thesis' subject, goals, contributions and context. Next, chapter 2 presents the general concepts and the state-of-the-art in the objects categorization field, providing also a brief glance over the HVS and over the lines and edges cortical attributes used in the model. Chapter 3 describes the development of the categorization model, explaining the way every step was designed; next, chapter 4 reveals the tests and results and chapter 5 completes the thesis main work by presenting some final conclusions and some future work ideas. In chapter 6 the Appendices are presented, consisting on images and tables which support chapters 3, 4 and 5, namely images related to the studied objects, and tables which present the categorization numerical results. Finally, after the Appendices, the Bibliography is presented.

# 2. General Concepts and State-of-the-Art

---

## ABSTRACT

Several models concerning object categorization are presented in this chapter – biological and non-biological -, with special focus on the ones which used the ETH-80 database to achieve their experimental results, as these models are more suited to be compared to the model developed in this thesis. In addition, since the latter model is bio-inspired and a cortical attributes detection model was used as basis for the categorization model developed in this thesis, a description over the HVS and over the cortical attributes detection model was given; furthermore, some motivations regarding the use of the HVS as inspiration were covered.

---

Computer Vision is spread among several sub-areas; some of the most important are objects detection, categorization and recognition, as well as face detection and recognition. More research took place in the recognition area than in the categorization area “in the past” (Pinz, 2006); although the majority of recognition algorithms isn’t directly applicable to categorization, the success achieved in the recognition field influenced several categorization approaches. As examples of the evolution in the object recognition field, Lowe, 2001 developed a system which uses SIFT (Scale Invariant Feature Transform) features, that is able to robustly recognize 3D objects in cluttered natural images, and, Su et al., 2006 implemented a framework for 3D object recognition in 2D images, by means of a similarity-based and aspect-graph approach. Passalis et al., 2007 showed that face recognition is connected to object retrieval, being the latter topic concerned about the search for objects - in a database - which are similar to a given query object; the latter authors presented a method for intraclass retrieval of 3D objects, which is appropriate for nonrigid object classes, and was applied to the face recognition domain. Also regarding object retrieval, Lam & du Buf, 2011 developed a 3D polygonal objects retrieval method which uses two sets of multiresolution signatures.



The content based 3D retrieval field is covered in Tangelder & Veltkamp, 2008, where the literature on methods concerning this subject is surveyed.

Biologically speaking, at the visual cortex level, the objects invariant detection, categorization and recognition depends on a cortical stages hierarchy where invariance is gradually built (Grossberg et al., 2011); this involves two data streams - “bottom-up” and “top-down” - in the “what” e “where” subsystems (Deco and Rolls, 2004), including also the integration of both (Farivar, 2009). Among the existent several cortical areas, there is area V1; the latter area contains simple and complex cells which are adjusted to different spatial frequencies, orientations and disparity (Hubel, 1995). Rodrigues & du Buf, 2009b presented a model for lines and edges (LEs) detection in V1, which is based on simple and complex cells responses. The same authors exemplify “the use of the multi-scale line/edge representation” for several processes: visual reconstruction (brightness perception), automatic scale selection and object segregation; finally, they tested a two-level 2D object categorization setting, and presented “a multi-scale object and face recognition model”. Also placed in cortical area V1, the end-stopped cells combine responses of complex cells tuned to different orientations (Rodrigues and du Buf, 2006); this latter cell type allows detecting line and edge crossings, singularities and points of large curvature. In Rodrigues and du Buf, 2006 the importance of multi-scale keypoint representation is investigated, and also object segregation, automatic scale selection, saliency maps building and face detection are approached. Having as basis these neuronal processing systems, it was possible to create a cortical architecture for 3D faces recognition (Rodrigues et al., 2011) and a first approach to 3D object recognition (Rodrigues et al., 2012). Rodrigues et al., 2012 explore the use of 2D projections in cortical 3D face and object recognition, as a development from the work presented in Rodrigues & du Buf, 2009b; in other words, Rodrigues and du Buf, 2009b presented a biological model in 2D which evolved to 3D by using a group of 2D cortical patterns (Rodrigues et al., 2012) for 3D recognition. More biologically inspired publications include Rodrigues & du Buf, 2011 where a HVS inspired framework for scene categorization was presented, the thesis by Sousa, 2009 about human facial emotions recognition, the cortex integrated multi-scale architecture by Rodrigues, 2007 and the “biologically plausible model to obtain a saliency map for Focus-of-Attention (FoA)” by Martins et al., 2009.

In the next sections the main general concepts presented in the above description are explored in detail, as well as the state-of-the-art for the main topics of this thesis.

## 2.1 - Categorization approaches gist and non-biological approaches description

Several categorization-related publications enriched the CV state-of-the-art, as for example the one by Leibe & Schiele, 2003 who studied the efficiency of some appearance- and contour-based recognition methods for multi-class object categorization; in the following year, Leibe & Schiele 2004 presented a multi-scale object categorization approach “using scale-invariant interest points and a scale-adaptive Mean-Shift search”; also, Jiang et al., 2007b explored various aspects of Bag of Features (BoF) for object categorization and semantic video retrieval, Wu et al., 2009 proposed a “scale invariant visual language model,” where the supervised information of the image level labeling is used to learn the object categories, Kinnunen et al., 2010 proposed an approach for unsupervised visual object categorization, Kim & Medioni, 2011 presented a scalable framework for 3D object categorization in range images, and, Han et al., 2011 evaluated a canonical correlation- and view-based approach both for object recognition and categorization; moreover, Petre & Zaharia, 2011 presented a scheme for categorization of 2D objects present in still images, which resorted to categorized 3D models, and, Duchenne et al., 2011 presented an object categorization approach which sees image matching as an energy optimization problem. For further categorization methods see also Savarese & Fei-Fei, 2010 and Stöttinger et al., 2012. We also refer to Galleguillos & Belongie, 2010 for a survey regarding the usage of different kinds of contextual information for a strong object categorization, and mention as well the article by Pinz, 2006, which presents foundations, original research and tendencies in object categorization by computer vision methods. Moreover, some biologically inspired categorization approaches can be named: Mundhenk et al., 2004 developed a method for clustering features - collected from the environment - into classes of objects, Chen et al., 2012 proposed an algorithm for feedforward categorization of objects (particularly human postures) in real-time video sequences and Mutch & Lowe, 2008 investigate the role of sparsity and localized features in a biologically-inspired model of visual object classification. Finally, as already mentioned, Rodrigues & du Buf, 2009b presented a model which was used as basis for this thesis: a model for lines and edges (LEs) detection in V1, which was based on simple and complex cells responses; the latter authors also tested a biologically inspired and two-level object categorization setting. For more biologically based publications on categorization, see, for example, Serre et al., 2007, Jiang et al., 2007a and Buckley & Sigala, 2010.

Along this chapter some of these approaches are described in detail: Leibe & Schiele, 2003, Han et al., 2011, Petre & Zaharia, 2011 and Rodrigues & du Buf, 2009b - being the latter the only biologically based from the mentioned four. All four publications are related to this thesis work; however, the publications by Leibe & Schiele, 2003, Han et al., 2011 and Rodrigues & du Buf, 2009b are more suited for comparing the final results because these three authors used the ETH-80 database (Leibe & Schiele, 2003) to test their approaches as it was done for evaluating the categorization model developed in this thesis.

Leibe & Schiele, 2003 presented ETH-80 – an 80 objects database specifically intended for object categorization. As referred in the latter paragraph, this database was used to test the biologically based categorization model developed in this thesis; particularly, the *eth80-cropped256* set (ETH-80, 2013) was used to test the latter model; the database is available for download in ETH-80, 2013. ETH-80 comprises 8 different categories, each one being represented by 10 objects; the objects show significant intra-class variations, even though they clearly belong to their respective categories. Each object is represented by 41 images from viewpoints equally spaced over the upper viewing hemisphere (at distances of 22.5 - 26°), hence, this database comprises a total of 3280 color images (of high-resolution); besides, a high-quality segmentation mask is provided for each image and also, a blue chromakeying background was used to collect the views (Leibe & Schiele, 2003). In order to explore the categorization of natural and human-made objects, ETH-80 includes: 3 “fruits & vegetables” categories – apples, pears and tomatoes -, 3 animal categories – cow, dog and horse -, one “human-made, small (graspable)” category – cups - and one “human-made, big” category – cars. Leibe & Schiele, 2003 used ETH-80 as a benchmark to individually evaluate the performance of several appearances and contour based methods for a multi-class object categorization purpose; the methods are as follows:

- a) The first method uses color histograms (Swain & Ballard, 1991); in this approach they collected a global RGB histogram over all image pixels belonging to the object; by comparing histograms, the test image is finally assigned to the category which has the closest matching histogram (Leibe & Schiele, 2003).
- b) Moving on to the second used method which is texture-based, Leibe & Schiele, 2003 used “a generalization of the color histogram approach to histograms of local grayvalue derivatives at multiple scales” (Schiele & Crowley, 2000). Two versions of this approach were compared: the first is a rotation-variant descriptor, and uses first

derivatives in  $x$  and  $y$  directions over 3 different scales; the second uses the gradient magnitude and the Laplacian – rotation invariant features -, also over 3 scales. Finally, in these approaches, the matching histograms alternative was used (Leibe & Schiele, 2003).

- c) In order to use the objects' global shape as a tool for categorization, Leibe & Schiele, 2003 used PCA-based methods (PCA: Principal component analysis), only reporting results in one of them; the authors reported results from a PCA-based method which consisted in building separate eigenspaces for each category and calculating the reconstruction error (Turk & Pentland, 1991), i.e., the class-specific eigenspace capability of representing the test image; two approaches were tested: in the first experiment they applied PCA to the segmentation masks, and on the second they applied it to the segmented grayvalue images (Leibe & Schiele, 2003).
- d) Finally, the last method was local shape based, or, more specifically, it was a contour based method. Leibe & Schiele, 2003 chose as reference a method based on the Shape Context (Belongie et al., 2001); in this method, an object view is represented by a discrete points set, and the referred points are regularly sampled along the internal or external contours. Referring to Belongie et al., 2001, Leibe & Schiele, 2003 compared two simpler approaches: (a) a dynamic programming approach was used to search a continuous path around the object contour; adjacent points in a contour are matched to the correspondent point in another contour; (b) the second approach comprises a “one-to-one matching between contour points using a greedy strategy”. In both approaches, the final score is given by the sum over all individual matching costs, and a comparison between shape context histograms is made (Leibe & Schiele, 2003).

Leibe & Schiele, 2003 reported their best results obtained using each of the 7 approaches here described. Using a leave-one-object-out cross validation, they used 79 objects for training, and tested the methods with the one remaining unknown object; besides, their results were averaged over the 80 available test objects. They also mention their experimental conditions: “we use the database for a best case analysis: categorization of unknown objects under the same viewing conditions, with a near-perfect figure-ground segmentation, and known scale.” The highest average categorization rate regarding the results obtained using the described approaches was 86.4% - a percentage achieved by both contour approaches. Evaluating the results, Leibe & Schiele, 2003 observed that no single method (mentioned above) is superior

for all categories, and that almost all of the referred methods are the best choice for at least one category. At this point, Leibe & Schiele, 2003 inferred that regarding multi-class object categorization, multiple features should be used as well as different combinations of these. This way, they explored four different cues - colour, rotation invariant texture, PCA on segmentation masks and contours -, obtaining a grouping hierarchy for each of them: “by iteratively grouping together those categories that are confused most often, we obtain a hierarchy of groupings.” While analysing the grouping hierarchies of the four different cues in Leibe & Schiele, 2003, we could notice that the ‘dog’ and ‘horse’ categories were considered a best grouping in terms of colour, rotation invariant texture, global shape and contours, and so were ‘apple’ and ‘pear’ in terms of rotation invariant texture; likewise, ‘apple’ and ‘tomato’ were considered a best grouping regarding global shape and contours. Finally, Leibe & Schiele, 2003 proved the significant potential in resorting to multiple features by using a decision tree (Duda et al., 2001) which at each level bases its decisions on one cue only; using an optimal multi-cue decision tree, the global categorization rate was improved to 93.02% (Leibe & Schiele, 2003).

In Han et al., 2011, a canonical correlation- and view-based approach was presented and evaluated both for object recognition and categorization; this approach explores local feature sets which are common in different views of the same object, as these sets can be “an effective representation of an image”. In order to measure the similarity between two local features sets, a canonical correlation analysis is performed, so that the similarity between two images can be calculated for object categorization. RGB SIFT is used to represent a local region of a colour image; also, an orthogonal subspace is extracted from a local feature set of an image, so that the similarity between two images can be measured by using a canonical correlation analysis on the orthogonal subspace for image representation. Two different view-based object datasets were used for different purposes: COIL-100 (Nene et al., 1996) was used for an evaluation in object recognition, and ETH-80 supported the object categorization tests. As in this thesis the results for the biological categorization model here developed were achieved using ETH-80, the experimental setup and results from Han et al., 2011 concerning the latter database were considered. As referred before, ETH-80 is organized in 8 different categories, each with 10 objects (Leibe & Schiele, 2003); Han et al., 2011 used - from each category - 9 objects for the reference (training) set and the remaining object for the test set. The training set consists of 5 views from each training object, and the testing set contains all views from the test objects. To accomplish object categorization, a canonical analysis was

used to compare the input orthogonal subspace to the reference orthogonal subspaces, where the similarity between the input and the reference subspaces is represented by the averaged correlation value. Several subspace dimensions were explored in the experiments, and an 82% categorization success rate was reached (Han et al., 2011).

The approach by Petre & Zaharia, 2011 is also described in detail, so that another non-biological categorization approach is provided - besides the ones by Leibe & Schiele, 2003 and Han et al., 2011. Petre & Zaharia, 2011 developed a novel 2D object categorization method, where by means of a matching between unknown 2D objects and categorized 3D models, the former objects (present in still images) are semantically labelled. A motivation presented by the authors relies on the fact that automatically identifying the semantics of an image's elements allows a machine to easily retrieve the required content. With the purpose of classifying a 2D object, the latter is compared to all 3D models, being the semantical matching mentioned above achieved by using 2D/3D indexing methods. The principle of 2D/3D indexing comprises the presentation of the 3D model as a 2D views set, in turn, each view is characterized using a 2D shape descriptors set. To provide a matching between 2D objects and 3D models, the same 2D shape descriptor is used for the query objects - which were beforehand manually segmented from still images. As to obtain 2D views from the 3D models, several projection strategies were also evaluated. The model was tested using (a) 115 randomly chosen 2D web images against the MPEG-7 3D model database (Zaharia & Prêteux, 2004) – which has 23 available categories -, and using (b) 65 of the latter 2D objects against the Princeton Shape Benchmark (PSB) (Shilane et al., 2004) – which has 161 available categories (Petre and Zaharia, 2011); the categorization success rate was explored in several cases - regarding the various 2D shape descriptors, projection strategies and number of candidate categories for final answer (1, 2 or 3) -, reaching 70.4% on case (a) and 64.6% on case (b). These maximum recognition rates were both achieved in the case of 3 accepted categories for final answer (Petre and Zaharia, 2011).

## **2.2 - Choosing the HVS as an inspiration**

As we could see in Leibe & Schiele, 2003, Han et al., 2011 and Petre & Zaharia, 2011, these non-biological categorization models show respectable results, but still our HVS is the fastest system for object categorization, being also the most effective, reliable and versatile in this regard; various reasons at the level of Neuroscience and Cognitive Psychology substantiate

this fact. However, for starters, a common sense justification is the fact that our HVS deals with real world complex images, still being able to categorize objects in real-time. We are able to decide if a quickly flashed image corresponds to a certain category in less than 150ms (Pinz, 2006), while the referred categorization models don't assure real-time speed. The HVS is selective, as we are capable of being indifferent to most information that is daily presented to us, and as well able to be very aware and immediately notice an object moving in our peripheral vision. This selectivity has a very specific name: "attention." Attention is therefore one of the most important perception mechanisms; by having this practical ability, our gaze is foremost and impulsively oriented to the most interesting zones in our visual field (Ruesch et al., 2008). Our brain is, in fact, a selective organ, because as visual attention orients our focus to the most interesting areas in the visual field, the brain is also able to choose among these areas data its most relevant information, thus storing it (Saleiro, 2011). Humans' attention is quickly drawn towards "eye-catching" objects and regions in the visual field; the ability of identifying these objects/regions is important for humans' survival (Elazary & Itti, 2008). Attention provides HVS with another significant characteristic: moderation. This feature was mentioned by Rensink, 2000, Brady et al., 2008 and Saleiro, 2011 in different words. As mentioned by Saleiro, 2011 the level of a person's attention and concentration conditions the visual perceptions' memory storage. A bigger attention level originates, this way, more detailed memory storage. Thereby, the HVS varies the quantity of absorbed detail depending on the circumstances (Brady et al., 2008). If it weren't like that, i.e, if our brain would analyse every single detail that is brought to our senses, it would be constantly busy (Rensink, 2000). This capability of just attending to certain parts of the visual field, selecting these areas' most important data and being capable of varying the detail absorption, are some of the many quality-features of the HVS. As selective and moderated, the HVS is also versatile; while walking on the beach, the country or other different scenarios, we are capable of recognizing and categorizing various objects which in turn may be different in appearance, or even belong to different classes. Cognitive psychology evidence shows that humans deal with approximately 30.000 different categories (Pinz, 2006 referring to Biederman, 1995), however, "this would require solving currently intractable computational complexity" (Pinz, 2006). Bornstein et al., 2010 stated that young infants have noteworthy categorization skills, being able to easily categorize objects - like faces, animals, furniture, vehicles, tools and plants. Also interestingly, object categorization in 6-month-old infants is flexible in its capacity of transcending variation in object-context relations (Bornstein et al., 2010). Having been presented the arguments which denote the superiority of the HVS, the following

paragraphs are intended to complement the information regarding object categorization and its link to Neuroscience and Cognitive Psychology.

While Bornstein et al., 2010 stated that robust object categorization is adaptive, Leibe & Schiele, 2003 - referring to Rosch et al., 1976 -, stated that categories are a learned representation which depends on education and experience (Leibe & Schiele, 2003). Bornstein et al., 2010 also wrote that object identification, memory, and categorization may depend on contextual cues; in this regard, Galleguillos & Belongie, 2010 denote the work by Palmer, 1975 who found that the object categorization task was eased if the target object was presented after an adequate scene, and impaired if the object and scene combination wasn't coherent. Analysing an object's surrounding context contributes for the object's categorization; a proper visualization for this statement is given by Fig. 1 in Galleguillos & Belongie, 2010, where these authors refer that the features of a certain object - which is shown isolated from its environment - are not enough to classify it. For more information about the use of contextual data in object categorization see Galleguillos & Belongie, 2010.

Kourtzi & Connor, 2011 refer to object vision as an incredible ability that is still mostly unexplained concerning the neural coding mechanisms. As Leibe & Schiele, 2003, Kourtzi & Connor, 2011 also denoted the importance of experience, mentioning that the latter is central in shaping structural and categorical coding for object perception; structure and category are distinct domains, however, it's still not clear how these interrelate (Kourtzi & Connor, 2011). Stating the last authors "learning optimizes the neural processes that mediate binding of local elements and parts into objects, recognition of objects across image changes that preserve identity (e.g., position, orientation, clutter), and selection of behaviorally relevant features for object categorization." Understanding the mechanisms which mediate the adaptive coding of classes is very important to comprehend our capacity of making flexible perceptual decisions (Kourtzi & Connor, 2011). Again citing the latter authors: "extensive behavioral work on visual categorization (e.g., Goldstone et al. 2001) suggests that the brain learns the relevance of visual features for categorical decisions rather than simply representing physical similarity."

The HVS was often thought as a processes sequence - detection, segregation, categorization and recognition; however, these processes cannot occur entirely in a series (Bar et al., 2006); as mentioned by Saleiro, 2011, recent researches suggest that objects categorization and segregation occur simultaneously, or that categorization takes place before segregation



(Rodrigues & du Buf, 2009a). Despite the researches made in this regard, the processes exact sequence is still unknown as are the cases of processes parallelization (Saleiro, 2011). Understanding the HVS, proposing hypotheses about its behaviour and trying to integrally reproduce it, is a subject of much work, once the brain's behaviour is not entirely known.

As mentioned in chapter 1, this thesis goal is to categorize 2D objects using a biologically inspired model which is based on V1 LEs cortical attributes. In the next section, some basic concepts regarding the HVS are presented.

### **2.3 - General concepts about the Human Visual System**

Our eyes gather exterior information in the form of light which is projected in the photosensitive retina and converted in electrical signals; the signals from the two retinas are forwarded to the optic nerve and next to “two peanut-size nests of cells” – the lateral geniculate bodies (Lateral Geniculate Nucleus - LGN). The LGN's ending fibres connect to the visual cortex -, or more specifically, the striate cortex (primary visual cortex) -, from which the information is sent to several neighbouring higher visual areas (Hubel, 1995).

Retina is a plate composed by rods and cones - cells which are photosensitive in different ways. The cones provide us our detailed and colourful vision, while the other photosensitive cells – the rods – make us capable of perceiving the environment in dim light (Hubel, 1995). The human retina is composed by one single type of rod, and yet, three types of cones (Troncoso et al., 2011). The latter photoreceptors, “responsible for color vision, are called L (or red) cones, M (or green) cones, and S (or blue) cones”, having each cone type its maximum sensitivity in different wavelengths of the light spectrum. Rods represent a bigger ratio in the retina than the cones; the concentration of the latter cells varies across the retina: although cones are found spread in the entire retina, only these cells are found in this organ's central area - named fovea -, a rod-free area, where our fine-detail vision is the best (Hubel, 1995). Additionally, more cell types take place in the retina, such as the ganglion cells, bipolar cells, horizontal cells, and amacrine cells (Hubel, 1995).

Nerve cells, or neurons, are composed mainly by their nucleus, dendrites and axons. A nerve cell has several dendrites for receiving the input signals from the other cells, and one axon (a nerve fibre) which allows the first to transmit a response. The optic nerve, which makes the

connection between retina and the LGN, is composed by a group of axons from “the third stage retinal cells” - the ganglion cells (Hubel, 1995).

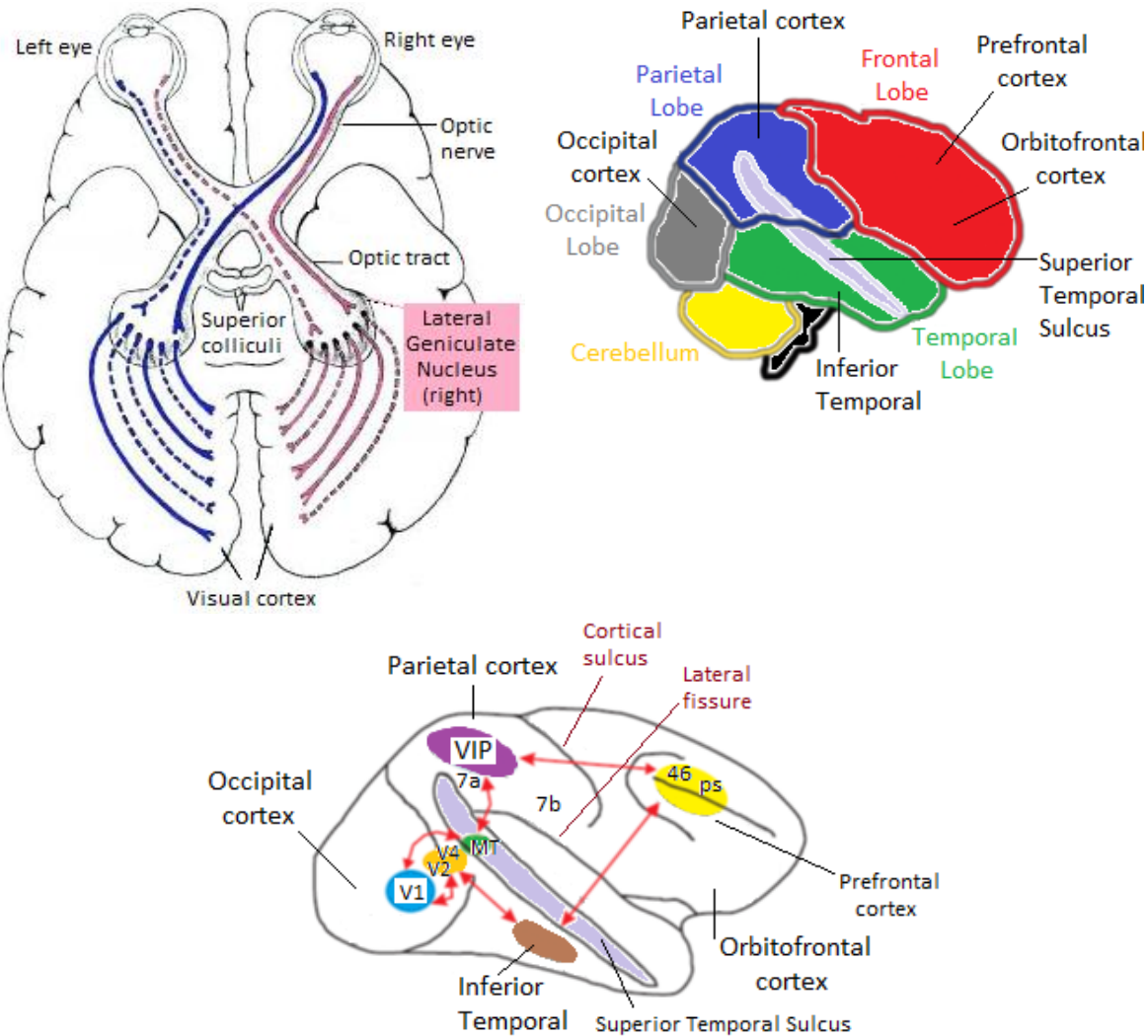
Moving from the retina, and through the optical nerve, the signals are forwarded to the LGN; the latter is found in the thalamus and is a part of the central nervous system (Hubel, 1995). The LGN receives its main input from the retina, being indirectly connected to a certain population of rods and cones; this clustered cells population in the retina, which feeds into a certain cell in the visual pathway, comprises the receptive field of this cell; in other words, the receptive field is “the outer world as seen by a single cell”, contrasting with the concept of visual field which is our perceptions of the external environment. In addition, the LGN cells’ receptive fields are characterized by the same disposition of ganglion cells that feed into them (Hubel, 1995).

When in presence of a stimulus in the receptive field (RF) region, the neuron may fire a signal. As an example, ganglion cells - which have a circular shape and two concentric regions - are divided in two types: on-center and off-center cells. The on-center cells fire a signal if light-stimulated on their RF, suppressing the response when stimulated outside this region, and firing again when the light is turned off; the two last phenomena – suppression followed by signal discharging comprise an “off response”, being the first referred occurrence denominated “on response”. The off-center cells have an inverse reaction: their RF provided off responses, and its surroundings presented on responses (Hubel, 1995). In addition, citing Hubel, 1995: “because convergence occurs at every stage, receptive fields tend to become larger: the farther along the path we go, the more fuzzy this representation-by-mapping of the outside world becomes”; also, the receptive fields of the primary cortex cells are more complex than those “behind in the visual pathway” from the retina and LGN, increasing their complexity as far as they are localized in the pathway (V1, V2, V3, etc.).

As referred before, the data moves forward and mainly to an area in the visual cortex named primary cortex V1, nevertheless, the visual cortex is also composed by other cortical areas. For further details concerning the HVS, see Hubel, 1995 and Troncoso et al., 2011. Also noteworthy, is the fact that the existing connections between visual cortex and LGN are retinotopic projections, i.e, the left-right and up-down relations in the spatially transported data are preserved in the retina, LGN and visual cortex (Sousa, 2009). For a better visualization of the named regions, consider the top-left image from Fig. 2.3.1, which shows the brain’s horizontal cross section, giving some perspective about the LGN position in the

brain as well as the optic nerves' and the visual cortex's; consider also the top-right image which shows the anatomical regions of the human brain.

At the visual cortex level, the objects invariant detection, categorization and recognition depends on a cortical stages hierarchy where invariance is gradually built (Grossberg et al., 2011); this involves two data streams - "bottom-up" and "top-down" - in the "what" e "where" subsystems (Deco and Rolls, 2004), including also the integration of both (Farivar, 2009). For a proper visualization, consider the bottom image in Fig. 2.3.1 which illustrates the datastream based on the cortical architecture from Deco & Rolls, 2004.



**Figure 2.3.1** Top-left: Brain's horizontal cross section, adapted from Fisiologia ocular, 2013. Top-right: the brain's anatomical regions, adapted from Fig. 3.3 in Sousa, 2011; bottom: the datastream based on the cortical architecture from Deco and Rolls, 2004, adapted from Fig. 1 from Deco & Rolls, 2004.

Cortical area V1 is composed by a stack of selective spatiotemporal filters; accordingly, among other spatiotemporal functions, the latter are able to execute several processing tasks concerning frequency, orientation, movement, direction, velocity (Olshausen & Field, 2005). Cortical area V2 receives signals from V1, forwarding them to V3, V4 and to the middle temporal area (MT). Although V2 is characterized by more complex functions than V1, both areas are known for having similar functions (Qiu & von der Heydt, 2005); object segregation, a complex function, is an example for the latter consideration.

V3 takes place in the dorsal datastream, receiving inputs from V2 and V1; some of its cells are sensitive to orientation, being many of them tuned to movement and depth (Gegenfurtner et al., 1997). Cortical area V4 – the third area in the ventral datastream - is the first area which shows strong attention modulation (Chelazzi et al., 2001). Although dealing with more complex shapes, V4 deals with the same attributes that concern V1: orientation, scale and colour. Another cortical region is the Inferior Temporal area (IT); this region deals with more complex shapes than cortical area V4 (Hubel, 1995), and it's found upper in the ventral datastream, dealing with shapes' and objects' visual representations (Logothetis et al., 1995); additionally, IT sends data to the prefrontal cortex (PF) which is involved in the connection between perception, and, memory and action (Miller, 2000). Finally, finishing the visual cortex areas description, the previously mentioned MT area is composed by many neurons which are movement selective, and also sensitive to complex attributes, such as lines' and curves' endings; this region processes more complex movement stimuli (Hubel, 1995).

In terms of “brain networks” regarding categorization, “a large network of cortical and subcortical areas has been implicated in visual category learning” (Kourtzi & Connor, 2011). These authors mentioned that areas in the prefrontal cortex, the basal ganglia, the medial temporal cortex and occipitotemporal regions are related to categorization (see Fig. 2.3.1, top-right and bottom): areas in the prefrontal cortex have been connected to rule-based tasks in which the category structure is defined by a single stimulus dimension, the basal ganglia have been linked primarily to information-integration tasks, the medial temporal cortex has been connected to category-learning tasks which rely on memorization, and, occipitotemporal regions are engaged by prototype-distortion tasks (Kourtzi & Connor, 2011).

Some general concepts about the HVS were covered in this section; the following section was reserved for the explanation about the V1 LEs detection model (Rodrigues & du Buf, 2009b) which was used as a basis for the categorization model developed in this thesis.

## 2.4 - Cortical attributes: LEs detection and classification

Hubel and Wiesel report the existence of several cell types in V1 and V2; among these are the simple, complex and hypercomplex cells (Hubel, 1995). These cells can be modelled using Gabor filters which are also used in Computer Vision applications such as texture, document analysis, retina identification, coding, and image representation (e.g. Weldon et al., 1996). The complex cell's modelling can be done through simple cells, and hypercomplex cells can be modelled by complex cells; the hypercomplex cells can be also named 'end-stopped' (Rodrigues & du Buf, 2009a).

Gabor filters possess a very proximate response to the simple cells' (Lee, 1996), being as well the best solution - presented until now - to mathematically model this cell's type (Sousa, 2009). These cells have several scales and orientations, being type, orientation and scale of the stimuli influence factors for their response; also, simple cells can either be even or odd (Sousa, 2009). The receptive field of simple cells can be approximated by function  $g_{\lambda,\sigma,\theta,\varphi}(x,y)$ , which is centred in the origin:

$$g_{\lambda,\sigma,\theta,\varphi}(x,y) = \exp\left(-\frac{\tilde{x}^2 + \gamma\tilde{y}^2}{2\sigma^2}\right) \cos\left(2\pi\frac{\tilde{x}}{\lambda} + \varphi\right). \quad (2.4.1)$$

In the simple cells' receptive field function,  $\tilde{x}$  and  $\tilde{y}$  are respectively given by  $\tilde{x} = x\cos\theta + y\sin\theta$  and  $\tilde{y} = y\cos\theta - x\sin\theta$ ;  $\gamma$  is the spatial aspect ratio - responsible for determining the ellipticity of the RF ( $\gamma = 0.5$ );  $\sigma$  is the standard deviation of the Gaussian factor, which directly influences the size of the RF;  $\lambda$  is the wavelength;  $1/\lambda$  is the cosine factor's spatial frequency;  $\sigma/\lambda$  is a ratio which reflects the bandwidth of the spatial frequency, or, in other words, the number of parallel excitatory and inhibitory stripe regions (see left image in Fig. 2.4.1);  $\theta$  defines the orientation;  $\varphi$  defines the symmetry of the RF's function (Eq. 2.4.1) with respect to the origin (Grigorescu et al., 2003). The ratio  $\sigma/\lambda$  was used equal to 0.56, which yields a half-response width of one octave (Rodrigues & du Buf, 2009b); as a reference to the angle parameter  $\theta$ , 8 orientations were used; the  $\varphi$  variable was used equal to 0 for an RF's even symmetry, and  $-\pi/2$  for an odd one (Grigorescu et al., 2003).

The left image in Fig. 2.4.1 presents a Gabor function's intensity map (adapted from Grigorescu et al., 2003) which models a simple cell's RF profile; Grigorescu et al., 2003 explain that grey levels lighter than the background are positive values, being the darker the

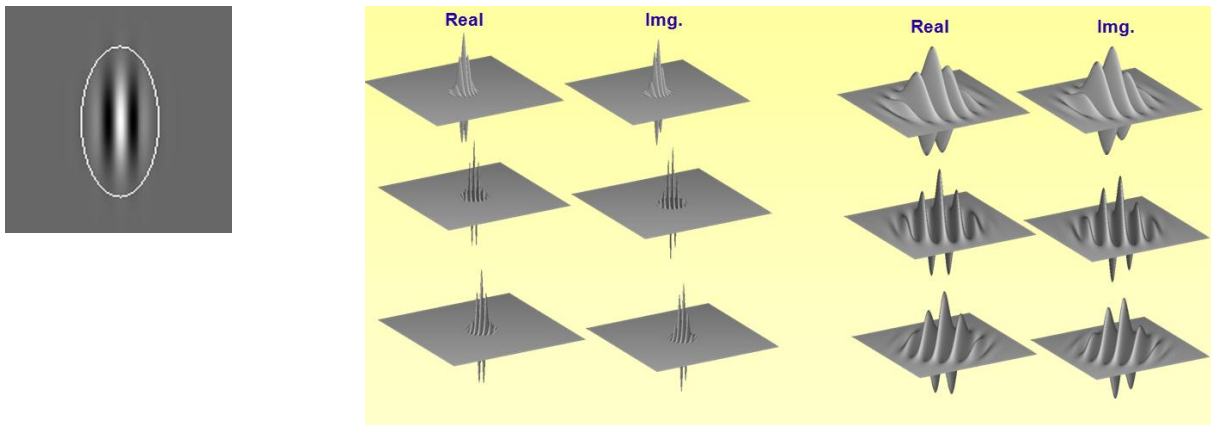
negative ones. Also the visible denoted ellipse designs “the boundary of the (classical) receptive field outside which the function takes negligibly small values” (Grigorescu et al., 2003). As said before, simple cells can either be even or odd. As an illustration for the modelling of these two cell types, the even simple cells’ are represented in the first and third column of the right image in Fig. 2.4.1, and the odd simple cells’ are in the remain columns. As well in this figure, the two most left columns represent a certain scale, being the remaining columns illustrating another scale, and, each of the three rows are linked to a different orientation (Sousa, 2009).

Considering the receptive field function given by Eq. 2.4.1, the response of a simple cell to an input image with luminance spatial distribution  $f(x,y)$  can be given by

$$R_{\lambda,\sigma,\theta,\phi}(x,y) = f(x,y) * g_{\lambda,\sigma,\theta,\phi}(x,y) = \iint_{\Omega} f(u,v)g_{\lambda,\sigma,\theta,\phi}(x-u,y-v)dudv . \quad (2.4.2)$$

Considering even and odd simple cells, the responses of both are given, respectively, by  $R_{s,i}^E(x,y)$  (for  $\varphi=0$ ) and  $R_{s,i}^O(x,y)$  (for  $\varphi = -\pi/2$ ), which are correspondent to the Gabor filter’s real and imaginary parts (Rodrigues & du Buf, 2009b);  $s$  refers to the scale of analysis which was given in terms of  $\lambda$  expressed in pixels, where  $\lambda=1$  corresponds to 1 pixel (Rodrigues & du Buf, 2009b); moreover,  $i$  is the orientation ( $\theta_i = i\pi/N_\theta$ ) and  $N_\theta$  is the number of orientations ( $N_\theta = 8$ ) (Rodrigues & du Buf, 2009b). This way, complex cells are modeled by two simple cells responses with a phase difference of  $\frac{\pi}{2}$  (Rodrigues & du Buf, 2009b)

$$C_{s,i}(x,y) = \left[ \left\{ R_{s,i}^E(x,y) \right\}^2 + \left\{ R_{s,i}^O(x,y) \right\}^2 \right]^{\frac{1}{2}} . \quad (2.4.3)$$



**Figure 2.4.1** - Left: Intensity map of a Gabor function, which models the receptive field profile of a simple cell (adapted from Fig. 3 in Grigorescu et al., 2003); Right: Modeling even/real simple cells (first and third column), and odd/imaginary simple cells (second and fourth column) (adapted from Fig. 3.4 in Sousa, 2009).

Several attributes can be obtained using the simple and complex cells, such as lines and edges (LEs). These cells respond beyond line and edge terminations; also, at line or edge crossings and junctions, detection leads to continuity of the dominant events with biggest amplitudes but to gaps in the sub-dominant events – gaps which must be reduced for the purpose of rebuilding continuity (Rodrigues & du Buf, 2009b). These issues were solved by using a lateral ( $L$ ) and cross-orientation ( $C$ ) inhibition (Rodrigues & du Buf, 2009b)

$$I_{s,i}^L(x, y) = \left[ C_{s,i}(x + d\hat{C}_i, y + d\hat{S}_i) - C_{s,i}(x - d\hat{C}_i, y - d\hat{S}_i) \right]^+ + \left[ C_{s,i}(x - d\hat{C}_i, y - d\hat{S}_i) - C_{s,i}(x + d\hat{C}_i, y + d\hat{S}_i) \right]^+, \quad (2.4.4)$$

$$I_{s,i}^C(x, y) = \left[ C_{s,(i+N_\theta/2)}(x + 2d\hat{C}_i, y + 2d\hat{S}_i) - 2C_{s,i}(x, y) + C_{s,(i+N_\theta/2)}(x - 2d\hat{C}_i, y - 2d\hat{S}_i) \right]^+, \quad (2.4.5)$$

with  $(i + N_\theta/2) \perp i$ ,  $C_i = \cos \theta_i$ ,  $S_i = \sin \theta_i$ ,  $d=0.6s$  and  $[.]^+$  denoting a halfwave rectification to suppress negative responses (Rodrigues & du Buf, 2009b). Also, inhibition is applied to the responses of complex cells, where  $\beta$  controls the strength of inhibition. Rodrigues & du Buf, 2009b used  $d=0.6s$  and  $\beta=1$ ,

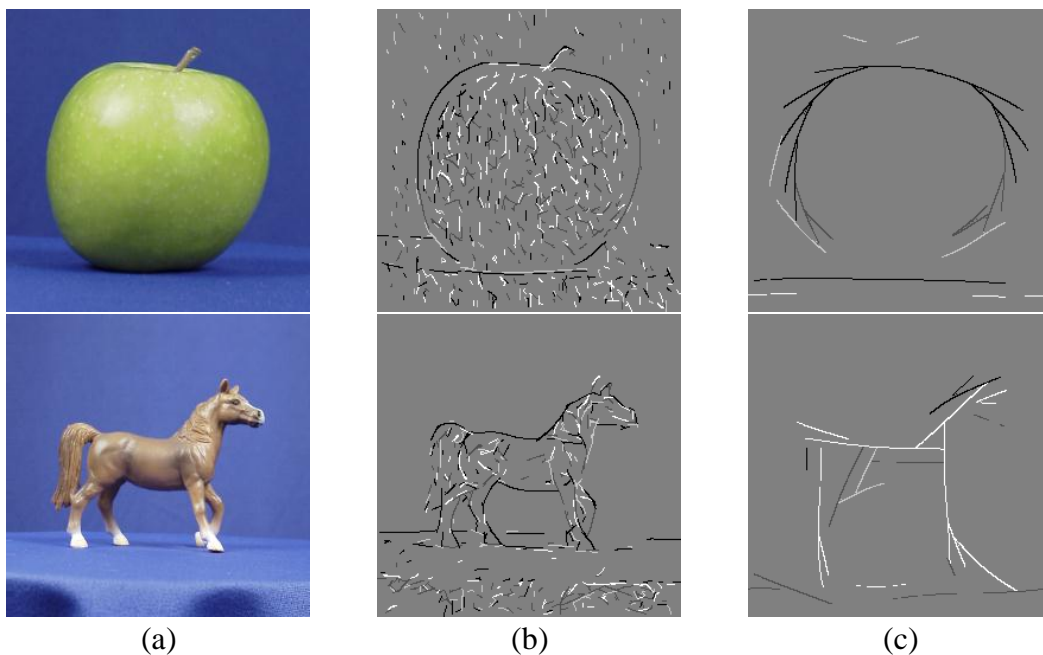
$$\hat{C}_{s,i} = [C_{s,i}(x, y) - \beta(I_{s,i}^L(x, y) + I_{s,i}^C(x, y))]^+. \quad (2.4.6)$$

Once detected, events (lines and edges), can be found in four distinct combinations of type and polarity in an image: positive lines, positive edges, negative lines and negative edges. Each of these events corresponds to three simultaneous occurrences regarding the different cells used; see Table 2.4.1. As an example, positive lines represent (simultaneously) a maximum response from even simple cells, an odd simple cells zero-crossing, and a complex cells maximum response after inhibitions were applied (Rodrigues & du Buf, 2009b). A positive line corresponds to a white line in black background, and a negative line to a black line in a white background; moreover, a negative edge corresponds to a white to black transition, and the inverse sequence links to the positive edge. As an example, the images in Fig. 2.4.2 illustrate LEs detection in an apple and a horse from ETH-80 (Leibe and Schiele, 2003), using two different scales: a fine ( $\lambda = 5$ ) and a coarse one ( $\lambda = 47$ ); the positive lines are represented in white, the negative lines in lighter gray, the positive edges in dark gray, and

the negative edges in black. For more details about the LEs cortical attributes, see Rodrigues & du Buf, 2009b.

**Table 2.4.1** - Cells responses combination for lines and edges detection (Max. - Maximum, Min. - Minimum, Z.C. – Zero-Crossing). Further details can be seen in Rodrigues & du Buf, 2009b.

Cell:	Simple, even	Simple, odd	Complex
Positive line	Max.	Z.C.	Max.
Positive edge	Z.C.	Max.	Max.
Negative line	Min.	Z.C.	Max.
Negative edge	Z.C.	Min.	Max.



**Figure 2.4.2** - Lines and edges detected in (a) an apple and a horse from ETH-80 (Leibe and Schiele, 2003), using (b) a fine scale  $\lambda = 5$  and (c) a coarse scale  $\lambda = 47$ ; four event types can be spotted: positive lines (white), negative lines (light-gray), positive edges (dark-gray) and negative edges (black).

After the LEs model's description, the biologically-based categorization scenario from Rodrigues & du Buf, 2009b which was also based on the V1 LEs detection model (Rodrigues & du Buf, 2009b) is provided next, in the following section, among other biologically inspired categorization models.



## 2.5 - Biologically inspired categorization models

The last section described the model for LEs detection in V1 by Rodrigues & du Buf, 2009b – an important model for this thesis, since it was used as basis for the biologically inspired categorization model here developed. As referred before, the LEs detection model by Rodrigues & du Buf, 2009b was based on simple and complex cells responses (Rodrigues & du Buf, 2009b), being as well multi-scale “with no free parameters” (Rodrigues & du Buf, 2009b). Rodrigues & du Buf, 2009b demonstrated that very promising results can be obtained using a cortical model, having exemplified “the use of the multi-scale line/edge representation” for several processes: visual reconstruction (brightness perception), automatic scale selection, and, also, object segregation (Rodrigues & du Buf, 2009b); Rodrigues & du Buf, 2009b also applied the detected LEs to object categorization and recognition: they tested a two-level 2D object categorization setting, using coarse scales for pre-categorization, and, coarse and fine scales for a final categorization level; besides, the latter authors presented “a multi-scale object and face recognition model”. Finally, processing schemes were “discussed in the framework of a complete cortical architecture” (Rodrigues & du Buf, 2009b). Once given a gist of the work by Rodrigues & du Buf, 2009b, we focused specially on the categorization scenario that was created.

Rodrigues & du Buf, 2009b used the ETH80 database (Leibe & Schiele, 2003) to test their two-level categorization approach. They chose 10 distinct images from 8 categories – dogs, horses, cows, apples, pears, tomatoes, cups/mugs and cars – which were a total of 80 images. The explored two-level categorization approach worked as follows: “three types of objects (horses, cows, dogs) are first grouped (animal), which we call *pre-categorization*, after which *categorization* determines the type of animal” (Rodrigues & du Buf, 2009b). The first level, pre-categorization, was based on LEs contours templates (solid objects) which were obtained through segregation, in order to generalize shape and remove surface detail. Once the object views were normalized, “and because different objects within each group are characterized by about the same line/edge representations at coarser scales, group templates can be constructed by combining randomly-selected images” (Rodrigues & du Buf, 2009b). The multi-scale LEs representation was made within  $\lambda \in [4, 32]$  by means of 8 equally spaced scales. The aim of the pre-categorization task was to select one among the following groups: animal, fruit, cup or car. At this point, the three coarsest scales within  $\lambda \in [4, 32]$  were used: 24, 28 and 32 pixels. Group templates were created in 3 ways: (a) all 80 images were used - 30 animals, 30 fruits,

10 cups, 10 cars -, (b) half of the images were randomly selected and used as templates, where a 30 images group was reduced to 15 and a 10 images group was reduced to 5, and, finally, (c) a third of the images were randomly selected and used as templates, where a 30 images group was reduced to 10, and a 10 images group was reduced to 3. Stating Rodrigues & du Buf, 2009b: “for each group template, at each of the three scales, a positional relaxation area was created around each responding event cell, by assuming grouping cells with a dendritic field size coupled to the size of underlying complex cells (Bar, 2003). These grouping cells sum the occurrence of events in the *input images* around event positions in the *templates* (a sort of local correlation). Mathematically, at every scale and at every position where events of a template are stored in memory, local grouping cells with circular dendritic fields at a certain position are activated. The “activities of all activated local grouping cells”, grouped together, are calculated through a global correlation. Next, the “global groupings were summed over scales” and the template which held the highest response was chosen (Rodrigues & du Buf, 2009b). Without performing spatial relaxation, the model remains almost the same with the exception of the circular dendritic field which was reduced to 1 pixel, and the fact that the local grouping works fundamentally as the logical AND function. Table 2.5.1 presents the categorization results in the form of mean (standard deviation); furthermore, the majority of the errors occurred between car/animal and cup/fruit (Rodrigues & du Buf, 2009b).

**Table 2.5.1** – This table corresponds to Table 1, in Rodrigues & du Buf, 2009b.

	All	Half	Third
Pre-categorization template construction	30/10	15/5	10/3
Error without relaxation	0.0%	5.7%(0.6)	8.0%(1.7)
Error with relaxation	0.0%	3.0%(1.0)	4.3%(0.6)
Categorization template construction	10	5	3
Error with relaxation	0.0%	9.3%(2.1)	12.7%(4.0)

For the next level - categorization -, the tests were performed using the 8 templates, all 80 images, and also the multi-scale LEs maps were applied at all 8 scales of the real input images. Like in pre-categorization, templates were built by means of random selections. Global correlations were compared considering the 8 scales, being selected the one with highest number of correspondences. Finally, Table 2.5.1 presents the results achieved for the categorization tests using positional relaxation. Rodrigues & du Buf, 2009b mentioned that “typical miscategorizations were dog/cow, horse/dog, horse/cow and apple/tomato” (Rodrigues & du Buf, 2009b). Finally, stating Rodrigues & du Buf, 2009b: “only about 9

errors in 80 images (the “50/50 training and testing” scenario) is a very promising starting point for refining the algorithms” (Rodrigues & du Buf, 2009b).

There is more literature in CV concerning object categorization, however much less in biological vision (Rodrigues and du Buf, 2009b). Regarding biologically inspired methods, as mentioned before, we also name those designed by and Mundhenk et al., 2004, Mutch & Lowe, 2008 and Chen et al., 2012. Mundhenk et al., 2004 developed a method for clustering features - collected from the environment - into classes of objects; these features (including intensity, orientation and colour) were collected from the image’s most salient points; for the latter purpose, a biologically inspired saliency program was used (Mundhenk et al., 2004). Mutch & Lowe, 2008 investigated “the role of sparsity and localized features in a biologically-inspired model of visual object classification” (Mutch & Lowe, 2008). The latter authors used Gabor filters at every position and scale, and also built up feature complexity and position/scale invariance (Mutch & Lowe, 2008). Increasing sparsity was found beneficial to improve generalization performance; sparsity was “increased by constraining the number of feature inputs, lateral inhibition, and feature selection” (Mutch & Lowe, 2008). Finally, Mutch & Lowe, 2008 denote as well the importance of “retaining some position and scale information above the intermediate feature level” (Mutch & Lowe, 2008). Chen et al., 2012 proposed “an algorithm for feedforward categorization of objects” (particularly human postures) “in real-time video sequences” (Chen et al., 2012). Stating Chen et al., 2012: “the system employs an innovative combination of event-based hardware and bio-inspired software architecture” (Chen et al., 2012); the software module collects size and position invariant line features which were inspired by primate visual cortex models. The latter authors employed “a simplified line segment Hausdorff distance scheme” for the purpose of measuring similarity, and achieved size and position invariance “by deriving size and position information from event clusters” (Chen et al., 2012). For more details on the models by Mundhenk et al., 2004, Mutch & Lowe, 2008 and Chen et al., 2012, please take a look at the respective publications.

This chapter - which covered some general concepts and state-of-the-art -, started by providing a gist regarding object categorization publications; detailed descriptions of some of these publications were given, namely of the publications Rodrigues & du Buf, 2009b, Leibe & Schiele, 2003, Han et al., 2011 and Petre & Zaharia, 2011. Given the fact that the LEs detection model by Rodrigues & du Buf, 2009b was used as basis for the categorization

model developed in this thesis, a proper description about the former was as well provided. Being the categorization model developed in this thesis biologically inspired, some general concepts related to the HVS were also described. Next, the following chapter comprises a detailed description of the biologically inspired categorization model developed in this thesis.

# 3. Categorization Model

---

## ABSTRACT

**This chapter presents a detailed description of the categorization model's development.**

---

The previous chapter covered the state-of-the-art and the main concepts related to this thesis: detailed descriptions of some categorization approaches were given – biologically inspired or not. As our HVS is the best for the categorization task, the categorization model here developed was based on the LEs detection model by Rodrigues & du Buf, 2009b, thus, a proper description of the latter model was already provided (section 2.4), as well as some HVS biological background. The developed categorization model is thereby described in this chapter, and the proof of concept – which was made using the ETH-80 database (Leibe and Schiele, 2003) - is provided in the next chapter.

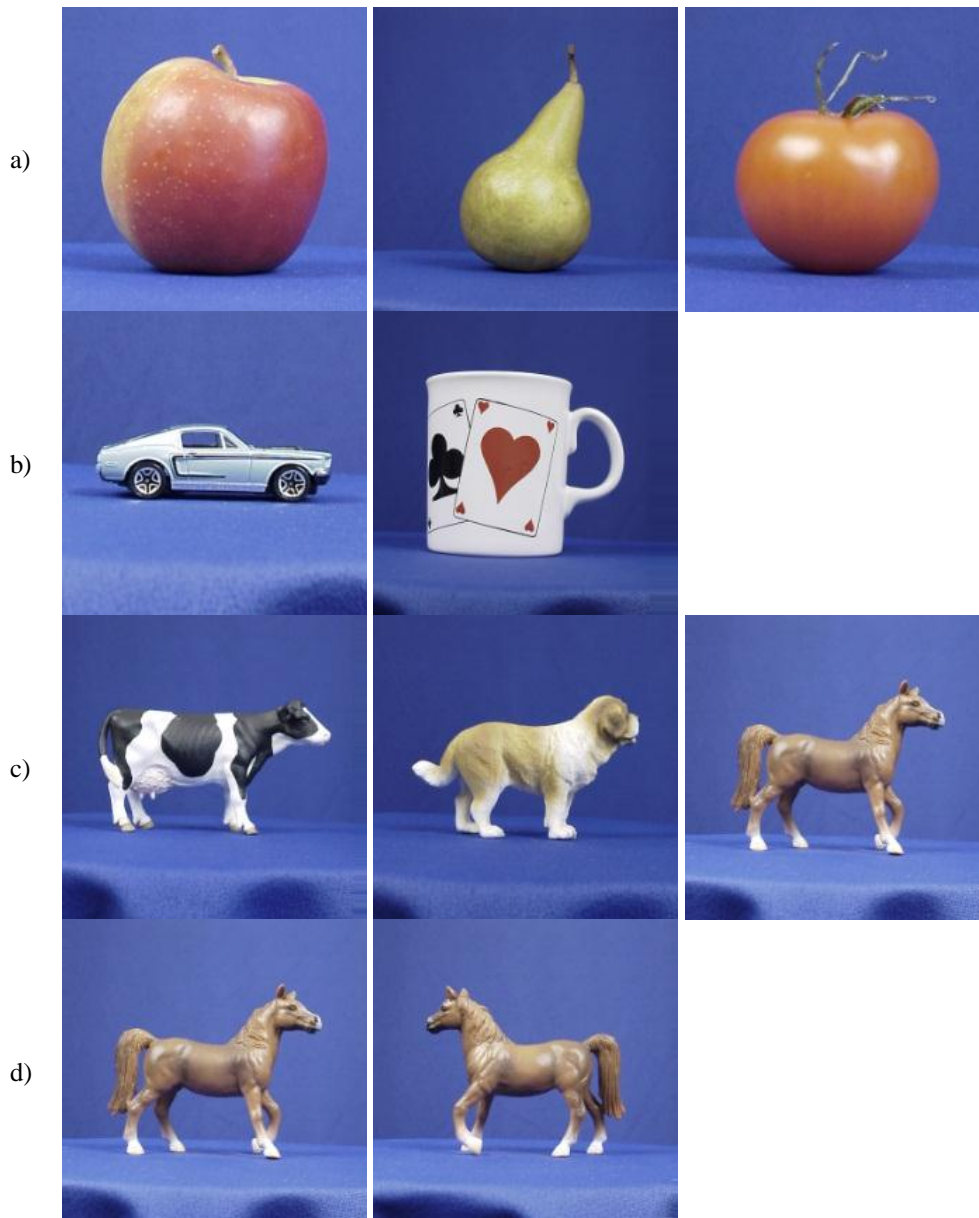
### 3.1 – The categorization model: an introduction

There are several available databases that could be used to test the model, such as Caltech 101 (Fei-Fei et al. 2004) or ETH-80 (Leibe & Schiele, 2003). The latter database was chosen for the proof of concept because: (a) the objects in this database are ‘almost’ segregated (this topic is brought up latter in this section). (b) ETH-80 has different available categories and several objects per category and (c) the achieved results can be easily compared with previous categorization results from Rodrigues and du Buf, 2009b obtained with ETH-80.

As referred before, the *eth80-cropped256* set was used; in this dataset, 8 categories were available: apples, cars, cows, cups, dogs, horses, pears and tomatoes (Leibe & Schiele, 2003; ETH-80, 2013). As an example, Fig. 3.1.1 shows a sample object from each category: a) fruits and vegetables, b) man-made objects and c) animals.

By observing Fig. 3.1.1, it can be noticed that each image in the *eth80-cropped256* dataset is composed by one centred object; moreover, all the images are rescaled to a size of  $256 \times 256$  pixels, they have a 20% border area, and all images of a particular object have equal scale.

For more information on this dataset see Leibe and Schiele, 2003 and ETH-80, 2013; the database is available for download in ETH-80, 2013.

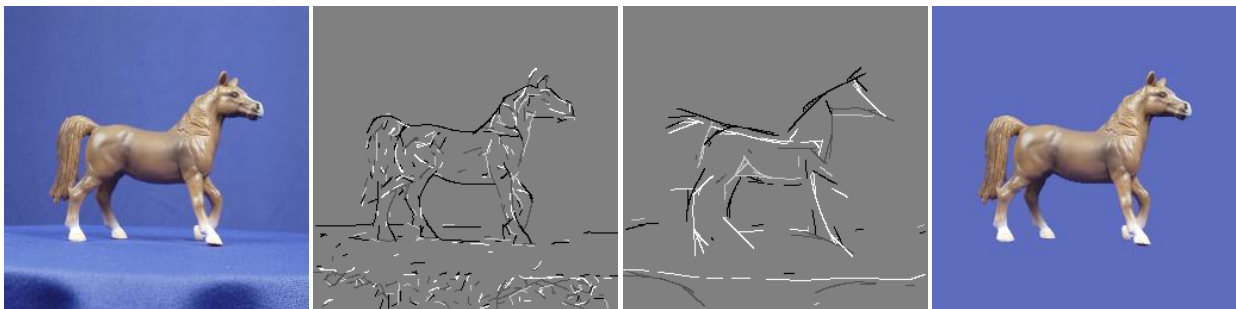


**Figure 3.1.1** - Objects samples from the eth80-cropped256 set (ETH-80, 2013). a) to c) a sample object from each category in the eth80-cropped256 set: (a) fruits and vegetables - apple, pear and tomato, (b) man-made objects - car and cup and (c) animals: cow, dog and horse. Bottom line (d): horse7 shown in left and right profile.

As already mentioned, in the ETH-80 database each category has 10 objects, being each of the latter available in 41 views (see examples in Fig. 6.1.1, Appendix 6.1). From all the 41 views, two specific ones were chosen to work with: left and right profile - see in Fig. 3.1.1 d) horse7 in both views. This way, there were chosen 160 objects images: 80 objects in left and right profile. The algorithm was trained using a sub-group from the chosen 160 images. Half of

these images - 80 images (40 objects in both views) - was used for training (templates images), and the other half (80 images) was used to evaluate the validity of the built model (test objects images). In summary, from the original database all the objects are used in two views; there are 10 objects per category, from which 5 were used as template/training objects, and the remaining 5 were used as test objects.

It was mentioned above that the objects were ‘almost segregated’, as Leibe & Schiele, 2003 applied “a blue chromakeying background for easier segmentation” (Leibe & Schiele, 2003). However, in the figures distinguished until this point in this chapter, it can be noticed that the ETH-80 objects images’ background isn’t truly homogeneous; moreover, when the lines and edges detector algorithm from Rodrigues and du Buf, 2009b was used directly on ETH-80’s objects, many lines and edges were detected in the images background, especially at the bottom - on the base on which the objects were positioned. This is illustrated in Fig. 3.1.2, where a horse is presented in right profile, and also the horse’s LEs detected for  $\lambda = \{5, 20\}$  are shown. Furthermore, a previous example of LEs detection in objects images was already shown in Fig. 2.4.2, in section 2.4. We also refer to the same section for a detailed explanation about these cortical attributes.

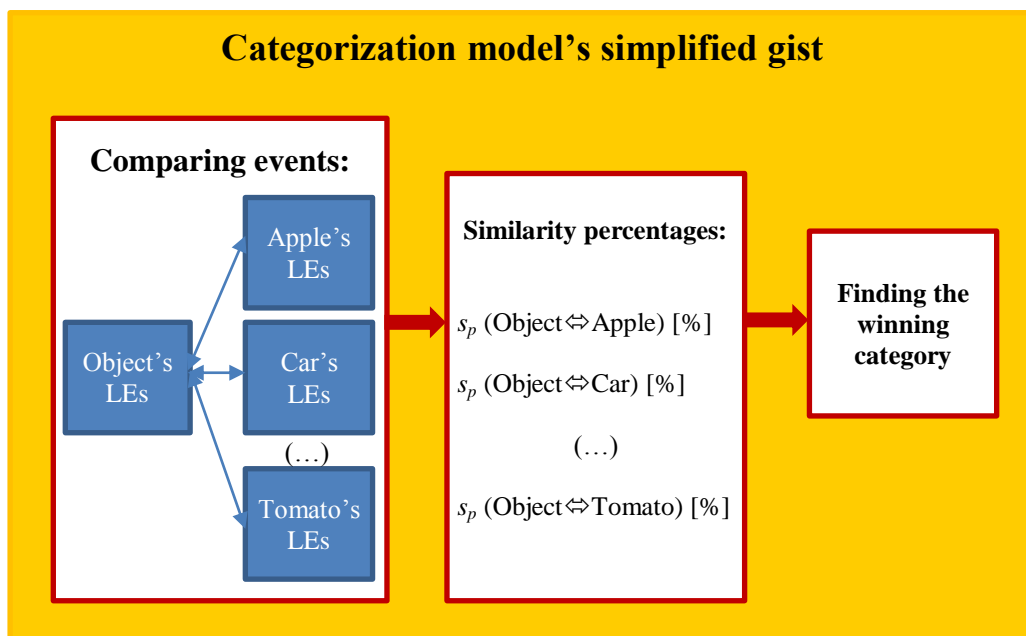


**Figure 3.1.2** - From left to right, an horse shown in right profile, corresponding lines and edges detected using  $\lambda=5$  and  $\lambda=20$ . In the right most image the same horse image with the background homogenized.

Being objects categorization the goal of the thesis and segregation a topic out of the focus, the background of all 160 images was manually homogenized, so that the found LEs in the images would only belong to the used objects. This homogenization is shown in the right most image of Fig. 3.1.2; more examples are shown in Fig. 6.1.2 in Appendix 6.1. In the following section we present a simplified gist of the categorization model which was developed surrounding the 8 formerly presented categories: apple, car, cow, cup, dog, horse,

pear and tomato. The model's explanation is gradually completed throughout this chapter, i.e., more details are presented as far as this chapter is read.

This model aims to categorize an object, i.e., to calculate which of the 8 categories in ETH-80 the latter is more similar to. This computation is made using the LEs cortical attributes collected from the object and from all the template objects in several lambdas. Examples of detected LEs in objects images are provided in Appendix 6.2. The first main idea behind this model was to compare an object's LEs to a category's LEs and generate a similarity percentage  $s_p$  which represents the resemblance degree between both. As 8 categories were available in ETH-80 (Leibe & Schiele, 2003), an object could be assigned 8 similarity percentages (later, we mention that in fact 24 similarity percentages are assigned to an object; however, for now, it's important to consider 8 similarity percentages for an easier explanation about the model). Afterwards, these 8 percentages could be analysed in order to decide which category the object is more similar to. For an illustration about this description, see Fig. 3.1.3.



**Figure 3.1.3** - Diagram describing a simplified gist of the categorization model.

It's important to refer that the diagram on Fig. 3.1.3 is just a very simplified illustration provided to explain the model more easily, because this diagram doesn't exactly represent the model; however, this diagram represents very well the fact that in order to obtain the similarity percentages, object's and categories' LEs must be compared, and afterwards, the final category is found; the way these processes occur is explained along this chapter. The LEs of a certain category - or a category's templates set - are originated through the LEs of



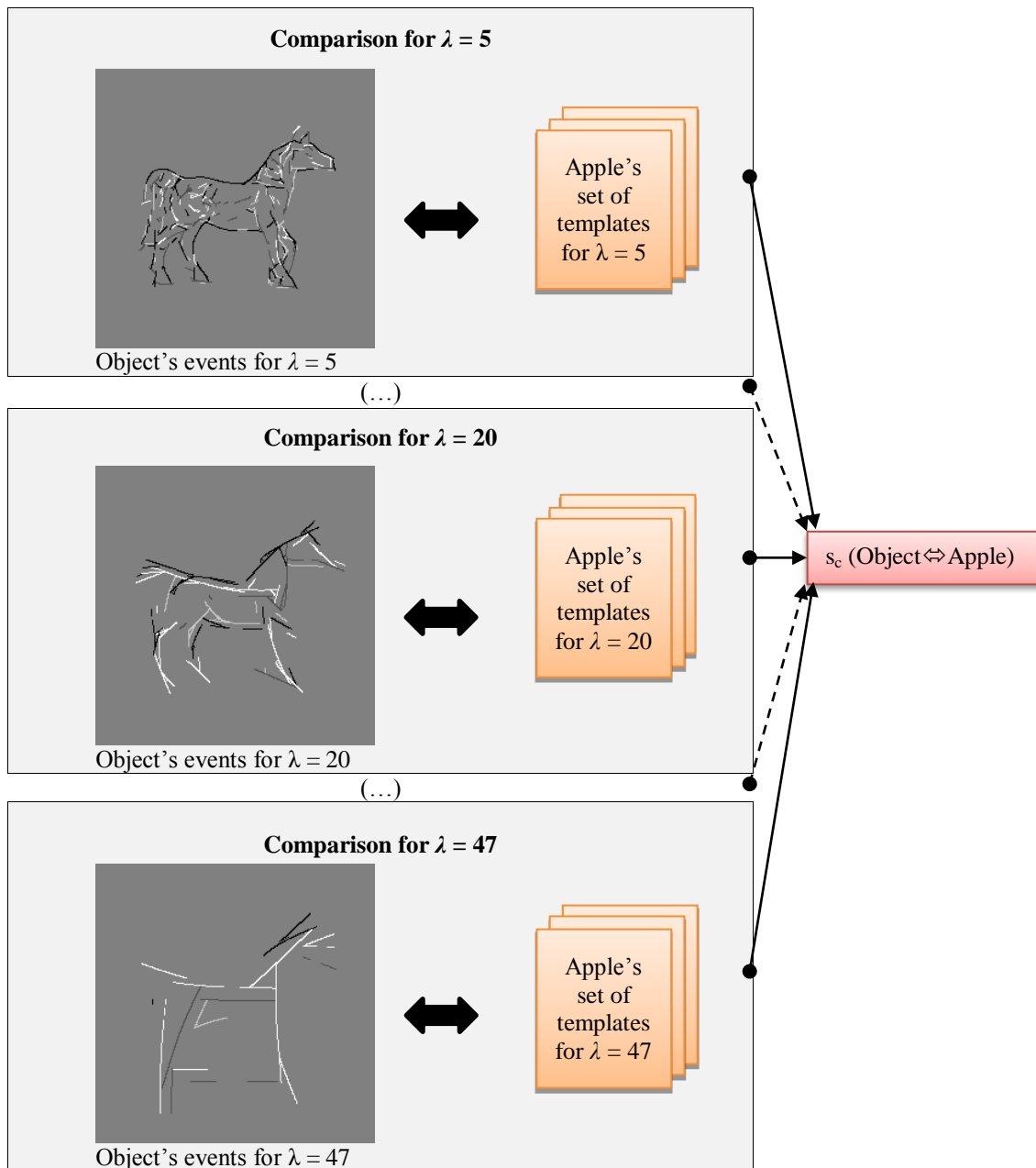
this category's templates objects. The category's LEs can be seen as 'memory' located in pre-frontal cortex. Detailed information about how a category's templates set is built is given in section 3.2. As referred before, the LEs are explored in several scales: the categorization model works with 15 values of  $\lambda$ , in order to give some robustness to the model; this way, the comparison between an object and each category's templates set is made in 15 levels. The  $\lambda$  values are given by

$$\lambda_i = 5 + 3 \times i, \quad i = \{0,1,2, \dots,14\}. \quad (3.1.1)$$

In order to calculate a similarity percentage between an object and a category, it's necessary to calculate a similarity score  $S_c$ , which is the numerator of the similarity percentage. The similarity score is the number of common events found between object and category templates in all these lambdas. Fig. 3.1.4 provides a simplified diagram that shows how the similarity score is obtained.

Analysing the diagram in Fig. 3.1.4, it can be noticed that a comparison between object's LEs and a category's LEs (set of templates) is made at every  $\lambda$ ; as illustrated, all of these comparisons contribute for the final similarity score, and therefore, for the final similarity percentage. As can be seen, the object's events and each category's events are only compared when both belong to the same view and were achieved using the same  $\lambda$ ; when testing an object in left profile, the available categories' template sets for comparison are as well in left profile, as the latter were previously built using the template objects in left profile; if the input test object is in right profile, it's only compared with right profile sets. As it was not the focus to recognize a view automatically, the view is manually chosen while using the algorithm, and, having this data, the algorithm does the rest automatically, i.e., it picks the corresponding categories files that contain the information for comparison. Basically, the model will be applied to each view independently, so that the former is demonstrated to work independently of the input view.

Unknown object:

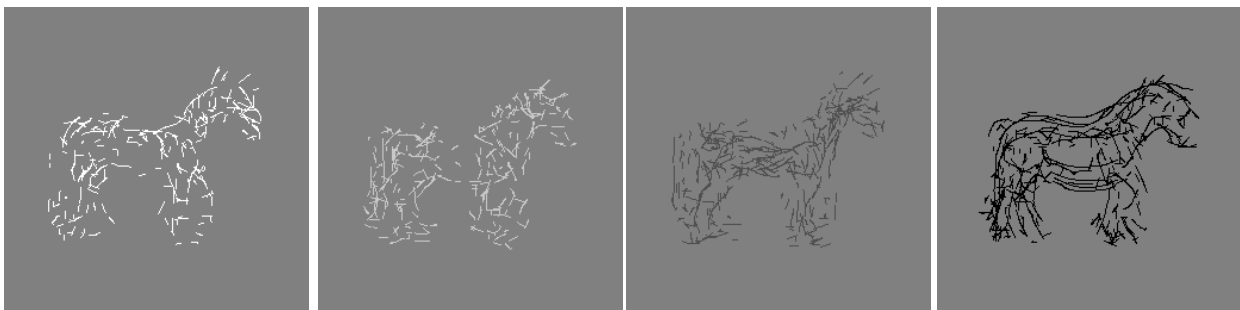


**Figure 3.1.4** – Simplified illustration for a similarity score calculation; the referred score represents the similarity between the ‘unknown object’ and the apple category.

### 3.2 - Obtaining each category's template set

Regarding Fig. 3.1.4, this section explains how to obtain a category's set of templates for a certain  $\lambda$  and view (e.g. Apple's set of templates for  $\lambda = 5$ ). Considering a certain  $\lambda$  and a view, a category's templates set has 4 templates – each one providing different information about the category. Each template is composed by different information from the 5 template objects LEs within a category, view, and  $\lambda$ : (a) the 1<sup>st</sup> template combines all positive responses of line cells, (b) the 2<sup>nd</sup> template combines all the negative responses of line cells, (c) the 3<sup>rd</sup> template combines all positive responses of edge cells, and (d) the 4<sup>th</sup> template combines all negative responses of edge cells, see Fig. 3.2.1.

The combination is done considering each cell response as active. All active cells of the same type and polarity (within the same  $\lambda$ ) contribute to a certain template; active cells of another combination of type/polarity contribute for another template for the category. As there are 4 built templates for each category, view and  $\lambda$ , there is a total of 60 templates per category and view:  $4(\text{templates}) \times 15(\text{scales}) = 60\text{templates} / \text{category and view}$ . The number of categories templates per view can as well be calculated: as 8 categories were available (Leibe & Schiele, 2003), the templates sets for all categories have 480 templates for each view:  $15(\text{scales}) \times 4(\text{templates}) \times 8(\text{categories}) = 480\text{ templates} / \text{view}$ . As an illustration, Fig. 3.2.1 shows for the horse category in right profile the templates used for scale  $\lambda=5$  (from left to right) - the positive lines, the negative lines, the positive edges and the negative edges templates. More category sets examples can be found in Appendix 6.3.



**Figure 3.2.1** - Templates set for the horse category in right profile, for  $\lambda = 5$  - from left to right: positive lines, negative lines, positive edges, negative edges templates.

### 3.3 - Calculating the similarity percentages between an object and a category

As mentioned in section 3.1, we consider, for now, that an object is assigned one similarity percentage for each of the 8 categories (we will see later that in fact an object is assigned 3 different similarity percentages for a single category). This section describes several conceptualized ways for calculating a similarity percentage between an object and a category, hence explaining as well how to obtain the similarity score and other parameters.

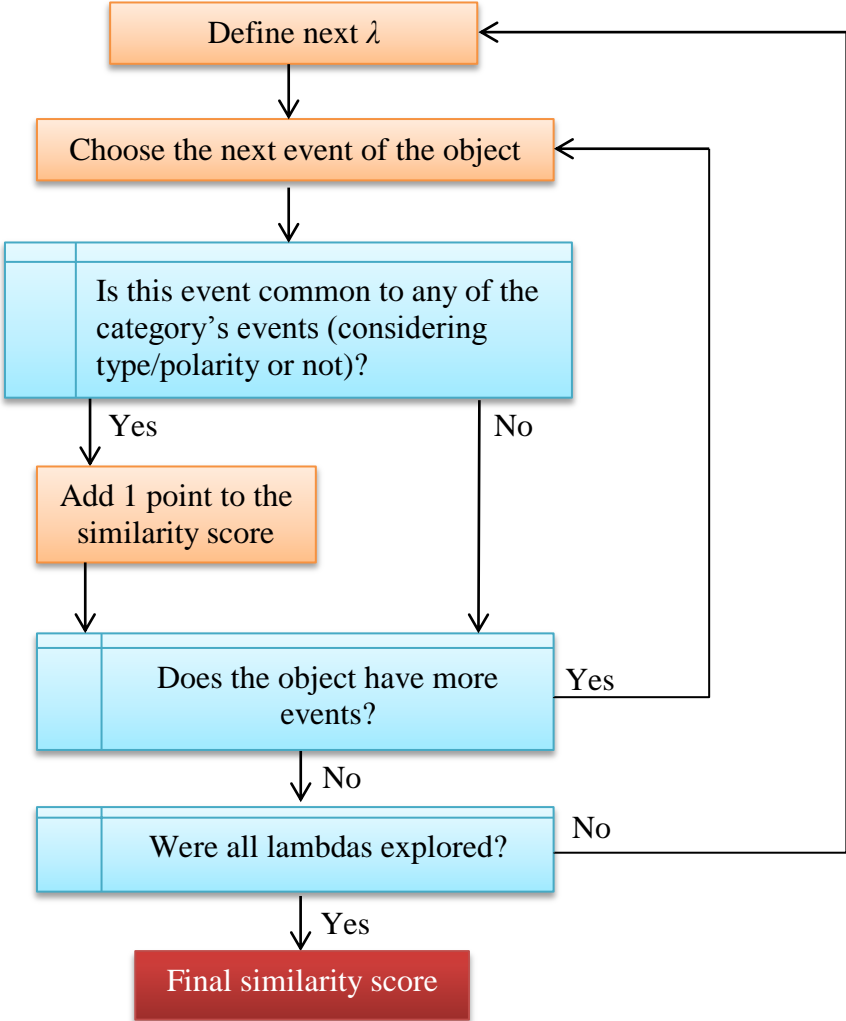
Independently of the way through which a similarity percentage is calculated, the number of common events that exist between object and category's templates is important to achieve. In order to calculate a similarity score, several comparisons are made between the object's events and the templates sets' events for each  $\lambda$  (object and category's templates should have the same view); all these comparisons contribute to the referred integer number (Fig. 3.1.4). So, considering a certain  $\lambda$ , this calculation is done choosing every object's event, and testing if each event exists in a certain region of interest (RoI) of the (previously prepared) category's templates; the RoI is defined by a circular dendritic field whose centre has equal coordinates to the searched event of the object.

The calculation of the common events number is based on accumulation of evidence: if the type and polarity of the event is not considered, if at least one event (line or edge cell response) is found inside this area in any of the category's templates, this category's score is increased by 1 - a common event was found. However, if the events type and polarity are considered, in order for the category to score 1 point this event has to be found within the RoI localized in the category's template correspondent to the same type and polarity. For the latter case, similar line and edge cells are tested in terms of being firing (or not) in the same scale in the object and template within the same RF. If the score is increased by 1, it is said that there's one common event between the category and the object; also, a certain category scores a maximum of 1 point per searched event.

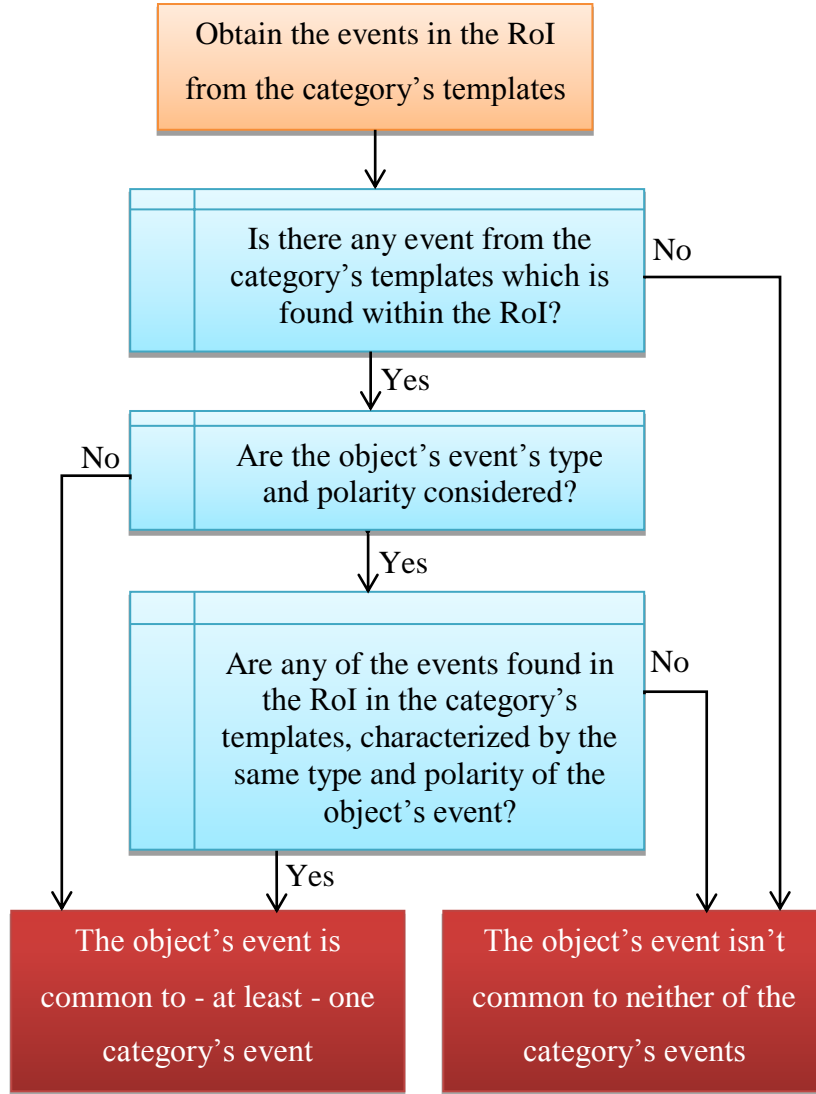
Consider Fig. 3.3.1 which presents a diagram showing the way the categorization model calculates the similarity score between an object and a category, and also Fig. 3.3.2 which presents a diagram showing the way the categorization model decides if an object's event – in a certain  $\lambda$  – is found common to at least one category's event. As can be seen, while testing if an object's event is common to any of a category's (within a certain  $\lambda$ ), if no event is found in the RoI of the category's templates, the category doesn't score; if one or more events exist in

the RoI of one or more category templates, and the type and polarity are considered, the category scores if at least one of these events has the same type and polarity of the object's event; if one or more events are found within the RoI of one or several category templates and the type and polarity are not considered, the category scores 1 point, no matter the type and polarity of the object's event and of the category templates' events in the RoI.

Another illustration is provided in Fig. 3.3.3, where a positive line is searched within a circular dendritic field (delineated in dark red) in the positive lines apples' template for  $\lambda=5$  (see below for the details about the radius of the circular dendritic field function deduction). At the top, the event is found and the apples category scores 1 point (there's 1 common event), and at the bottom the event is not found and the apples category doesn't score.



**Figure 3.3.1** – Diagram presenting the way the categorization model calculates the similarity score between an object and a category.



**Figure 3.3.2** - Diagram presenting the way the categorization model decides if an object's event – in a certain  $\lambda$  – is found common to at least one category's event.

As mentioned before, in order to attain the similarity score between an object and a certain category – or the number of common events -, several comparisons are made between the object's events and the category's template set for each  $\lambda$ . This score is achieved by summing all the common events found, considering fifteen values of  $\lambda$ ,

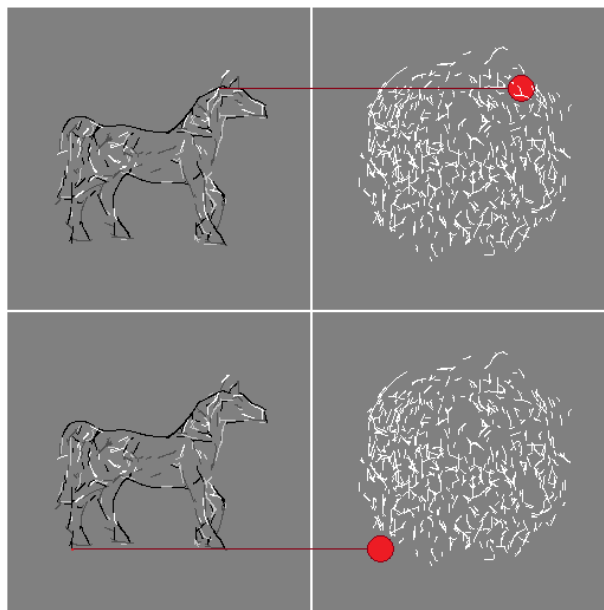
$$S_c(e_{t,p}) = \sum_{\lambda=5}^{\lambda=47} N_{ce}(\lambda, e_{t,p}), \quad (3.3.1)$$

being  $S_c$  the similarity score and  $N_{ce}$  the number of common events. The number of common events, which depends on  $\lambda$ , also depends if the events' type and polarity  $e_{t,p}$  should be considered ( $e_{t,p} = ON$ ) or not ( $e_{t,p} = OFF$ ). As mentioned before, considering the events type and polarity ( $e_{t,p} = ON$ ), for an object's event to be common to a category's, this event

not only needs to exist inside the previously mentioned RoI, but also to be of the same type and polarity. The score will, logically, tend to be lower in this case than in the situation when the event type and polarity are not considered in order to score ( $e_{t,p} = OFF$ ). Concluding, two ways for calculating the number of common events between an object and a category - or similarity score  $S_c(e_{t,p})$  – were found:  $S_c(ON)$  and  $S_c(OFF)$ .

As mentioned before, the final categorization is made using similarity percentages, and not just scores: a decision that is explained at this point. Imagine that a certain object<sub>A</sub> has 25.000 events, and a certain object<sub>B</sub> has 30.000, and that these objects are compared to category<sub>X</sub>. Consider that the similarity score  $S_c(\text{object}_A \leftrightarrow \text{category}_X)$  was 25.000, and the same number was achieved for the  $S_c(\text{object}_B \leftrightarrow \text{category}_X)$  score. Considering that the object<sub>A</sub>'s events were all common to those of category<sub>X</sub>, this object is more entitled to category<sub>X</sub> than object<sub>B</sub> whose common events found were  $25.000/30.000 \approx 83\%$  of its own events. Following this logic, the score ( $N_{ce}$ ) is important when compared to the number of events of the object in question. Hence, a new parameter was introduced: object reference ( $O_R$ ), which is calculated through the sum of all objects events along all considered values of  $\lambda$ ,

$$O_R = \sum_{\lambda=5}^{\lambda=47} N_{eo}(\lambda). \quad (3.3.2)$$



**Figure 3.3.3** - At the left a horse's LEs ( $\lambda=5$ ) are shown; a positive line cell response is searched within a RoI (circular dendritic field) in the positive line apples template. At the top: the event is found and the apples category scores 1 point; at the bottom: the event is not found and the apples category doesn't score.

Referring to Eq. 3.3.2 which calculates the object reference  $O_R$ ,  $N_{eo}$  is the number of an object's events for a certain  $\lambda$ . The object reference was decisive to find a first similarity percentage formula ( $S_p$ ),

$$S_p(e_{t,p}) = \frac{S_c(e_{t,p})}{O_R}. \quad (3.3.3)$$

Recalling that the  $S_c$  represents two options -  $S_c(ON)$  and  $S_c(OFF)$  -, Eq. 3.3.3 actually represents two similarity percentages:  $S_{p1} = S_p(ON)$  considering the events' type/polarity, and  $S_{p2} = S_p(OFF)$  not considering the events' type/polarity. Consider, this way, Eqs. 3.3.4 and 3.3.5 which represent the two first conceptualized ways of calculating a similarity percentage between an object and a category,

$$S_{p1} = \frac{S_c(ON)}{O_R}; \quad (3.3.4)$$

$$S_{p2} = \frac{S_c(OFF)}{O_R}. \quad (3.3.5)$$

Like  $O_R$ , another reference emerged. The previous equations represent the percentage of common events found among the object events; in the same logic, it was thought that it would be interesting to use the percentage of common events found among the evaluated category's events, i.e., the ratio of events in a category that were found common to the objects events. Accordingly, a category reference ( $C_R$ ) is calculated by summing all the events found in all 4 category templates (for all lambdas) without repetition, i.e.: considering a certain  $\lambda$ , and the 256x256 size of the category templates, a single pixel counts as 1 point if an event is found in this pixel in one or more category templates; this sum for a certain lambda is added the other sums of the remaining lambdas. Consider, this way, Eq. 3.3.6 that was used to calculate a category's reference.

$$C_R = \sum_{\lambda=5}^{\lambda=47} N_p(\lambda) \quad (3.3.6)$$

Referring to Eq. 3.3.6,  $N_p(\lambda)$  is the number of events found in one or more category templates within a certain  $\lambda$ ; as an example to clarify this parameter, if (for a certain  $\lambda$ ) each of the 256x256 pixels was found to correspond to one to four events (four category templates), the value of  $N_p(\lambda)$  would be 256x256. With the new reference -  $C_R$  - two other similarity percentages equations emerge:



$$S_{p3} = \frac{S_c(ON)}{C_R}, \quad (3.3.7)$$

and

$$S_{p4} = \frac{S_c(OFF)}{C_R}. \quad (3.3.8)$$

$S_{p1}$ ,  $S_{p2}$ ,  $S_{p3}$  and  $S_{p4}$  – the four similarity measures – are called, correspondingly, the similarity percentage for concept 1, concept 2, concept 3 and concept 4. Summarizing, four concepts were considered:

- **Concept 1** - The similarity percentage between an object and a category is calculated as the quotient between the similarity score and the sum of the number of object's events in all  $\lambda$ . The similarity score is calculated considering the events' type and polarity ( $S_{p1}$ ).
- **Concept 2** –The similarity percentage between an object and a category is calculated as in concept 1, with the exception that the similarity score is calculated not considering the events' type nor polarity ( $S_{p2}$ ).
- **Concept 3** - The similarity percentage between an object and a category is calculated as the quotient between the similarity score and the category reference. The similarity score is calculated considering the events' type and polarity ( $S_{p3}$ ).
- **Concept 4** - The similarity percentage between an object and a category is calculated as in concept 3, with the exception that the similarity score is calculated not considering the events' type nor polarity ( $S_{p4}$ ).

At this point another parameter needed to be defined: the radius of the dendritic field (RoI). The latter parameter was defined as a function of  $\lambda$ . Considering  $e_{t,p} = OFF$ , for a category to score 1 point when compared to an object, it means that what happened was while searching a certain object's event within a RoI in the category's set's templates, this event was found in one or more templates. This area is circular and centred in a pixel of a category's set's templates, which has the same coordinates as the searched object's event. The radius of the circle depends on the  $\lambda$  that was used. This radius is called dendritic field radius, as the event is being searched in a circular area which is a group of pixels, and not just in the analogous pixel.

The dendritic field radius, being equal to the number of pixels, and, being also a function of  $\lambda$ , was conceptualized as a linear function, with integer values, starting from the value 3 (pixel) for  $\lambda=5$ , and reaching the value 10 (pixel) for  $\lambda=47$ , being this calculated empirically. So, the dendritic field radius function was calculated, starting from the general formula of a linear function (3.3.9), in which  $m$  is the slope and  $b$  the ordinate of the linear function that crosses the vertical axis,

$$r(\lambda) = m \times \lambda + b. \quad (3.3.9)$$

Considering the two points mentioned before, whose coordinates were (5, 3) and (47, 10), the variables  $m=1/6$  and  $b=13/6$  were calculated through a two equations system. Given the fact that the radius unity is given in pixel (an integer number), the radius was calculated using the round half up rule (“int”),

$$r(\lambda_i) = (\text{int}) \left( \frac{1}{6} \times \lambda_i + \frac{13}{6} \right), \text{ with } \lambda_i = \{5,8,11, \dots, 47\}. \quad (3.3.10)$$

As all the parameters were defined, the four concepts were tested separately, using the template objects images chosen from ETH-80 (Leibe & Schiele, 2003) for training and testing the model. It was considered that the final category attributed to an object is simply the one which had the highest similarity percentage. The results of these experiments were helpful to decide which concepts are the best, and how to use these similarity percentages to categorize; the following parameters were used: (a)  $r(\lambda_i)$ , (b)  $\lambda_i = \{5,8,11, \dots, 47\}$ , (c) the tested objects were the template objects: 40 objects in left and right profile (80 images) and (d) the same template objects on left and right profile were used to build the categories templates sets.

The tests returned that the first three concepts ( $S_{p1}, S_{p2}$  and  $S_{p3}$ ) - when testing individual templates objects - were 100% successful, and the 4<sup>th</sup> ( $S_{p4}$ ) had most objects well categorized (80%). Further information about the similarity percentages between each object and each category is presented in detail in Appendix 6.4, in the Tables 6.4.1 to 6.4.5. As predicted before, the similarity percentages are lower when, using the same reference, the events’ type and polarity are considered. This can be realized when comparing Tables 6.4.2 and 6.4.3 in Appendix 6.4, which respectively correspond to concept 1 and concept 2. It was also noticed in each object’s similarity percentages that the latter follow a ‘similarity percentages pattern’ depending on the object’s true category; as an example, an apple always has a higher similarity percentage for the tomato category rather than the car category, as a cow has a

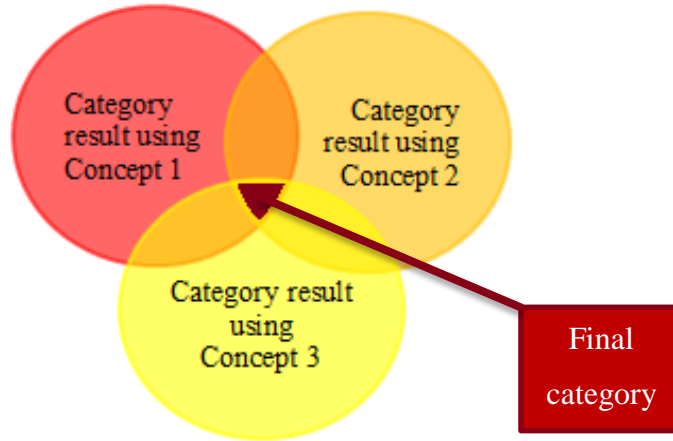
higher similarity percentage for the horse category rather than the pear category. Considering the 100% success rate of concept 1, concept 2 and concept 3, and the 80% result achieved for the 4<sup>th</sup> concept, the next step is to combine the best concepts in order to build a robust model.

### 3.4 - Final categorization model

Choosing one of the first three concepts to finally categorize the test objects is an option, because, theoretically, all had a 100% categorization rate. But, what if one of the three concepts is the best since the proof of concept will be made using test objects? There's a chance that this will happen, and, at this point, there's no evidence of one concept being better than the other two; the only evidence is that the 4<sup>th</sup> concept had an 80% result, which is lower than 100%, and therefore it was not used.

Continuing the deliberation, choosing one of the concepts is, metaphorically, taking a step blindfolded. With this consideration, another option emerged: what if all the three concepts were used? This way all of the three could be taken advantage of. In this regard, Fig. 3.4.1 illustrates the main idea of the proposed model: intersecting the three concepts' answers to obtain the final category and provide robustness to the model. Thus, 24 similarity percentages are calculated for each object, i.e., there are 3 similarity percentages calculated to define the resemblance between an object and a category. So, if concept 1 ( $S_{p1}$ ) says 'it's an apple', if concept 2 ( $S_{p2}$ ) says 'it's an apple', and if concept 3 ( $S_{p3}$ ) says 'it's an apple', the three agree and that is the final answer. Also, following the previous example, in the eventuality of concept 3 ( $S_{p3}$ ) having the highest similarity percentages linked to the "apple" and "tomato" categories, the final intersection would, as well, be "apple."

If the result of the intersection is considered as the final category result, this will, supposedly, build a more robust model. An analogy that can be done is the fact that if it is asked to one person: "which category do you think it belongs to?" and the person answers "I'm certain it's a cow", one will be almost convinced because of this person's certainty. But, if three people answer saying "I'm certain it's a cow", the certainty degree of being a cow is increased, because, not one, but three people said the same thing, with a very good level of conviction. Hence, the intersection of the three concepts presents itself as a good option.



**Figure 3.4.1** - The model’s main principle: intersecting the categories answers of concept 1, concept 2 and concept 3, in order to find a common agreed category for classifying an object.

However, it’s possible that a basic intersection will have an empty answer; therefore, the final category answer was not calculated only as a simple intersection. As an example, concept 1 and 2 would say “it’s an apple” but concept 3 would say “it’s a tomato”. One can ask “why would concept 3 say it’s a tomato if the odds of the category result to be right are 100%?”; well, given the fact that test objects are not equal to the template objects, it’s possible that they will have different similarity percentages; following the same thought, there’s the possibility that the highest percentage will not correspond to the true category. Even though the probability of this to happen is not known, it’s still possible, and, given that assumption, the three concepts may disagree in the final answer, and a basic intersection will just originate an empty answer in this case - and that’s not what’s intended. As the previous supposition takes its place, a new variable was defined: the threshold percentage ( $T_p$ ).

The threshold percentage was introduced in order to define a margin of error around the maximum similarity percentage achieved for an object within a concept. To better understand this parameter’s usage, let’s consider the example given in Table 3.4.1 which presents the similarity percentages for a hypothetical object ( $object_x$ ) which is an apple.

**Table 3.4.1** - Hypothetical similarity percentages between an object ( $object_x$ ) and the 8 categories in ETH-80, using concept 1, 2 and 3.

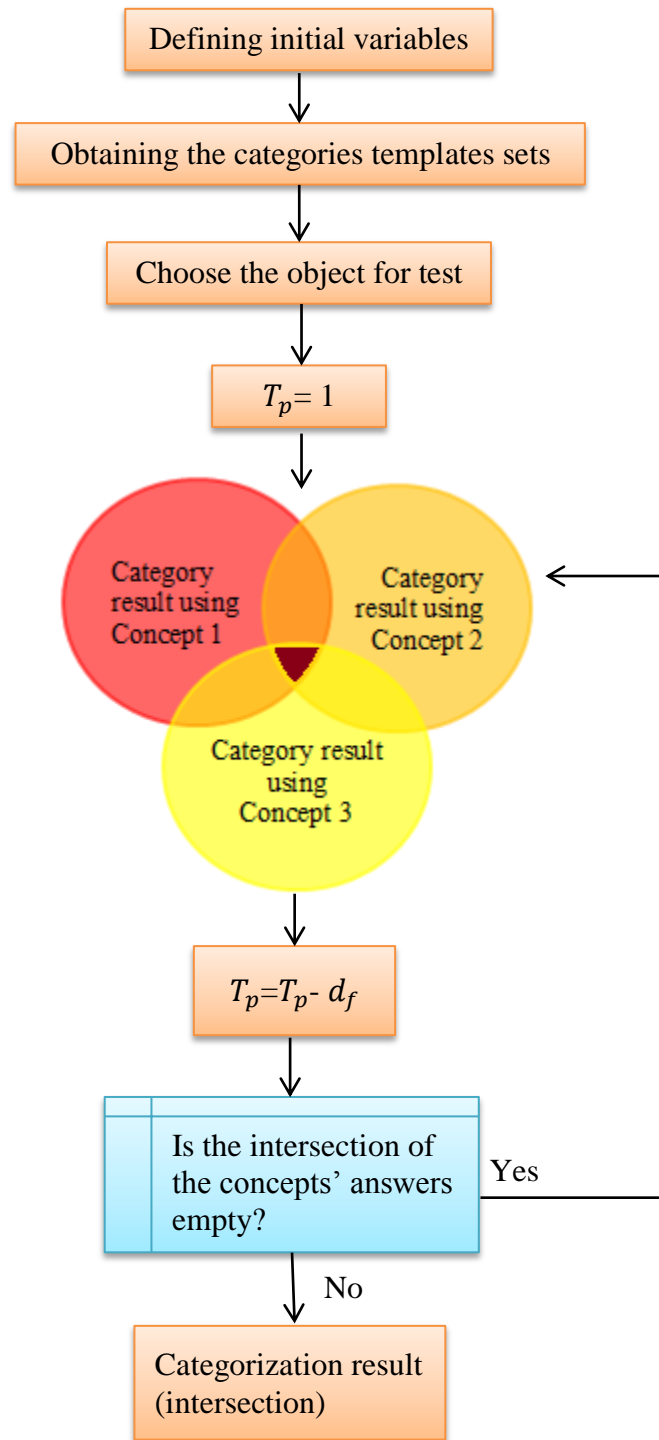
[%]	Apple	Car	Cow	Cup	Dog	Horse	Pear	Tomato
Concept 1 ( $S_{p1}$ )	75	25	30	50	35	30	60	76
Concept 2 ( $S_{p2}$ )	71	20	25	45	30	25	55	70
Concept 3 ( $S_{p3}$ )	56	5	10	30	15	10	40	50

Table 3.4.1 shows that this object, which is an apple, is - at the moment - being categorized as a tomato (76%) regarding concept 1 ( $S_{p1}$ ), and as an apple regarding concept 2 ( $S_{p2}$ ) and 3 ( $S_{p3}$ ), because these are the categories which have the highest percentage within each concept; the final category would be an empty answer if a simple intersection was applied. However, the intersection wouldn't be null if a margin of error was defined. For this, we defined the minimum percentage or threshold ( $T$ ), which is achieved by multiplying the maximum similarity percentage by the threshold percentage. It should be noticed that although the threshold percentage is a constant, the threshold  $T$  is object dependent, as Eq. 3.4.1 suggests:

$$T_p = \frac{T}{\text{MAX}(s_p)}. \quad (3.4.1)$$

Regarding the example in Table 3.4.1, in order for the apple to be included in concept 1's answer ( $S_{p1}$ ), this threshold percentage ( $T_{p1}$ ) should have the value of  $T_{p1} = \frac{75}{76} \approx 0.987$ . Considering two decimal places, for this example,  $T_{p1}$  is rounded down to 0.98; using this threshold percentage in the three concepts, the two first concept's final answer would be "apple or tomato" as  $76 \times 0.98 < 75$  and  $71 \times 0.98 < 70$ ; also, the third concept's final answer would be "apple", and therefore, the final intersection would also be "apple" which is the true category. Given the logic presented in this example, the proposed solution to achieve a non-empty answer is shown in Fig. 3.4.2. In order for an object to be categorized, the three concepts have to agree, at least, in a common category. In the meantime, each time the intersection is null, the threshold percentage is diminished by a decreasing factor  $d_f$ ; this will tangentially increase the number of categories in one or more concept's answers, providing, at one point, a categories common ground. Furthermore, the initial threshold percentage is 1, and the used decreasing factor ( $d_f$ ) was 0.01.

As mentioned before, Fig. 3.4.2 shows a solution for the problem in which the three concept's intersection was empty. In an initial iteration, if the three concepts don't agree in a common category when  $T_p$  is 1, the next iteration is done with a threshold percentage equal to 0.99 for the 3 concepts, and, if this doesn't solve, 0.98 is used, etc., i.e., the threshold percentage is decreased in 1 percentual point (pp) each time, until the three concepts agree in a categories common ground. The ideal situation is that the intersection is just one final category, and that this category is the true one; however, it's possible to obtain more than one common category in the intersection result – an issue that will be solved next.



**Figure 3.4.2** – Solution presented to the problem in which the three concept’s intersection was empty. The diagram considers only one object for test.

Imagining now that the categories resulting from the intersection are more than one, let’s consider Table 3.4.2 which shows some hypothetical cases in this regard. In case 1, cow, horse and dog (originated from the intersection) have different similarity percentages for each

concept. In order to select the winning category, one might be lead to the conclusion that the category with the highest similarity percentages on the three concepts is the winning category: that's the example represented in case 1; by selecting the category with the best percentage within concept 1, cow would be the winner with a percentage of 95%, as well as in concept 2 with 85% and in concept 3 with 90%. In this case, finding the category with the best score in each concept is enough to find a winning category.

**Table 3.4.2** – Examples showing the similarity percentages between a hypothetical object and the categories which originated from the intersection.

<b>Case 1</b>			
	Concept 1 ( $S_{p1}$ )	Concept 2 ( $S_{p2}$ )	Concept 3 ( $S_{p3}$ )
Cow	95%	85%	90%
Horse	80%	75%	65%
Dog	50%	55%	53%
<b>Case 2</b>			
	Concept 1 ( $S_{p1}$ )	Concept 2 ( $S_{p2}$ )	Concept 3 ( $S_{p3}$ )
Cow	80%	85%	90%
Horse	80%	75%	65%
Dog	50%	55%	53%
<b>Case 3</b>			
	Concept 1 ( $S_{p1}$ )	Concept 2 ( $S_{p2}$ )	Concept 3 ( $S_{p3}$ )
Cow	95%	85%	90%
Horse	95%	95%	65%
Dog	50%	95%	53%
<b>Case 4</b>			
	Concept 1 ( $S_{p1}$ )	Concept 2 ( $S_{p2}$ )	Concept 3 ( $S_{p3}$ )
Cow	95%	85%	90%
Horse	80%	90%	65%
Dog	50%	55%	96%
Car	49%	53%	63%
<b>Case 5</b>			
	Concept 1 ( $S_{p1}$ )	Concept 2 ( $S_{p2}$ )	Concept 3 ( $S_{p3}$ )
Cow	80%	70%	60%
Horse	80%	70%	60%
Dog	65%	55%	50%

However, there are other case scenarios. Exploring other possibilities, what if there's a tie in one of the concepts, having cow and horse the same percentage? For this, consider case 2 in Table 3.4.2. Applying the same logic as in case 1, the winning category should be found in each concept; in concept 1 cow and horse are both winners (80%), in concept 2 cow wins (85%), and in concept 3 cow wins again (90%). The tie between cow and horse in concept 1 doesn't allow applying the same logic as before, because no category is a single-winner for the three concepts; on the other hand this example allows moving on to a more flexible rule: as the concepts-score (the number of wins in a category) has a maximum value of 3 (3

concepts), the category whose concepts-score is equal to 2 or 3 is the winner. Therefore, in case 2 cow wins in 3 concepts, horse wins in 1 and dog in 0, being cow the winning category.

Case 3 in Table 3.4.2 has a special challenge for the latter rule: cow wins in 2 concepts (1 and 3), horse wins also in 2 concepts (1 and 2) and dog wins in concept 2; even if dog is put aside by having the lowest concepts-score, cow and horse have both a concepts-score equal to 2. For this case, a solution was found: for each category the average percentage between concept 1 and 2 can be found,

$$A = \frac{S_{p1} + S_{p2}}{2}. \quad (3.4.2)$$

Therefore, choosing the categories with the best concepts-score – cow and horse – their averages can be calculated:  $A_{cow} = \frac{95\% + 85\%}{2} = 90\%$  and  $A_{horse} = \frac{95\% + 95\%}{2} = 95\%$ .

These two categories were then untied, as the horse's average is higher than the cow's, being horse the final winning category. Another example for a tie is shown in case 4 where cow wins in concept 1, horse wins in concept 2, dog wins in concept 3, and car wins in none: there's no category with concepts-score equal to 2 or 3. For this case, the solution found will be the same theorized for case 3: cow, horse and dog are chosen as they had the highest concepts-score, and then their average percentage between  $S_{p1}$  and  $S_{p2}$  is calculated. The average percentages achieved were 90% for cow, 85% for horse and 52.5% for dog, thus cow is the winning category in case 4.

Case 5 shows the last presented challenge: if two categories - which resulted from the intersection – have exactly the same concepts-score, and also the same average percentage (75%), how can the final category be obtained? The immediate solution found is choosing the category which had the highest similarity percentage in concept 3; however, in the presented case the latter percentages are equal, therefore the winning category can't be obtained through that way. Therefore, the final category is proposed to be chosen by selecting which category had the highest similarity score when considering the events type and polarity,  $S_c(ON)$ . The similarity score between an object and a category is a guiding line; furthermore, this quantity is unique for each category in this situation. To prove the latter statement, let's consider Eq. 3.3.7 which allows calculating the similarity percentage for concept 3,  $S_{p3}$ ; according to Eq. 3.3.7, if the  $S_{p3}$  percentages are equal for all categories (Table 3.4.2, case 5) while the category references are all different (see Appendix 6.4, Table 6.4.5), it's certain that the

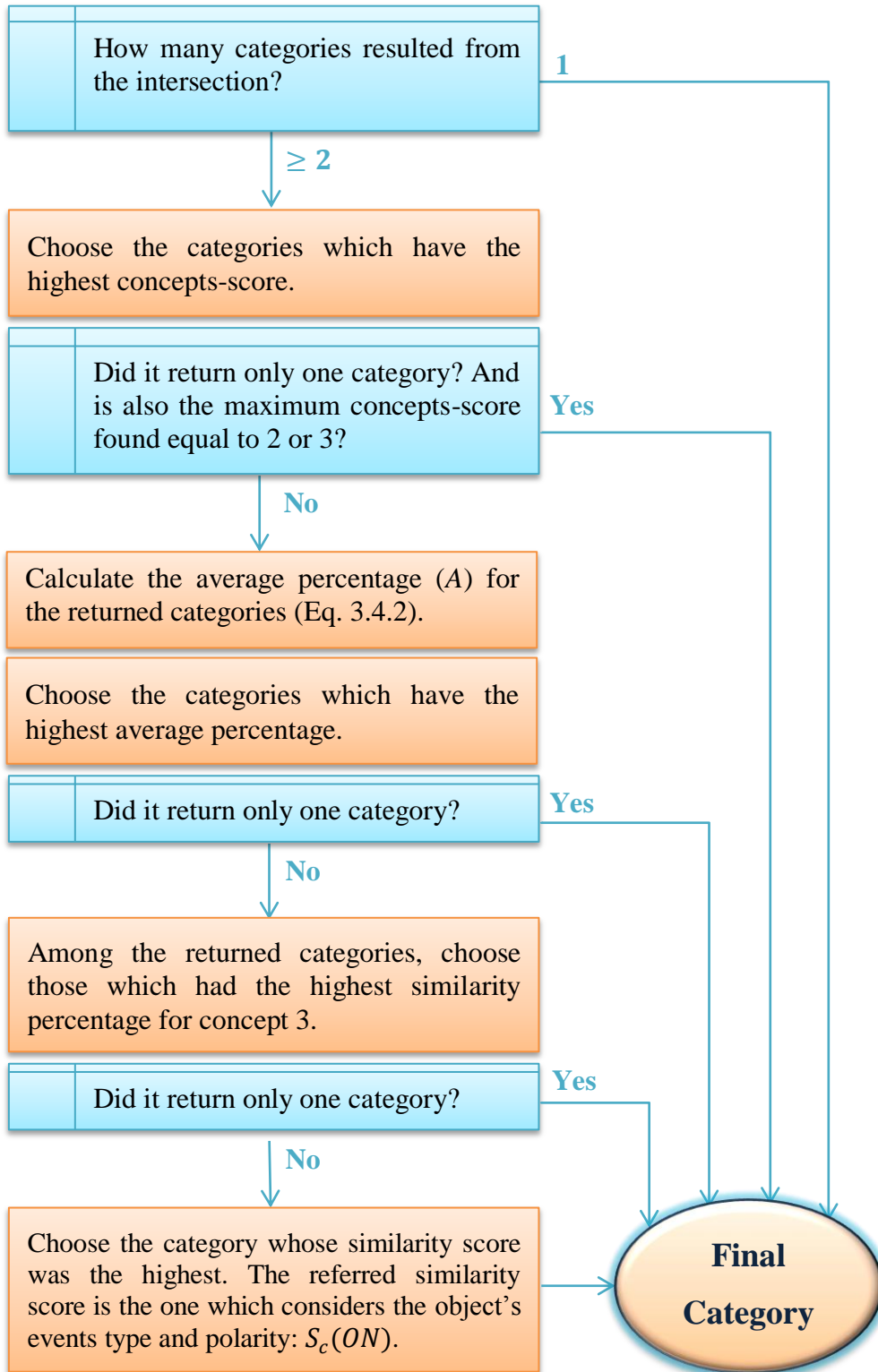


similarity score is different, allowing definitely to choose a final winning category. The situation considered in case 5 - where several categories have the same percentage in the three concepts - is highly unlikely considering the results achieved with the template objects, presented in Appendix 6.4, where the similarity percentages tend to be all different; either way, this case was already predicted and solved.

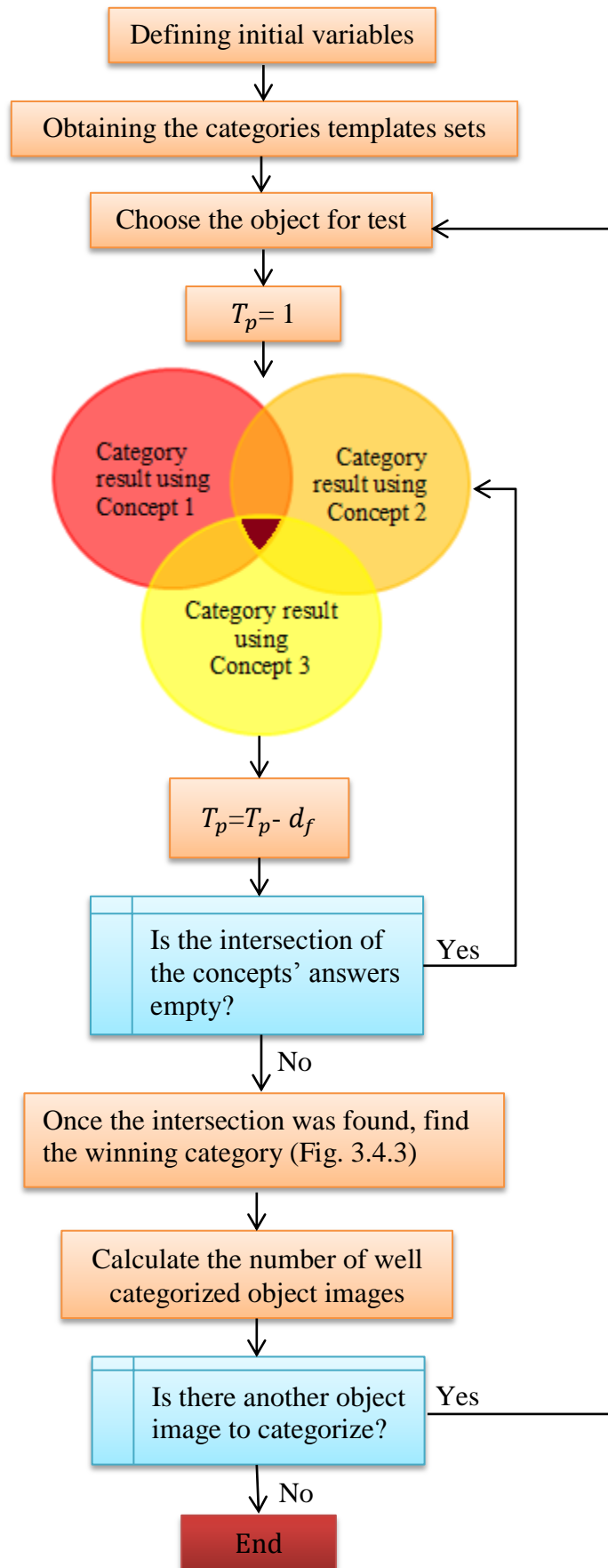
The previously presented cases from Table 3.4.2 consisted in eventual situations that could be faced while in the categorization process – cases which now have a solution. The solutions previously discussed are shown in Fig. 3.4.3 in a compendium-diagram which describes how the final category is chosen once a common ground (one or more categories) is found among the three concepts. Moreover, Fig. 3.4.4 presents a more general diagram which summarizes the way a final category is chosen for one or several objects: essentially, after finding a non-empty intersection between the 3 concepts' answers through the diagram on Fig. 3.4.2, the final category is found through the solution shown on the diagram in Fig. 3.4.3.

The model built in this section was tested using the template objects. The following parameters were used: (a)  $r(\lambda_i)$ , (b)  $\lambda_i = \{5,8,11, \dots, 47\}$ , (c) the tested objects were the template objects: 40 objects in left and right profile (80 images) and (d) the same template objects on left and right profile were used to build the categories templates sets. A final categorization rate of 100% was obtained; this was expected, considering the already referred similarity percentages obtained from the template objects, which are presented in Appendix 6.4; i.e., for each object, in either concept, the highest similarity percentage found was unique and linked to the true object's category, thus the three concepts' intersection was always the true category.

Next, the model was evaluated using the test objects. In chapter 4, a description of this evaluation is provided, as well as the final results of the performed tests.



**Figure 3.4.3** - Diagram that illustrates how to find the winning category using the similarity percentages, once a non-empty intersection among concepts' answers was found.



**Figure 3.4.4** – General diagram of the model, which summarizes the way a final category is chosen for one or several objects.

# 4. Tests and Results

---

## ABSTRACT

**This chapter describes an evaluation made to the categorization model: several tests were performed in order to assess it by comparison to other models' results and also further tests took place to test the model in terms of rotation, noise and scale invariance.**

---

In order to complete the proof of concept, the model was tested using 80 test objects images, which correspond to 40 objects shown in left and right profile. Furthermore, other tests took place so that the model could be evaluated in terms of rotation, noise and scale invariance.

### 4.1 – Comparing the categorization model's results to other models' results

Summarily, 75 of the 80 images were correctly categorized, representing a 93.75% success rate (see Table 6.4.6 in Appendix 6.4, which shows the test objects references, Table 6.4.7 which presents the final category results for each of the 80 tested objects, and also Table 6.4.8 which summarizes the mentioned results). Just 5 of the objects images were 'miscategorized'; however, the false categories attributed were relatively comparable to the true ones: apple10 (left profile) was mistaken for being a tomato, and dog4 and dog9 (both views) were mistaken for being horses. The apple10's error (left profile) was considered minor, because tomatoes are round as apples, and, we only analyse LEs, and not colour nor texture; in this case "tomato" had the highest similarity percentage in the three concepts, validating the assumption made in section 3.4: the test objects are different than template objects, and there's a chance that the highest similarity percentage is not connected to the true category. This also happened, for example, with dog4. As referred by Rodrigues & du Buf, 2009b, the animals are relatively similar regarding their heads, necks and tails, therefore, the dog/horse error was considered as well a minor one. In a view perspective, 95% of the 40 objects in right profile were correctly categorized, as were 92.5% of the 40 objects in left profile (see Table 6.4.8 in Appendix 6.4). Moreover, considering a category perspective, 90% of the apples and 60% of the dogs were well categorized; also, positively, the majority of the categories was 100% successfully categorized: car, cow, cup horse pear and tomato.

Considering the performed tests and the diagram shown in Fig. 3.4.3, when the algorithm reached the point ‘How many categories resulted from the intersection?’, the answer was always ‘1’. The diagram – which was available in the algorithm – just needed technically, for these specific tests, to return the intersection of the categories because all the intersections comprised just one category in each of the 80 cases. At this point it was thought that the entire diagram just comprised a good plan B, but later in this chapter we discuss that indeed it was necessary.

Other publications where the ETH-80 database (Leibe & Schiele, 2003) was also used for categorization tests were Rodrigues & du Buf, 2009b, Han et al., 2011 and Leibe & Schiele, 2003 (see Chapter 2). The first publication comprises the best comparison link to this thesis work, as Rodrigues & du Buf, 2009b also used the LEs cortical attributes (V1) detection model as a basis for their categorization approaches. In order to test their categorization model, Rodrigues & du Buf, 2009b selected 10 images from each 8 categories: a total of 80 images. They explored a two-level categorization approach: ‘pre-categorization’ and ‘categorization’. The ‘categorization’ test set was more similar to this thesis’ experimental setup, because (a) on pre-categorization, LEs contour templates were used (differently from the performed tests in this thesis), (b) the pre-categorization level just divides between animal, fruit, cup or car, and, in our tests we distinguished among all categories similarly to what happens in the ‘categorization’ level of Rodrigues & du Buf, 2009b, and (c) on pre-categorization only coarse scales were used, whereas on the ‘categorization’ level of Rodrigues & du Buf, 2009b, coarse and fine scales were used – similarly to our tests; however, while we used 15 scales, Rodrigues & du Buf, 2009b used 8. Similarly to this thesis experimental setup, in Rodrigues & du Buf, 2009b 40 objects were chosen for training (for the ‘half’ test); however, while 40 objects images per view were used for test in this thesis, Rodrigues & du Buf, 2009b used all 80 images from 80 objects in the same view for this purpose. Another comparison point is the fact that while the background of the test/training object images used in this thesis was previously homogenized, Rodrigues & du Buf, 2009b used the images directly from ETH-80. Here, the obtained categorization rate considering two views was 93.75%, while Rodrigues & du Buf, 2009b achieved for the ‘half’ categorization test scenario a categorization rate of about 88.75% (71/80). Also, while the same authors mentioned that “typical miscategorizations were dog/cow, horse/dog, horse/cow and apple/tomato” (Rodrigues & du Buf, 2009b), the categorization model in this thesis was able to successfully distinguish among almost all categories (regarding the test described in this

section) - the exception resided in the dog/horse and apple/tomato miscategorizations. Finally, the test conditions used for both models weren't exactly the same; however, the average rate obtained in this thesis is about 5 percentage points (pp) higher than the one obtained by Rodrigues & du Buf, 2009b.

Leibe & Schiele, 2003 proved the potential of resorting to multiple features in categorization, by using a decision tree which at each level bases its decisions on one cue only. In order to compare the results obtained in this thesis to the ones from Leibe & Schiele, 2003, both experimental setups were evaluated. The background of the images used in this thesis was manually homogenized, which is an advantage for the significance of the results of Leibe & Schiele, 2003, once these authors categorized unknown objects with "a near-perfect figure-ground segmentation" (Leibe & Schiele, 2003). Leibe & Schiele, 2003 presented a more accurate result regarding the ETH-80 database, once they tested all 41 views from each test object and the 93.75% categorization rate obtained in these thesis' tests corresponds to testing 2 views per test object; however, while they used a "leave-one-object-out crossvalidation" - 79 objects for training and one for test -, in this thesis only 40 objects were used for training (40 objects images per view), and the remaining 40 objects (40 objects images per view) were used to test the categorization model. So, while less objects in less views were tested to obtain the 93.75% average categorization rate of this thesis, Leibe & Schiele, 2003 used a higher number of training images. Comparing the average categorization rates, the one obtained in this thesis was 93.75% and the one obtained by Leibe & Schiele, 2003 was 93.02%; these rates are numerically very close, however, they aren't directly comparable due to the differences in the experimental setups. Finally, considering these thesis results, 5 objects images were miscategorized among all 80 tested images: 4 dog images were mistaken for horse images, and 1 apple image was mistaken for a tomato image. Interestingly, 'dog and horse' and 'apple and tomato' were best category groupings regarding the grouping hierarchies obtained for the cues explored in Leibe & Schiele, 2003.

Finally, Han et al., 2011 presented a canonical correlation- and view-based approach; the latter was evaluated both for object recognition and categorization. Using ETH-80 for the categorization tests, Han et al., 2011 used - from each category - 9 objects for the reference (training) set and the remaining object for the test set. Moreover, the training set comprised 5 views from each training object, and the testing set contained all views from each test object. The average rate obtained considering all 80 possible test objects was 82% (Han et al., 2011).

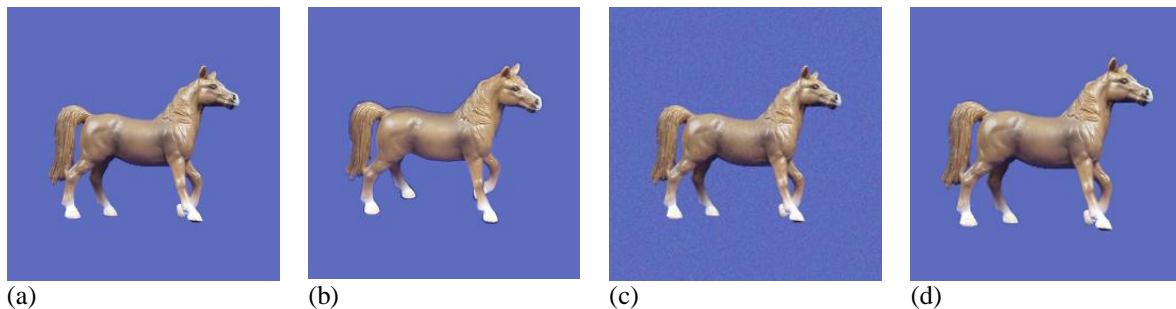
As understood, Han et al., 2011 used 9 objects per category for training, or, more precisely,  $9 \times 5 = 45$  objects images per category; moreover, 80 objects were used for test, or more precisely,  $80 \times 41 = 3280$  objects images. Comparing the referred experimental setup to the one in this thesis, here there were used 5 objects per category for training (5 training objects images per category and view) and 40 objects for test in two views (a total of 80 training objects images and 80 test objects images), and the average result of 93.75% was obtained. On one hand, there were used less training objects per category in this thesis' performed tests than in Han et al., 2011 ( $5 < 9$ ), being this thesis' categorization model 'less experienced' with objects – which is an advantage in terms of the significance of the 93.75% result here obtained; moreover, Han et al., 2011 used 5 views per training object while here only 2 views were used in this regard; on the other hand, they tested their model using more views per object (41 images/object) which is an advantage for the significance of their 82% result. Also an advantage for the significance of their 82% result is the fact that the background of the training/test objects images used in this thesis was homogenized, while in Han et al., 2011 the objects images were directly used from ETH-80. Hence, the 93.75% rate obtained in these thesis tests isn't directly comparable to their 82% result; however, 93.75% is 11.75 pp higher than the result by Han et al., 2011. Considering the latter 3 comparisons – to Rodrigues & du Buf, 2009b, Leibe & Schiele, 2003 and Han et al., 2011 – we consider 93.75% a very acceptable rate. Besides, 100% of the objects in most objects categories were well categorized in both views - cars, cows, cups, horses, pears and tomatoes (see Table 6.4.7 in Appendix 6.4).

## **4.2 - Evaluating the categorization model in terms of invariance**

As referred in the beginning of this chapter, we conclude this model's evaluation by performing other tests which assessed the model in terms of rotation, noise and scale invariance. For that, we used, as reference, the test objects in right profile, and also, the same right profile template sets used before; in other words, the template and test objects groups remain the same, being the test images the only difference between the test described in section 4.1 and the ones here described.

For examining the model in terms of rotation invariance, using the right profile view as reference, the available rotation variations available in the *eth80-cropped256* set (ETH-80, 2013) were left and right pan variations and a 'one-way' tilt variation; moreover, 22 degrees

is the angle difference that defines these slight variations. For the rotation test, we used the same 40 test objects tilted in minus 22 degrees (instead of using the ‘090-000’ view, another was used: ‘068-000’); consider Fig. 4.2.1 which shows in (a) a test object (horse7) in right profile, and in (b) the same test object image tilted in minus 22 degrees.



**Figure 4.2.1** - Test object (horse7) shown (a) in right profile. The same object is shown in (b) with a minus 22 degrees tilt rotation, in (c) with a 10% Gaussian noise rate addition, and in (d) with a 10% size increase rate.

As we can see by Fig. 4.2.1, the background of these 40 images was also homogenized. The results of this test are shown in Table 6.4.9, Appendix 6.4: a successful 95% rate was achieved, like in the tests before in section 4.1 regarding the right profile view (see Table 6.4.8, Appendix 6.4). In a class perspective, the most successful categories which had a 100% categorization rate were: apple, car, cow, cup, horse, pear, tomato (the majority of them); once more, the ‘dog’ category, which represents 12,5% of the test objects, had a 60% success rate, because 2 dog images were mistaken for horses. At this point, regarding the categorization success rate (95%), we can say that this model is rotation resistant when considering a -22 degrees tilt variation. Consider Table 6.4.13 in Appendix 6.4, which compares the results between the tests in section 4.1 (a) and the rotation tests described at this point (b). Regarding the results of the tests (a) and (b), the practical difference between these results resides in terms of threshold percentage needed in order to categorize the objects images. As expected, because the test objects in (b) are slightly rotated from right profile, the majority of the objects needed the same or a decreased threshold percentage (a decreased threshold percentage means a bigger margin for the concepts to agree) so that the objects could be categorized; however, there were “exceptional objects” for which there was an increase of the necessary threshold percentage - cow8, horse4 and horse7 – being still these objects correctly categorized, as the test in (a). Considering Table 6.4.9, we can notice that the objects which used less resources in order to reach a categorization result were cow, cup, pear and tomato, as they had the highest average necessary threshold percentage: 1. After those



objects, the ones which used less resources were, sorting by an increased resource use: horse (0.998), apple (0.994), dog (0.990) and car (0.890). Moreover, considering all categories, the average necessary threshold was 0.984, which is lower than the necessary average threshold needed in (a) for right profile: 0.997; a lower number means that a bigger margin for agreement between concepts was needed, which makes sense, once the template sets' view was not exactly the same as the objects'. Summarily, this test came to prove that the model is still stable even when the objects don't exactly have the same position as the template sets'.

For the noise tests, a 10% Gaussian noise (GN) rate was added to the 40 test objects in right profile (40 images) in order to test the resistance of the model in this regard. As an illustration, consider Fig. 4.2.1c) which shows a test object (horse7) in right profile when added a 10% GN rate. The results of this test are shown in Table 6.4.10 in Appendix 6.4: again, a successful 95% categorization rate was achieved, and the success rate of each category is once more maintained – the majority of them had a 100% categorization rate; again, the same two dogs were mistaken for horses. Therefore, considering the achieved rate, we can consider that the model is GN resistant as regards to a 10% GN rate. As mentioned for the previous test, the practical difference among the results in test (a) and the GN tests described at this point (c), resides in terms of threshold percentage that was necessary for categorizing the objects images. For this consideration, see Table 6.4.13 in Appendix 6.4. As expected, the necessary threshold from (a) to (c) was maintained or decreased for each object; this was expected because, although the objects position is the same, the fact that the background is different alters the LEs of the tested image - influencing the similarity percentages within the concepts, and, therefore, the final categorization result; also, as referred before, a smaller threshold means that a bigger margin for agreement among concepts was needed: more resources were expended. Nevertheless, for most objects the necessary threshold was exactly the same with the exception of dog9 and horse4. Consider again Table 6.4.10; the categories which needed a maximum threshold percentage (1), or in other words, the ones which used fewer resources in order to reach a categorization result were apple, cup, pear and tomato, being the latter 3 categories common in this regard to the latter test (b). After those objects, the ones which used fewer resources were, sorting by an increased resource use: car (0.998), dog (0.996), cow and horse (0.990). Furthermore, considering all categories, the average necessary threshold percentage was approximately 0.997: the same necessary average threshold percentage needed in (a) (right profile); once the categorization rates in (a) (right profile), (b) and (c) were the same, and for (a) (right profile) and (c) the average necessary

threshold percentage was 0.997 while for (b) was 0.984, the model seems to be even more resistant to a 10% GN rate (c) than to a minus 22 degrees tilt rotation (b), i.e., being the concepts more stable in their agreement for test (c).

Finally, the model was tested in terms of scale (size) invariance. In order to do so, the size of the 40 test objects in right profile (40 images) was altered, while the objects remained centred in the images (the background is homogeneous); see for e.g. Fig. 4.2.1d), which shows a test object whose size was increased by 10%. Considering the categorization tests (d) where the 40 test objects in right profile were increased in size by 10%, the results are shown in Table 6.4.11 in Appendix 6.4: 92.5% of the test objects were correctly categorized. Once more, the majority of the categories had a 100% categorization rate (apple, car, cup, horse, pear, tomato), with the exception of dog (as before) and cow; the same two problematic dogs were again mistaken for horses, and, this time, a cow (cow8) was also mistaken for a horse. These are acceptable errors, as dog/horse and cow/horse inaccuracies are within the animal category (they have roughly the same shapes). At this point we can consider that the model is scale resistant regarding a +10% scale change. Once more, the main practical difference among the results in test (a) and (d) resides in terms of threshold needed in order to categorize the objects images; for this consideration, see Table 6.4.13, Appendix 6.4. Again, in the results from (a) to (d) most of the necessary threshold percentages were maintained or decreased, for the same reasons stated before (an altered test image influences the LEs representation, and therefore the final categorization results). Consider again Table 6.4.11; the categories which used fewer resources in order to reach a categorization result were apple, car, pear and tomato; pear and tomato seem to be the winning categories in this regard until now. After those objects, the ones which used fewer resources were, sorting by an increased resource use: dog (0.994), cup and horse (0.992), and cow (0.984). Furthermore, considering all categories, the average necessary threshold was approximately 0.995, which is higher than the same parameter measured in (b) (0.984), but lower than the one in (c) or (a) (0.997). Considering the fact that the rate difference between 92.5% and 95% is just one image, we can say that the model is more stable in terms of a +10% scale variation than it is in a -22 degrees tilt rotation. However, among all tests described in this section until this point, the model performed best with a 10% GN rate adding, regarding categorization rate and average necessary threshold percentage.

Considering the categorization tests (e) where the size of the 40 test objects in right profile was decreased by 10% (40 images), the results are shown in Table 6.4.12 in Appendix 6.4: 82.5% of the test objects were correctly categorized. This time, half of the categories had a 100% categorization rate: car, cup, horse and pear; apple and cow had an 80% categorization rate, dog had a 40% rate, and tomato a 60% rate. The existing errors were apple/pear (1 error), cow/horse (1 error), dog/horse (3 errors), tomato/cow (1 error) and tomato/cup (1 error) – a total of 7 errors in 40 images (hence, an 82.5% categorization rate). Most errors until now were understandable because they existed among fruits/vegetables and among animals; however, we still had (only) one error between a fruit and an animal: tomato/cow. An increase in the scale seems to be better than a decrease, regarding the final categorization results in general, but we still consider that the model is moderately scale resistant regarding a -10% scale change. As referred before in section 4.1, the entire diagram - for choosing the winning category - on Fig. 3.4.3, was thought at first to be a good plan B, but it was indeed necessary; for example, in order to categorize dog7 – considering the conditions for the (e) categorization tests –, the intersection wasn't enough, because two categories remained: dog and horse. However, thanks to the diagram on Fig. 3.4.3, a final category was found: the answer to the questions 'did it return only one category? And is also the maximum concepts-score found equal to 2 or 3?' was 'yes: horse'. Furthermore, consider Table 6.4.13; comparing the necessary threshold percentages of test (a) to the ones of test (e), from (a) to (e) most have been maintained or decreased, as expected. Considering again Table 6.4.12, Appendix 6.4, this time, only car, cup and pear needed the fewest resources (average necessary threshold percentage = 1); after those objects, the ones which used fewer resources were, sorting by an increased resource use: cow (0.992), horse (0.988), dog (0.986), apple (0.964) and tomato (0.944). Furthermore, considering all categories, the average necessary threshold was approximately 0.984, as was the same parameter for the rotation invariance test (b). Finally, consider Table 6.4.14 from Appendix 6.4, where the rotation, noise and scale test results are summarized.

This chapter described a set of tests which evaluated our categorization model. We started by comparing the achieved results to other models' results, and described some extra tests made to evaluate the model in terms of rotation, GN and scale invariance. The next chapter will conclude this thesis work, presenting some final considerations and future work.

## 5. Conclusions and Future Work

---

### ABSTRACT

**This chapter summarizes this thesis work, presenting conclusions about the results obtained through testing the developed categorization model. Some final considerations and future work are presented.**

---

This thesis' goal was to build a biological model for object categorization, based on objects' 2D views. The model was based on multi-scale cortical attributes: V1 events, which are classified by type: line/edge, and polarity: positive/negative (Rodrigues & du Buf, 2009b). As a contribution, this model differs from the other categorization approaches - by Rodrigues & du Buf, 2009b, Leibe & Schiele, 2003 and Han et al., 2011 - thanks to its mutual agreement principle, created for the purpose of increasing the certainty degree of the given final category answer. The category assigned to an unknown object is obtained through a consensus among three different categorization concepts based on V1 events (LEs); each concept makes class suggestions/recommendations for a certain object, being the final assigned category present in the intersection of these suggestions. If the intersection is null, a 'bigger margin for consensus' is created - through a parameter named 'threshold percentage' -, in order to find at least one category on which 'all concepts agree on'; if the intersection is one category, that is the final answer; otherwise, if the intersection comprises several categories, the best scored one is then chosen through a proper process.

Each of the three concepts defines a unique way of calculating a similarity percentage between a certain object and a category. Therefore, as we used ETH-80 for our tests - a database which has 8 available categories (Leibe & Schiele, 2003) - an object is assigned 8 similarity percentages per concept (therefore, 24 in total). Ideally, for each concept, the true object's category would always be associated to the maximum similarity percentage; as this doesn't always happen, the 'threshold percentage' parameter was created to build a margin around the maximum similarity percentage found within the 8 similarity percentages of a concept; this is done for all three concepts, allowing each concept to give more categories

suggestions, thus facilitating a consensus. This way, the final category attributed to an object is the multi-scale convergence of three similarity concepts' responses.

The ETH-80 database (Leibe & Schiele, 2003) has 80 available objects: 40 were chosen for training (template objects) and the remaining for test. Moreover, 2 views were used for testing the model - left and right profile -, and also the background of all 160 used objects images was previously homogenized. The template objects were used to obtain left and right profile multi-scale templates sets for the 8 categories; these sets - whose events were used to compare to an unknown object' LEs - can be seen as the descriptors of the categories. In any of the three concepts, object's events were compared to the categories' sets in the same scale and view. Besides the fact that the template objects were used to train the model about each category (for both views), the former were also used to obtain some 'template results' for each concept; these results were crucial for structuring the model. When the model was finally built, it was evaluated using the template objects as test objects, reaching a 100% categorization success rate. Furthermore, the proof of concept comprised tests (a) and (b):

- a) The model was tested using left and right profile images from 40 test objects; the left and right profile multi-scale template sets were used to compare to each 'unknown object'. The 93.75% achieved average categorization success rate (Table 6.4.8, Appendix 6.4) was compared to other models' results (Rodrigues & du Buf, 2009b, Leibe & Schiele, 2003 and Han et al., 2011).
- b) The model was evaluated in terms of rotation, noise (GN) and scale invariance. As we only tested objects in right profile (or approximately in this view), the template sets used were the same as in (a) for that view. The right profile test objects were also the same as in (a), however, the used images were different: for the rotation test the objects were slightly rotated from right profile (a  $-22^\circ$  tilt rotation), for the Gaussian noise test the test objects images on right profile were added GN in 10%, and for the scale test the objects in right profile were rescaled in  $\pm 10\%$ . The background of the images used for the rotation and the scale tests was homogenized; the images for the GN test were obtained by adding a 10% GN rate to the right profile test images. On the rotation and GN tests, a 95% categorization success rate was achieved; on the +10% and the -10% scale tests, a 92.5% and an 82.5% categorization success rates were reached, respectively.

Regarding test (a) a categorization success rate was achieved: 93.75%; 92.5% was the success rate achieved for left profile, and 95.0% the one achieved for right profile. By comparison to other models' results, 93.75% was considered a very acceptable rate. 100% of the objects in most categories (6) were well categorized in both views and even the miscategorized objects were attributed a category which is similar to their true one: one apple image was mistaken for being a tomato image, and two dogs images were mistaken for being horses images. Besides the success of the tests in section 4.1 (categorizing the test objects in left and right profile), the model was also considered resistant to small rotation, noise and scale (size) variations (section 4.2).

Overall, the model was considered a success. In order to improve it, we propose the following future work:

1. In order to categorize an object, the model considers only one view. In the real world, when a person sees an object, several views are perceived. Thus, it is proposed that several views from an object are considered, with the aim of increasing the probability of a successful categorization.
2. Regarding this model's basic principle – mutual agreement among different concepts – other concepts could be added:
  - Once the object area is defined, its colors could be analyzed.
  - An object's contours could be analyzed in terms of derivatives, and encoded in order to describe this object in a certain view; using several descriptors from objects which belong to the same category and are shown in the same view, a category descriptor for that view could be obtained.
  - Concerning the latter suggestion, some “common points” in objects from a category could be obtained, which could be used for pre-categorization; these points can be used to define category common regions that can be useful to categorize an object.
  - As a pre-categorization descriptor, the quotient between an object's shortest length and its longest length could be calculated.

3. It was noticed that objects from the same category have always a pattern in their similarity percentages: some categories' percentages are generally higher than others. This being said, an object can be evaluated not only by finding the common category with the best similarity percentages, but also by evaluating all calculated similarity percentages (between this object and every category).
4. Implementing the suggestions 1 to 3 will expend more computational resources - (1) analyze more views, (2) implement more concepts, (3) enhance the 'winning category search system' by adding an additional module. In order to create a balance, as a next step, a higher performance model could be achieved by decreasing the number of explored scales to a minimum.
5. Finally, object segregation could be developed as a step to place before the categorization process.

# 6. Appendices

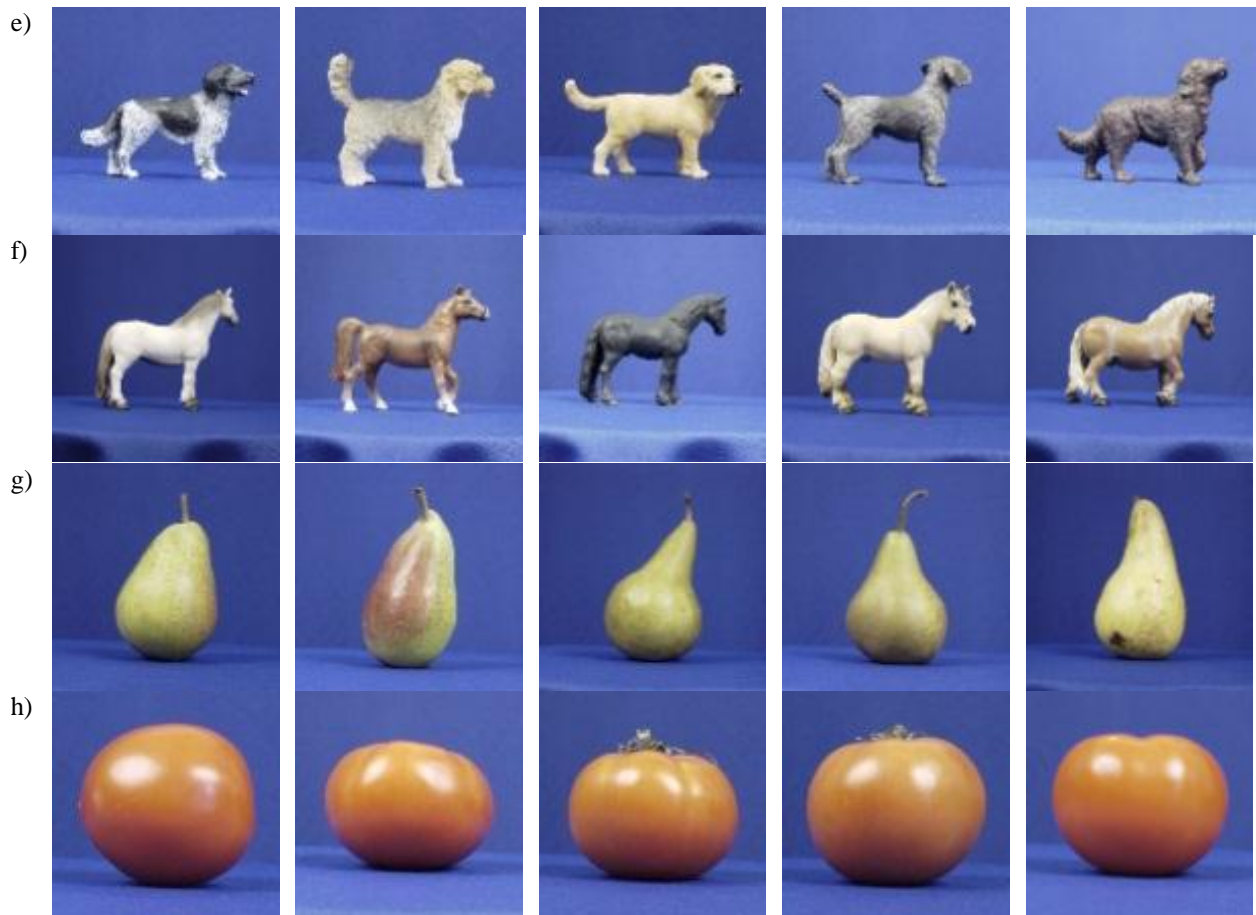
## Appendix 6.1 – ETH-80 database: some samples, and ‘segregated’ template objects

Some sample objects from the ETH-80 database (from the eight available categories) are presented (Leibe & Schiele, 2003). Therefore, Fig. 6.1.1 shows 5 category samples in right profile from each category in the ETH-80 database: a) apples; b) cars; c) cows; d) cups; e) dogs; f) horses; g) pears; h) tomatoes; the background of these objects images was manually homogenized, and they were posteriorly used as template or test objects (Fig. 6.1.2).



**Figure 6.1.1** - Sample objects from the ETH-80 database, in right profile. Top to bottom: a) apples; b) cars; c) cows; d) cups; e) dogs; f) horses; g) pears; h) tomatoes.





**Figure 6.1.1** (cont.) - Sample objects from the ETH-80 database, in right profile. Top to bottom: a) apples; b) cars; c) cows; d) cups; e) dogs; f) horses; g) pears; h) tomatoes.

As referred before, the background from the 160 ETH-80 object images was manually homogenized; these objects were divided in two groups - template and test objects; both sets have the same objects number in each category. Fig. 6.1.2 shows examples of the template objects in right profile, after the homogenization of the background (for the eight categories).



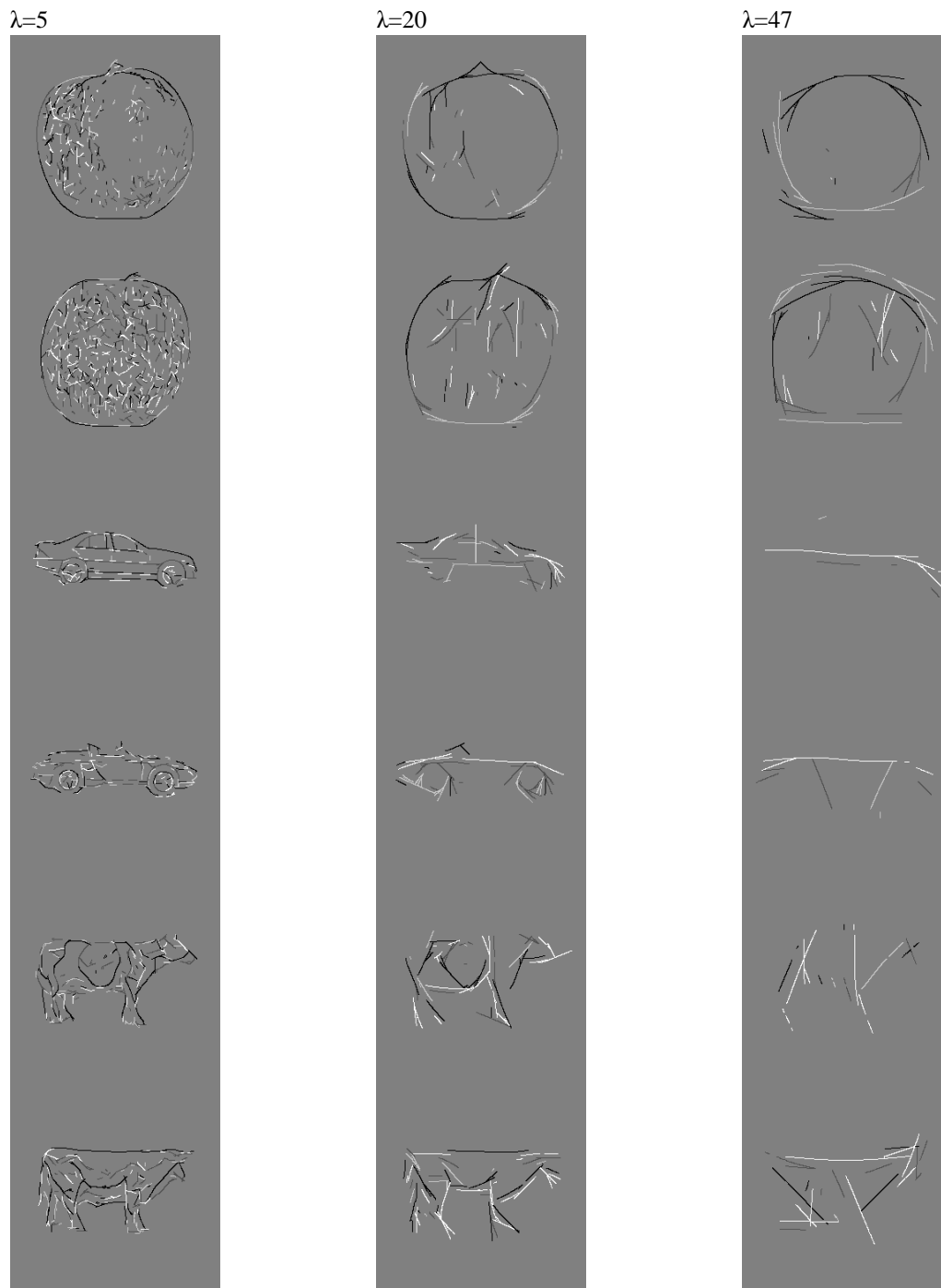
**Figure 6.1.2** – Some examples of template objects shown in right profile. These objects resulted from the manual ‘segregation’ of those in the ETH-80 database.



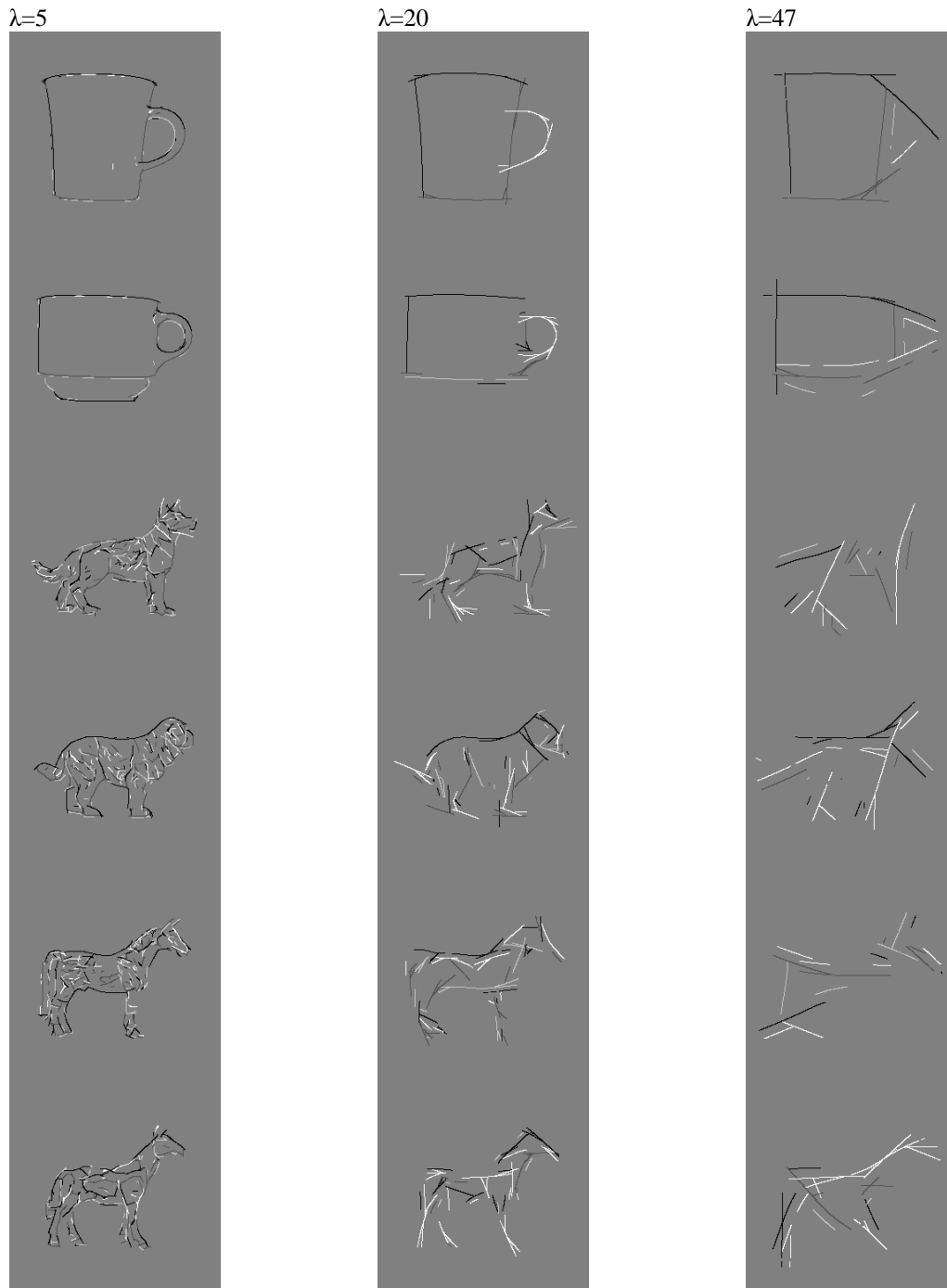
**Figure 6.1.2 (cont.)** - Template objects shown in right profile. These objects resulted from the manual 'segregation' of those in the ETH80 database.

## Appendix 6.2 - Examples of the lines and edges from some template objects

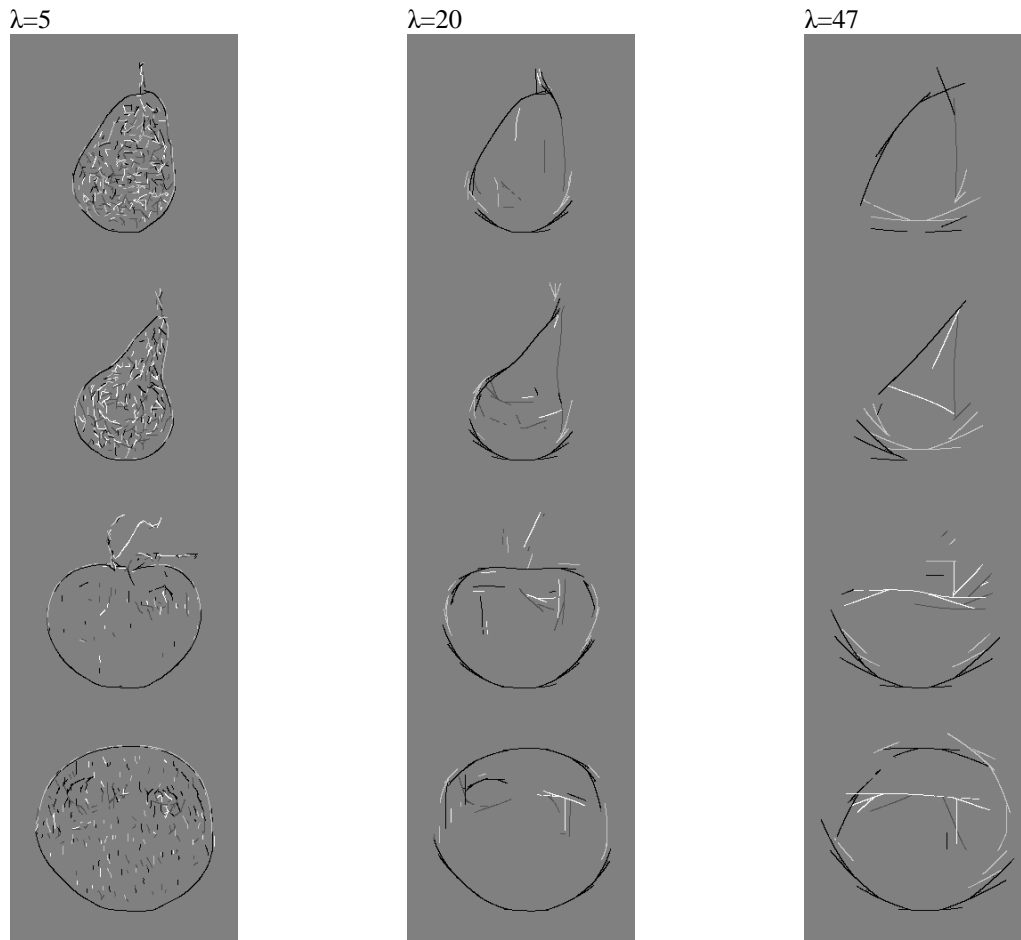
Two template objects from each category were chosen from Fig. 6.1.2, and their events (LEs) are shown in Fig. 6.2.1;  $\lambda=\{5,20,47\}$  were used to detect them, being fine to coarse scale shown from left to right.



**Figure 6.2.1** - The events of two template apples, cars and cows used from ETH-80, detected using  $\lambda=\{5,20,47\}$ ; fine to coarse scale is shown from left to right.



**Figure 6.2.1** (cont.) - The events of two template cups, dogs and horses used from ETH-80, detected using  $\lambda=\{5,20,47\}$ ; fine to coarse scale is shown from left to right.

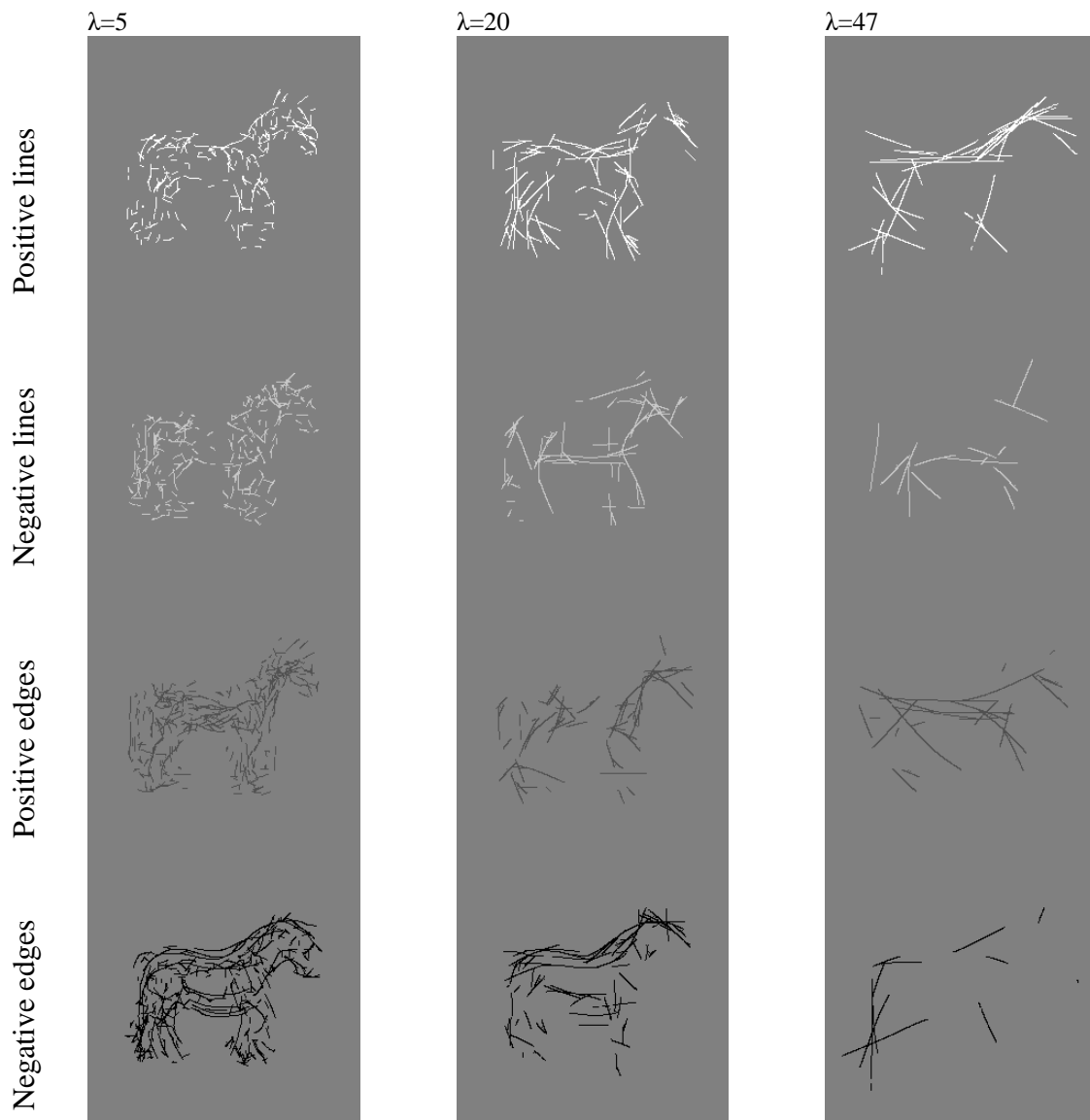


**Figure 6.2.1** (cont.) - The events of two template pears and tomatoes used from ETH-80, detected using  $\lambda=\{5,20,47\}$ ; fine to coarse scale is shown from left to right.

### Appendix 6.3 - Example of the animals' templates sets

Template sets were used as a descriptor that allows distinguishing one category from another. As the objects events were obtained in several scales and two views, a category has several sets, being the number of sets obtained by multiplying the number of scales by the number of used views. Tables 6.3.1 to 6.3.3 present a few templates set examples for the animals categories in right profile, for three scales  $\lambda = \{5, 20, 47\}$ : some horses' sets are shown in Table 6.3.1, the cows' in Table 6.3.2, and the dogs' in Table 6.3.3.

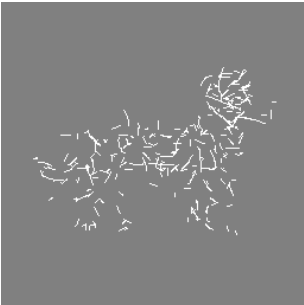
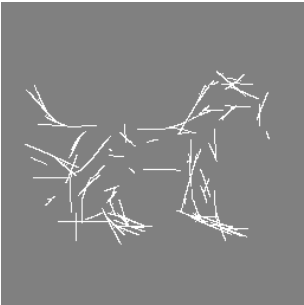
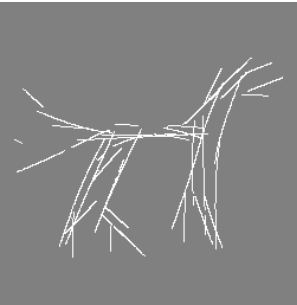

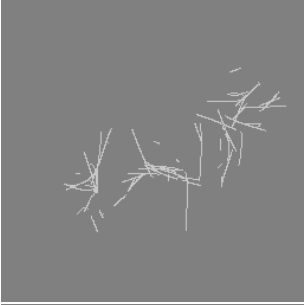
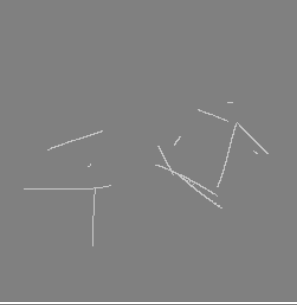




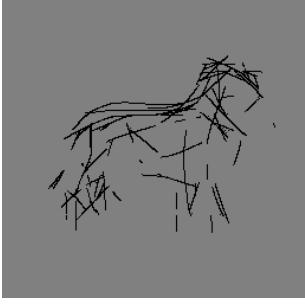

**Table 6.3.1** - Templates set examples for the horse category in right profile, using three scales  $\lambda = \{5, 20, 47\}$ .



**Table 6.3.2** - Template set examples for the cow category in right profile, using three scales  $\lambda = \{5, 20, 47\}$ .

	$\lambda=5$	$\lambda=20$	$\lambda=47$
Positive lines			
Negative lines			
Positive edges			
Negative edges			

**Table 6.3.3** - Template set examples for the dog category in right profile, using three scales  $\lambda = \{5, 20, 47\}$ .

	$\lambda=5$	$\lambda=20$	$\lambda=47$
Positive lines			
Negative lines			
Positive edges			
Negative edges			



## Appendix 6.4 - Detailed categorization results

The four concepts explored in section 3.3 were tested using the template objects; three of these were chosen to build the categorization model. Hence, Table 6.4.1 shows the categorization results achieved for the first three used concepts when the latter were still being explored; Tables 6.4.2 to 6.4.4 show the similarity percentages between each template object and category, for the three concepts that had the best score (100%) - concepts 1, 2 and 3.

**Table 6.4.1** - Categorization results achieved using the template objects to test the three concepts; the template objects were tested in left and right profile. A – apple, Ca – Car, Co-Cow, C - Cup, D – dog, H – horse, P – pear, T – tomato.

	Left profile	Right profile	
Object's id	Object reference	Object reference	Cat. Result
A,1	26075	28036	A
A,2	18526	36916	A
A,3	28510	31041	A
A,4	18238	18533	A
A,5	28552	27145	A
Ca,1	14715	14073	Ca
Ca,2	16115	16280	Ca
Ca,5	15508	15902	Ca
Ca,7	16235	16152	Ca
Ca,9	18499	16406	Ca
Co,1	19350	21275	Co
Co,2	22436	22060	Co
Co,5	19413	18148	Co
Co,6	24267	25872	Co
Co,7	23318	22480	Co
C,1	13661	14602	C
C,3	17758	16806	C
C,4	28236	27574	C
C,9	18924	18450	C
C,10	14192	14581	C
D,1	22284	23790	D
D,2	23477	24815	D
D,5	20513	21220	D
D,8	22344	24235	D
D,10	28362	28632	D
H,1	20746	23253	H
H,2	20705	19899	H
H,3	20614	20493	H
H,5	20967	20582	H
H,10	25000	23585	H
P,1	23561	20438	P
P,3	25209	22219	P
P,4	19808	19828	P
P,5	20607	18427	P
P,9	17870	19014	P
T,2	25166	24799	T
T,6	26994	27599	T
T,8	25387	23649	T
T,9	27385	28165	T
T,10	28305	29178	T

**Table 6.4.2** - Categorization results achieved using the template images (in left and right profile) to test concept 1 (A – apple, Ca – Car, Co-Cow, C - Cup, D – dog, H – horse, P – pear, T – tomato). Regarding this test, the dendritic field radius function deduced in section 3.3 was used, the events’ type and polarity are considered, and the reference is given by the objects.

Object's id	Similarity percentages [%]																Cat. Result
	Left profile								Right profile								
	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	
A,1	100	17.70	37.46	33.35	25.92	30.48	31.63	73.57	100	16.61	35.73	41.73	26.21	28.85	28.51	64.87	A
A,2	100	10.74	32.63	44.41	23.42	20.10	27.99	54.75	100	18.88	37.22	41.30	32.48	35.19	29.55	56.47	A
A,3	100	18.55	40.90	42.35	30.79	30.63	31.67	70.34	100	18.99	38.14	46.45	28.85	30.74	30.78	80.00	A
A,4	100	11.64	39.48	47.83	28.87	27.72	41.75	67.95	100	11.73	31.53	39.78	25.03	25.64	36.82	64.24	A
A,5	100	17.41	35.91	36.71	31.32	31.60	35.13	67.55	100	23.11	40.92	42.95	39.26	34.21	33.78	76.50	A
Ca,1	36.68	100	57.72	38.26	49.69	46.97	24.68	43.24	46.22	100	66.48	36.42	51.34	48.93	23.86	47.74	Ca
Ca,2	36.44	100	59.03	42.09	61.13	51.96	21.26	41.92	47.72	100	60.46	44.17	62.79	55.93	26.50	44.14	Ca
Ca,5	33.12	100	61.75	35.39	65.12	53.37	28.38	45.85	47.22	100	58.16	37.66	65.02	50.11	28.48	48.45	Ca
Ca,7	33.87	100	58.47	39.44	53.37	52.63	23.81	43.31	47.15	100	63.58	37.28	54.04	53.01	25.71	43.69	Ca
Ca,9	28.96	100	57.41	38.39	61.61	50.12	22.67	41.25	49.45	100	58.14	40.10	58.67	48.84	29.67	45.31	Ca
Co,1	34.94	27.95	100	45.29	53.08	60.33	31.52	42.89	48.31	33.35	100	44.24	58.44	66.13	23.34	46.31	Co
Co,2	36.13	31.77	100	43.47	58.70	61.79	29.36	44.08	44.82	27.12	100	44.07	58.01	58.93	30.50	44.13	Co
Co,5	32.19	36.49	100	46.87	60.11	57.38	33.63	30.65	46.69	44.06	100	44.87	70.55	58.44	30.00	40.35	Co
Co,6	39.15	42.36	100	48.56	47.34	46.16	30.45	42.49	44.36	41.38	100	48.03	48.33	44.51	31.89	40.95	Co
Co,7	33.94	40.41	100	47.39	67.82	72.68	33.88	40.56	48.75	35.34	100	44.07	65.06	67.73	31.53	45.67	Co
C,1	51.04	20.28	42.32	100	40.16	37.22	35.61	44.21	60.05	21.29	44.51	100	39.76	36.32	32.51	55.22	C
C,3	42.64	27.34	49.09	100	49.18	39.09	28.76	42.23	49.96	27.16	44.94	100	45.19	36.49	32.68	51.30	C
C,4	38.19	33.95	52.50	100	45.73	50.52	34.92	42.10	50.97	32.20	54.89	100	43.09	47.66	29.88	45.00	C
C,9	34.46	31.85	52.82	100	44.40	44.25	16.98	42.43	43.25	31.07	50.21	100	47.41	41.85	23.07	43.86	C
C,10	32.91	30.12	71.03	100	45.52	44.84	26.83	46.80	39.88	27.03	68.99	100	45.26	41.09	31.62	44.65	C
D,1	28.44	33.70	54.77	37.50	100	56.09	33.85	27.97	35.47	32.72	56.68	32.43	100	53.59	39.30	30.59	D
D,2	29.30	43.86	69.15	41.06	100	69.43	36.83	37.94	43.84	42.36	66.35	39.11	100	70.95	38.12	45.76	D
D,5	30.67	36.24	63.95	38.69	100	85.12	41.53	39.78	47.82	34.63	61.31	44.98	100	80.34	38.93	43.08	D
D,8	33.69	36.74	60.20	42.75	100	54.50	39.81	38.98	43.31	32.47	61.23	40.12	100	64.93	36.84	38.00	D
D,10	29.91	40.74	56.12	38.92	100	59.77	37.28	33.29	39.73	43.66	53.39	38.77	100	63.04	38.82	33.19	D
H,1	34.74	34.85	63.09	41.96	66.03	100	31.47	41.30	44.85	37.27	60.56	44.81	68.06	100	30.00	42.13	H
H,2	29.49	44.10	65.22	44.60	75.09	100	38.41	40.22	50.64	36.93	66.73	41.37	76.97	100	31.21	43.19	H
H,3	34.63	41.74	65.11	39.95	79.86	100	38.49	40.12	47.27	40.89	67.49	45.00	78.01	100	37.94	42.83	H
H,5	35.26	38.90	68.48	40.30	59.09	100	28.02	45.61	48.16	42.17	67.33	45.92	67.52	100	33.63	44.05	H
H,10	31.38	39.36	72.14	44.11	72.63	100	38.45	45.94	49.71	39.12	71.30	38.47	79.75	100	35.57	43.02	H
P,1	42.18	23.17	44.20	34.89	47.45	42.17	100	43.33	54.15	23.67	37.71	38.29	47.58	42.09	100	50.57	P
P,3	38.94	24.53	43.89	31.32	45.21	42.71	100	38.22	57.01	22.73	36.87	37.19	50.78	47.34	100	46.19	P
P,4	46.33	18.73	39.43	28.91	36.31	31.65	100	38.46	54.08	21.03	38.24	26.88	43.94	36.05	100	42.93	P
P,5	43.37	20.65	31.60	35.07	33.98	30.43	100	37.41	50.24	18.34	32.43	33.15	42.51	36.28	100	39.43	P
P,9	52.45	15.72	36.07	39.36	40.99	32.76	100	46.46	60.41	19.81	32.09	35.72	43.12	34.12	100	53.94	P
T,2	51.44	34.92	42.16	31.53	40.79	45.43	37.05	100	64.55	36.26	46.07	38.49	51.41	46.66	39.66	100	T
T,6	73.72	20.54	34.45	34.67	28.78	30.57	27.82	100	80.57	19.28	40.49	41.55	30.40	29.62	26.86	100	T
T,8	63.27	18.74	40.26	42.20	27.88	32.30	29.63	100	72.71	22.46	42.29	46.53	28.01	30.15	26.17	100	T
T,9	82.44	20.70	37.16	38.66	28.27	33.30	23.64	100	84.15	20.90	38.42	44.11	26.99	30.42	26.72	100	T
T,10	65.72	15.68	39.75	37.05	29.38	30.40	22.57	100	81.30	17.34	37.45	46.22	32.72	30.01	27.74	100	T

**Table 6.4.3** - Categorization results achieved using the template images (in left and right profile) to test concept2 (A – apple, Ca – Car, Co-Cow, C - Cup, D – dog, H – horse, P – pear, T – tomato). Regarding this test, the dendritic field radius function deduced in section 3.3 was used, the events' type and polarity were not considered, and the reference is given by the objects.

Object's id	Similarity percentages [%]																Cat. Result
	Left profile								Right profile								
	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	
A,1	100	33.23	57.28	64.30	52.59	46.18	51.02	92.32	100	29.04	56.55	64.44	54.57	47.88	39.55	91.29	A
A,2	100	21.61	51.51	60.09	49.23	42.73	35.10	92.27	100	32.23	60.01	68.05	61.10	57.96	54.97	90.45	A
A,3	100	30.10	62.97	67.83	60.44	53.13	40.77	91.88	100	30.86	62.03	69.45	58.17	53.55	39.13	94.91	A
A,4	100	21.89	52.75	65.08	55.59	49.70	48.27	96.15	100	24.26	44.63	55.18	51.38	42.15	44.56	96.32	A
A,5	100	27.66	61.71	65.73	63.01	54.12	49.40	93.02	100	35.12	64.03	66.44	65.79	55.13	43.72	92.79	A
Ca,1	74.56	100	91.99	76.34	78.95	64.42	43.59	83.45	86.59	100	94.73	73.18	78.68	66.73	48.28	87.91	Ca
Ca,2	70.30	100	90.67	83.69	89.46	69.13	48.58	77.94	84.08	100	95.39	83.78	88.73	73.05	49.83	84.91	Ca
Ca,5	66.20	100	93.04	77.61	87.11	74.07	58.62	79.12	83.30	100	91.05	78.25	86.76	72.96	57.22	82.33	Ca
Ca,7	63.32	100	91.91	78.57	86.60	70.53	54.50	80.14	81.08	100	93.80	77.09	86.09	69.89	57.47	81.28	Ca
Ca,9	61.99	100	92.06	76.08	85.35	67.33	54.25	79.10	80.92	100	95.21	77.05	84.38	67.94	53.60	80.70	Ca
Co,1	70.84	52.83	100	83.91	83.91	84.98	52.04	79.21	86.87	59.96	100	86.26	87.46	89.37	50.23	81.31	Co
Co,2	74.63	58.69	100	81.79	87.75	89.37	48.56	85.13	84.90	55.22	100	87.29	88.93	88.39	53.16	79.46	Co
Co,5	64.22	66.67	100	82.72	89.20	75.88	54.21	70.94	80.61	68.29	100	87.24	90.21	79.49	55.58	72.15	Co
Co,6	75.97	62.75	100	80.92	74.41	65.14	47.83	79.75	84.93	63.68	100	81.56	73.89	64.16	55.00	77.11	Co
Co,7	69.02	58.79	100	88.29	93.56	92.45	55.93	74.15	84.77	55.18	100	87.00	90.23	89.92	56.33	75.95	Co
C,1	74.69	33.18	61.45	100	70.87	61.24	52.77	80.07	81.56	35.17	63.87	100	70.78	67.46	49.88	82.65	C
C,3	69.54	43.59	81.95	100	75.36	61.80	49.17	75.36	78.12	42.64	82.45	100	77.01	64.52	49.39	75.27	C
C,4	65.17	53.53	79.54	100	78.17	73.46	65.42	75.89	83.55	49.95	77.48	100	73.83	72.90	58.78	78.23	C
C,9	79.95	47.33	91.69	100	76.98	67.08	43.73	80.59	85.34	47.62	91.36	100	78.44	68.68	48.03	83.18	C
C,10	64.51	46.41	97.27	100	80.24	70.15	47.96	73.70	81.61	42.78	95.77	100	79.15	70.26	46.46	75.22	C
D,1	64.93	59.71	83.36	74.68	100	77.35	65.19	72.16	78.56	55.51	83.14	73.70	100	83.10	66.42	70.35	D
D,2	68.44	59.57	94.01	81.04	100	87.38	61.30	73.83	81.89	60.50	93.65	80.90	100	89.62	59.47	73.56	D
D,5	68.47	53.93	86.09	78.05	100	96.89	74.08	77.65	84.59	56.08	85.18	80.66	100	97.00	64.67	79.33	D
D,8	67.58	54.16	89.90	80.04	100	75.29	63.73	70.97	81.14	50.87	86.26	77.51	100	82.03	61.79	72.95	D
D,10	61.19	61.77	87.33	77.70	100	77.84	71.55	66.24	79.25	62.21	84.73	75.37	100	81.09	67.92	71.24	D
H,1	75.30	56.47	91.07	81.44	95.75	100	62.87	82.76	88.49	53.83	87.56	82.07	92.63	100	56.34	81.59	H
H,2	62.83	63.14	90.16	82.91	96.09	100	75.79	71.70	84.60	56.68	86.57	83.73	95.84	100	64.02	73.67	H
H,3	69.02	56.45	88.28	79.37	95.31	100	72.37	74.95	84.50	58.54	88.25	85.18	94.36	100	71.89	73.88	H
H,5	69.40	64.48	97.06	80.18	91.59	100	62.24	82.01	83.10	65.87	97.07	85.38	92.84	100	65.23	80.95	H
H,10	70.65	54.92	92.90	83.60	94.46	100	62.61	79.07	84.05	57.97	92.02	83.85	96.09	100	60.93	75.63	H
P,1	72.20	41.31	69.30	73.53	72.97	68.34	100	75.49	83.17	39.42	60.32	73.52	71.27	65.19	100	77.60	P
P,3	64.77	42.36	65.22	68.55	74.87	71.55	100	72.80	84.06	40.02	60.21	68.86	76.66	73.17	100	73.34	P
P,4	71.69	34.09	59.24	65.63	63.87	61.40	100	72.80	84.78	38.04	58.10	61.67	63.14	58.98	100	76.85	P
P,5	68.90	35.60	53.25	66.85	58.70	57.21	100	73.72	83.44	31.48	50.71	68.70	58.92	59.43	100	76.79	P
P,9	74.61	29.44	52.13	71.35	58.23	54.52	100	80.17	88.19	32.98	50.37	65.74	61.77	56.26	100	81.04	P
T,2	81.79	46.72	74.48	70.80	70.41	68.41	60.55	100	89.97	49.70	71.94	68.84	71.32	65.38	62.95	100	T
T,6	91.77	34.03	62.10	65.24	58.71	50.62	43.62	100	94.91	33.61	60.99	64.46	58.55	51.28	38.34	100	T
T,8	88.46	33.69	69.13	76.87	58.86	52.22	48.08	100	93.79	38.21	68.98	76.27	60.49	56.23	46.17	100	T
T,9	93.85	35.99	70.97	72.74	61.56	54.14	44.92	100	96.29	36.77	62.00	67.52	58.87	52.31	43.92	100	T
T,10	92.02	33.43	72.52	76.85	62.24	52.00	40.68	100	96.31	30.68	63.28	75.41	61.18	53.43	41.87	100	T

**Table 6.4.4** - Categorization results achieved using the template images (in left and right profile) to test concept3 (A – apple, Ca – Car, Co-Cow, C - Cup, D – dog, H – horse, P – pear, T – tomato). Regarding this test, the dendritic field radius function deduced in section 3.3 was used, the events’ type and polarity were considered, and the reference is given by the category.

Object's id	Similarity percentages [%]																Cat. Result
	Left profile								Right profile								
	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes	
A,1	25.86	6.94	10.36	10.62	6.83	8.86	9.45	16.54	23.83	7.26	10.57	14.47	7.11	8.96	9.81	15.74	A
A,2	18.37	2.99	6.41	10.05	4.39	4.15	5.94	8.74	31.37	10.86	14.50	18.85	11.60	14.38	13.39	18.04	A
A,3	28.27	7.95	12.37	14.75	8.88	9.74	10.34	17.29	26.38	9.19	12.50	17.83	8.67	10.57	11.72	21.49	A
A,4	18.09	3.19	7.64	10.66	5.32	5.64	8.72	10.68	15.75	3.39	6.17	9.11	4.49	5.26	8.37	10.31	A
A,5	28.31	7.48	10.87	12.80	9.04	10.06	11.49	16.62	23.07	9.78	11.73	14.41	10.31	10.28	11.25	17.97	A
Ca,1	5.35	22.13	9.01	6.88	7.39	7.71	4.16	5.48	5.53	21.94	9.88	6.34	6.99	7.63	4.12	5.82	Ca
Ca,2	5.82	24.24	10.09	8.28	9.96	9.34	3.92	5.82	6.60	25.38	10.39	8.89	9.89	10.08	5.29	6.22	Ca
Ca,5	5.09	23.32	10.15	6.70	10.21	9.23	5.04	6.13	6.38	24.79	9.76	7.40	10.01	8.82	5.56	6.67	Ca
Ca,7	5.45	24.42	10.07	7.82	8.76	9.53	4.43	6.06	6.47	25.18	10.84	7.44	8.45	9.48	5.09	6.11	Ca
Ca,9	5.31	27.82	11.26	8.67	11.52	10.34	4.80	6.58	6.89	25.57	10.07	8.13	9.32	8.87	5.97	6.43	Ca
Co,1	6.70	8.13	20.52	10.71	10.38	13.02	6.99	7.15	8.73	11.06	22.46	11.64	12.03	15.58	6.09	8.53	Co
Co,2	8.04	10.72	23.79	11.91	13.32	15.46	7.55	8.52	8.40	9.32	23.29	12.02	12.39	14.40	8.26	8.43	Co
Co,5	6.20	10.65	20.59	11.11	11.80	12.43	7.48	5.13	7.20	12.46	19.16	10.07	12.39	11.75	6.68	6.34	Co
Co,6	9.42	15.46	25.73	14.39	11.61	12.50	8.46	8.89	9.75	16.69	27.31	15.36	12.10	12.75	10.12	9.17	Co
Co,7	7.85	14.17	24.73	13.50	15.99	18.90	9.05	8.15	9.31	12.38	23.73	12.25	14.16	16.86	8.70	8.89	Co
C,1	6.91	4.17	6.13	16.69	5.55	5.67	5.57	5.21	7.45	4.85	6.86	18.05	5.62	5.87	5.83	6.98	C
C,3	7.51	7.30	9.24	21.69	8.83	7.74	5.85	6.46	7.14	7.12	7.97	20.78	7.35	6.79	6.74	7.46	C
C,4	10.69	14.42	15.72	34.49	13.05	15.91	11.29	10.25	11.94	13.84	15.98	34.09	11.50	14.55	10.11	10.74	C
C,9	6.47	9.07	10.60	23.12	8.50	9.34	3.68	6.92	6.78	8.94	9.78	22.81	8.47	8.55	5.22	7.00	C
C,10	4.63	6.43	10.69	17.34	6.53	7.10	4.36	5.73	4.94	6.14	10.62	18.03	6.39	6.64	5.66	5.64	C
D,1	6.29	11.30	12.94	10.21	22.53	13.94	8.64	5.37	7.17	12.13	14.23	9.54	23.03	14.12	11.47	6.30	D
D,2	6.82	15.49	17.22	11.77	23.73	18.18	9.90	7.68	9.24	16.38	17.38	12.00	24.02	19.50	11.61	9.83	D
D,5	6.24	11.18	13.91	9.70	20.74	19.48	9.76	7.03	8.62	11.46	13.73	11.80	20.54	18.88	10.14	7.91	D
D,8	7.47	12.35	14.26	11.67	22.59	13.58	10.19	7.51	8.92	12.27	15.67	12.02	23.46	17.43	10.96	7.97	D
D,10	8.41	17.38	16.88	13.48	28.67	18.91	12.11	8.14	9.67	19.49	16.14	13.72	27.71	19.99	13.64	8.23	D
H,1	7.15	10.88	13.88	10.63	13.85	23.14	7.48	7.39	8.86	13.51	14.87	12.88	15.32	25.75	8.56	8.48	H
H,2	6.06	13.73	14.32	11.28	15.72	23.10	9.11	7.18	8.56	11.45	14.02	10.18	14.82	22.04	7.62	7.44	H
H,3	7.08	12.94	14.23	10.06	16.64	22.99	9.09	7.13	8.23	13.06	14.60	11.40	15.47	22.69	9.54	7.60	H
H,5	7.33	12.27	15.23	10.32	12.53	23.39	6.73	8.24	8.42	13.53	14.63	11.69	13.45	22.79	8.49	7.85	H
H,10	7.78	14.80	19.12	13.47	18.36	27.89	11.01	9.90	9.96	14.38	17.75	11.22	18.20	26.12	10.29	8.78	H
P,1	9.86	8.21	11.04	10.04	11.30	11.08	26.98	8.80	9.41	7.54	8.14	9.68	9.41	9.53	25.08	8.95	P
P,3	9.74	9.30	11.73	9.64	11.52	12.01	28.87	8.30	10.76	7.87	8.65	10.22	10.92	11.65	27.26	8.88	P
P,4	9.10	5.58	8.28	7.00	7.27	6.99	22.69	6.57	9.11	6.50	8.00	6.59	8.43	7.92	24.33	7.37	P
P,5	8.86	6.40	6.90	8.83	7.08	6.99	23.60	6.65	7.87	5.27	6.31	7.55	7.58	7.40	22.61	6.29	P
P,9	9.29	4.22	6.84	8.59	7.41	6.53	20.47	7.16	9.76	5.87	6.44	8.40	7.94	7.18	23.33	8.88	P
T,2	12.84	13.22	11.25	9.69	10.38	12.75	10.68	21.69	13.60	14.02	12.06	11.80	12.34	12.81	12.07	21.47	T
T,6	19.73	8.34	9.86	11.43	7.85	9.20	8.60	23.27	18.90	8.29	11.80	14.18	8.12	9.05	9.10	23.89	T
T,8	15.93	7.15	10.84	13.09	7.16	9.15	8.61	21.88	14.61	8.28	10.56	13.61	6.41	7.90	7.59	20.47	T
T,9	22.39	8.52	10.79	12.93	7.83	10.17	7.41	23.60	20.14	9.17	11.42	15.36	7.36	9.49	9.23	24.38	T
T,10	18.45	6.67	11.93	12.81	8.41	9.60	7.32	24.40	20.16	7.88	11.54	16.67	9.24	9.70	9.93	25.26	T

In order to achieve the results shown in Table 6.4.4, the categories references were used; these are shown in Table 6.4.5.

**Table 6.4.5** - Categories references shown for both views – left and right profile.

View	Category's reference							
	Apples	Cars	Cows	Cups	Dogs	Horses	Pears	Tomatoes
Right profile	117669	64154	94725	80878	103322	90301	81493	115532
Left profile	100838	66487	94299	81866	98914	89649	87313	116014

Finally, after the model was built (in section 3.4), other results were achieved using the test objects in left and right profile. Therefore, the test objects references are shown in Table 6.4.6, and the categories references shown in Table 6.4.5 should be also recalled, as both count for the results shown in Tables 6.4.7 and 6.4.8.

**Table 6.4.6** – Test objects' reference in both views: right (000) and left (180).

Object's name	Object's reference	Object's name	Object's reference	Object's name	Object's reference	Object's name	Object's reference
Apple6-090-000	43141	Cow3-090-000	22931	Dog3-090-000	24047	Pear2-090-000	18272
Apple6-090-180	44948	Cow3-090-180	23663	Dog3-090-180	25164	Pear2-090-180	17037
Apple7-090-000	27747	Cow4-090-000	25215	Dog4-090-000	24628	Pear6-090-000	15731
Apple7-090-180	33803	Cow4-090-180	23692	Dog4-090-180	27863	Pear6-090-180	16032
Apple8-090-000	26441	Cow8-090-000	24954	Dog6-090-000	25283	Pear7-090-000	17908
Apple8-090-180	27365	Cow8-090-180	22073	Dog6-090-180	23128	Pear7-090-180	20688
Apple9-090-000	32587	Cow9-090-000	21435	Dog7-090-000	31302	Pear8-090-000	15641
Apple9-090-180	28782	Cow9-090-180	21663	Dog7-090-180	28849	Pear8-090-180	15814
Apple10-090-000	25604	Cow10-090-000	21814	Dog9-090-000	27442	Pear10-090-000	19588
Apple10-090-180	31214	Cow10-090-180	23890	Dog9-090-180	24900	Pear10-090-180	15104
Car3-090-000	16154	Cup2-090-000	15857	Horse4-090-000	20156	Tomato1-090-000	24676
Car3-090-180	17184	Cup2-090-180	16438	Horse4-090-180	19261	Tomato1-090-180	23848
Car6-090-000	17463	Cup5-090-000	14669	Horse6-090-000	18617	Tomato3-090-000	24205
Car6-090-180	17634	Cup5-090-180	16101	Horse6-090-180	18782	Tomato3-090-180	25720
Car11-090-000	17450	Cup6-090-000	15470	Horse7-090-000	25880	Tomato4-090-000	19577
Car11-090-180	16790	Cup6-090-180	15613	Horse7-090-180	25624	Tomato4-090-180	21992
Car12-090-000	18333	Cup7-090-000	18115	Horse8-090-000	21113	Tomato5-090-000	22805
Car12-090-180	16332	Cup7-090-180	17237	Horse8-090-180	21315	Tomato5-090-180	22378
Car14-090-000	13867	Cup8-090-000	17123	Horse9-090-000	24081	Tomato7-090-000	29143
Car14-090-180	13842	Cup8-090-180	16654	Horse9-090-180	22151	Tomato7-090-180	29019

**Table 6.4.7** - Categorization results achieved by using the test objects, in right (090-000) and left (090-180) profile views. For the winning category (Wcat) results, the correspondences are as follows: A – apple, Ca – car, Co – cow, C – cup, D – dog, H – horse, P – pear, T – tomato. For each object, we show the necessary threshold percentage ( $NT_P$ ) that the model needed in order to provide a final categorization result. Also, the average  $NT_P$  for each category and view was calculated.

Object's name	Intersected categories (IC), and respective similarity percentages[%]				Wcat	$NT_P$	$A_{NT_P}$ LP	$A_{NT_P}$ RP
	IC	Concept 1	Concept 2	Concept 3				
Apple6-090-000	Apple	77.40	94.68	28.38	A	1	0.998	1
Apple6-090-180	Apple	63.47	87.15	28.29	A	0.99		
Apple7-090-000	Apple	85.58	96.06	20.18	A	1		
Apple7-090-180	Apple	63.78	87.81	21.38	A	1		
Apple8-090-000	Apple	86.34	96.21	19.40	A	1		
Apple8-090-180	Apple	72.24	90.46	19.60	A	1		
Apple9-090-000	Apple	74.92	94.04	20.75	A	1		
Apple9-090-180	Apple	74.47	87.91	21.26	A	1		
Apple10-090-000	Apple	68.63	94.67	14.93	A	1		
Apple10-090-180	Tomato	62.68	88.34	16.86	T	1		
Car3-090-000	Car	91.10	97.97	22.94	Ca	1	1	0.998
Car3-090-180	Car	90.57	98.39	23.41	Ca	1		
Car6-090-000	Car	83.14	95.75	22.63	Ca	0.99		
Car6-090-180	Car	83.49	96.46	22.14	Ca	1		
Car11-090-000	Car	77.19	96.65	20.99	Ca	1		
Car11-090-180	Car	78.34	95.84	19.78	Ca	1		
Car12-090-000	Car	85.19	97.64	24.34	Ca	1		
Car12-090-180	Car	85.86	97.62	21.09	Ca	1		
Car14-090-000	Car	84.69	99.35	18.31	Ca	1		
Car14-090-180	Car	87.57	99.87	18.23	Ca	1		
Cow3-090-000	Cow	79.47	97.46	19.24	Co	1	1	0.990
Cow3-090-180	Cow	77.70	95.48	19.50	Co	1		
Cow4-090-000	Cow	76.80	95.17	20.44	Co	1		
Cow4-090-180	Cow	74.39	96.10	18.69	Co	1		
Cow8-090-000	Cow	71.53	93.90	18.84	Co	0.95		
Cow8-090-180	Cow	79.38	98.53	18.58	Co	1		
Cow9-090-000	Cow	75.02	96.31	16.98	Co	1		
Cow9-090-180	Cow	75.03	97.49	17.24	Co	1		
Cow10-090-000	Cow	81.98	99.12	18.88	Co	1		
Cow10-090-180	Cow	76.42	97.85	19.36	Co	1		
Cup2-090-000	Cup	87.63	95.79	17.18	C	1	1	1
Cup2-090-180	Cup	82.05	94.91	16.47	C	1		
Cup5-090-000	Cup	74.70	96.31	13.55	C	1		
Cup5-090-180	Cup	71.23	95.04	14.01	C	1		
Cup6-090-000	Cup	85.79	98.07	16.41	C	1		
Cup6-090-180	Cup	87.19	97.53	16.63	C	1		
Cup7-090-000	Cup	66.16	90.18	14.82	C	1		
Cup7-090-180	Cup	69.52	93.08	14.64	C	1		
Cup8-090-000	Cup	88.49	98.95	18.73	C	1		
Cup8-090-180	Cup	89.07	98.92	18.12	C	1		

**Table 6.4.7** (cont.) - Categorization results achieved by using the test objects, in right (090-000) and left (090-180) profile views. For the winning category (Wcat) results, the correspondences are as follows: A – apple, Ca – car, Co – cow, C – cup, D – dog, H – horse, P – pear, T – tomato. For each object, we show the necessary threshold percentage ( $NT_P$ ) that the model needed in order to provide a final categorization result. Also, the average  $NT_P$  for each category and view was calculated.

Object's name	Intersected categories (IC), and respective similarity percentages[%]				Wcat	$NT_P$	$A_{NT_P}$ LP	$A_{NT_P}$ RP
	IC	Concept 1	Concept 2	Concept 3				
Dog3-090-000	Dog	84.83	97.85	19.74	D	1	0.996	0.998
Dog3-090-180	Dog	88.26	98.42	22.45	D	1		
Dog4-090-000	Horse	77.65	96.33	21.18	H	1		
Dog4-090-180	Horse	75.10	94.47	23.34	H	1		
Dog6-090-000	Dog	83.61	98.92	20.46	D	1		
Dog6-090-180	Dog	82.72	98.93	19.34	D	1		
Dog7-090-000	Dog	70.32	85.91	21.31	D	1		
Dog7-090-180	Dog	67.96	88.38	19.82	D	1		
Dog9-090-000	Horse	77.02	93.02	23.41	H	0.99		
Dog9-090-180	Horse	78.50	94.86	21.80	H	0.98		
Horse4-090-000	Horse	84.13	96.51	18.78	H	0.99	0.988	0.992
Horse4-090-180	Horse	83.55	98.41	17.95	H	1		
Horse6-090-000	Horse	88.02	99.41	18.15	H	1		
Horse6-090-180	Horse	91.43	99.65	19.15	H	1		
Horse7-090-000	Horse	81.02	94.24	23.22	H	0.97		
Horse7-090-180	Horse	76.27	93.01	21.80	H	0.94		
Horse8-090-000	Horse	79.73	99.06	18.64	H	1		
Horse8-090-180	Horse	83.54	99.39	19.86	H	1		
Horse9-090-000	Horse	82.95	97.93	22.12	H	1		
Horse9-090-180	Horse	86.93	98.14	21.48	H	1		
Pear2-090-000	Pear	72.07	96.64	16.16	P	1	1	1
Pear2-090-180	Pear	89.04	99.29	17.37	P	1		
Pear6-090-000	Pear	89.84	98.50	17.34	P	1		
Pear6-090-180	Pear	91.72	99.22	16.84	P	1		
Pear7-090-000	Pear	88.80	97.73	19.51	P	1		
Pear7-090-180	Pear	83.53	98.05	19.79	P	1		
Pear8-090-000	Pear	92.97	99.03	17.84	P	1		
Pear8-090-180	Pear	91.90	99.34	16.64	P	1		
Pear10-090-000	Pear	85.40	97.48	20.53	P	1		
Pear10-090-180	Pear	91.13	99.28	15.77	P	1		
Tomato1-090-000	Tomato	87.20	98.37	18.62	T	1	1	1
Tomato1-090-180	Tomato	86.03	97.95	17.68	T	1		
Tomato3-090-000	Tomato	81.17	95.65	17.01	T	1		
Tomato3-090-180	Tomato	84.11	94.57	18.65	T	1		
Tomato4-090-000	Tomato	81.01	96.48	13.73	T	1		
Tomato4-090-180	Tomato	77.80	96.64	14.75	T	1		
Tomato5-090-000	Tomato	85.53	97.97	16.88	T	1		
Tomato5-090-180	Tomato	82.65	97.85	15.94	T	1		
Tomato7-090-000	Tomato	83.26	94.35	21.00	T	1		
Tomato7-090-180	Tomato	82.30	94.90	20.59	T	1		

**Table 6.4.8** - Categorization results achieved by using the test objects, in right (090-000) and left (090-180) profile views: average  $NT_P$  and categorization success rate for each view.

Average $NT_P$		Categorization success rate (CSR)	
$A_{LP}$	$A_{RP}$	$A_{LP}$	$A_{RP}$
$\approx 0.998$	$\approx 0.997$	92.5%	95%
Average CSR: 93.75%			

The categorization model was also tested in terms of small invariance to rotation, Gaussian noise and scale. These tests were performed using as base the right profile view. These results are shown in Tables 6.4.9 to 6.4.14.

**Table 6.4.9** – Results achieved when testing the categorization model in terms of rotation invariance, more specifically, using the previous template sets for right profile to identify test objects slightly rotated (tilt rotation of -22°) from the latter position.

Object name	Final categorization result	$NT_P$	$A_{NT_P}$
Apple6-068-000	Apple	1	0.994
Apple7-068-000	Apple	1	
Apple8-068-000	Apple	1	
Apple9-068-000	Apple	1	
Apple10-068-000	Apple	0.97	
Car3-068-000	Car	0.88	0.890
Car6-068-000	Car	0.86	
Car11-068-000	Car	0.91	
Car12-068-000	Car	0.87	
Car14-068-000	Car	0.93	
Cow3-068-000	Cow	1	1
Cow4-068-000	Cow	1	
Cow8-068-000	Cow	1	
Cow9-068-000	Cow	1	
Cow10-068-000	Cow	1	
Cup2-068-000	Cup	1	1
Cup5-068-000	Cup	1	
Cup6-068-000	Cup	1	
Cup7-068-000	Cup	1	
Cup8-068-000	Cup	1	
Dog3-068-000	Dog	1	0.990
<u>Dog4-068-000</u>	<u>Horse</u>	<u>1</u>	
Dog6-068-000	Dog	1	
Dog7-068-000	Dog	0.98	
<u>Dog9-068-000</u>	<u>Horse</u>	<u>0.97</u>	
Horse4-068-000	Horse	1	0.998
Horse6-068-000	Horse	1	
Horse7-068-000	Horse	0.99	
Horse8-068-000	Horse	1	
Horse9-068-000	Horse	1	
Pear2-068-000	Pear	1	1
Pear6-068-000	Pear	1	
Pear7-068-000	Pear	1	
Pear8-068-000	Pear	1	
Pear10-068-000	Pear	1	
Tomato1-068-000	Tomato	1	1
Tomato3-068-000	Tomato	1	
Tomato4-068-000	Tomato	1	
Tomato5-068-000	Tomato	1	
Tomato7-068-000	Tomato	1	
<b>Categorization success rate:</b>			<b>Global average:</b>
95%			0.984



**Table 6.4.10** – Results achieved when testing the categorization model in terms of Gaussian noise (GN) invariance, more specifically, using the previous template sets for right profile to identify test objects in right profile which were previously subjected to a 10% GN rate.

Objects subjected to a 10% GN rate	Final categorization result	$NT_P$	$A_{NT_P}$
Apple6-090-000	Apple	1	1
Apple7-090-000	Apple	1	
Apple8-090-000	Apple	1	
Apple9-090-000	Apple	1	
Apple10-090-000	Apple	1	
Car3-090-000	Car	1	0.998
Car6-090-000	Car	0.99	
Car11-090-000	Car	1	
Car12-090-000	Car	1	
Car14-090-000	Car	1	
Cow3-090-000	Cow	1	0.990
Cow4-090-000	Cow	1	
Cow8-090-000	Cow	0.95	
Cow9-090-000	Cow	1	
Cow10-090-000	Cow	1	
Cup2-090-000	Cup	1	1
Cup5-090-000	Cup	1	
Cup6-090-000	Cup	1	
Cup7-090-000	Cup	1	
Cup8-090-000	Cup	1	
Dog3-090-000	Dog	1	0.996
Dog4-090-000	Horse	1	
Dog6-090-000	Dog	1	
Dog7-090-000	Dog	1	
Dog9-090-000	Horse	0.98	
Horse4-090-000	Horse	0.98	0.990
Horse6-090-000	Horse	1	
Horse7-090-000	Horse	0.97	
Horse8-090-000	Horse	1	
Horse9-090-000	Horse	1	
Pear2-090-000	Pear	1	1
Pear6-090-000	Pear	1	
Pear7-090-000	Pear	1	
Pear8-090-000	Pear	1	
Pear10-090-000	Pear	1	
Tomato1-090-000	Tomato	1	1
Tomato3-090-000	Tomato	1	
Tomato4-090-000	Tomato	1	
Tomato5-090-000	Tomato	1	
Tomato7-090-000	Tomato	1	
<b>Categorization success rate:</b>			<b>Global average:</b>
95%			≈0.997

**Table 6.4.11** – Results achieved when testing the categorization model in terms of scale invariance, more specifically, using the previous template sets for right profile to identify test objects in right profile which were previously subjected to a 10% size increase rate.

Objects subjected to a 10% size increase rate	Final categorization result	NT <sub>P</sub>	A <sub>NT<sub>P</sub></sub>
Apple6-090-000	Apple	1	1
Apple7-090-000	Apple	1	
Apple8-090-000	Apple	1	
Apple9-090-000	Apple	1	
Apple10-090-000	Apple	1	
Car3-090-000	Car	1	1
Car6-090-000	Car	1	
Car11-090-000	Car	1	
Car12-090-000	Car	1	
Car14-090-000	Car	1	
Cow3-090-000	Cow	1	0.984
Cow4-090-000	Cow	1	
Cow8-090-000	Horse	0.92	
Cow9-090-000	Cow	1	
Cow10-090-000	Cow	1	
Cup2-090-000	Cup	0.96	0.992
Cup5-090-000	Cup	1	
Cup6-090-000	Cup	1	
Cup7-090-000	Cup	1	
Cup8-090-000	Cup	1	
Dog3-090-000	Dog	1	0.994
Dog4-090-000	Horse	1	
Dog6-090-000	Dog	1	
Dog7-090-000	Dog	1	
Dog9-090-000	Horse	0.97	
Horse4-090-000	Horse	1	0.992
Horse6-090-000	Horse	1	
Horse7-090-000	Horse	0.96	
Horse8-090-000	Horse	1	
Horse9-090-000	Horse	1	
Pear2-090-000	Pear	1	1
Pear6-090-000	Pear	1	
Pear7-090-000	Pear	1	
Pear8-090-000	Pear	1	
Pear10-090-000	Pear	1	
Tomato1-090-000	Tomato	1	1
Tomato3-090-000	Tomato	1	
Tomato4-090-000	Tomato	1	
Tomato5-090-000	Tomato	1	
Tomato7-090-000	tomato	1	
<b>Categorization success rate:</b>			<b>Global average:</b>
92.5%			≈0.995

**Table 6.4.12** – Results achieved when testing the categorization model in terms of scale invariance, more specifically, using the previous template sets for right profile to identify test objects in right profile which were previously subjected to a 10% size decrease rate.

Objects subjected to a 10% size decrease rate	Final categorization result	NT <sub>P</sub>	$A_{NT_P}$
Apple6-090-000	Apple	1	0.964
Apple7-090-000	Apple	1	
Apple8-090-000	Apple	1	
Apple9-090-000	Apple	1	
Apple10-090-000	Pear	0.82	
Car3-090-000	Car	1	1
Car6-090-000	Car	1	
Car11-090-000	Car	1	
Car12-090-000	Car	1	
Car14-090-000	Car	1	
Cow3-090-000	Cow	1	0.992
Cow4-090-000	Cow	1	
Cow8-090-000	Horse	0.96	
Cow9-090-000	Cow	1	
Cow10-090-000	Cow	1	
Cup2-090-000	Cup	1	1
Cup5-090-000	Cup	1	
Cup6-090-000	Cup	1	
Cup7-090-000	Cup	1	
Cup8-090-000	Cup	1	
Dog3-090-000	Dog	1	0.986
Dog4-090-000	Horse	1	
Dog6-090-000	Dog	1	
Dog7-090-000	Horse	0.93	
Dog9-090-000	Horse	1	
Horse4-090-000	Horse	0.97	0.988
Horse6-090-000	Horse	1	
Horse7-090-000	Horse	0.97	
Horse8-090-000	Horse	1	
Horse9-090-000	Horse	1	
Pear2-090-000	Pear	1	1
Pear6-090-000	Pear	1	
Pear7-090-000	Pear	1	
Pear8-090-000	Pear	1	
Pear10-090-000	Pear	1	
Tomato1-090-000	Cow	0.91	0.944
Tomato3-090-000	Tomato	0.95	
Tomato4-090-000	Tomato	1	
Tomato5-090-000	Tomato	1	
Tomato7-090-000	Cup	0.86	
<b>Categorization success rate:</b>			<b>Global average:</b>
82.5%			≈0.984

**Table 6.4.13** - Categorization results comparison: joining results from Table 6.4.7 to 6.4.12, in order to compare them in terms of winning category (Wcat) and necessary threshold percentage (NT<sub>P</sub>). For the winning category results, the correspondences are as follows: A – apple, Ca – car, Co – cow, C – cup, D – dog, H – horse, P – pear, T – tomato. The global average of NT<sub>P</sub> (GA) and the categorization success rate (CSR) are also shown for each test.

Test objects (reference view: right profile)											
Object's name	Objects in right profile		Objects with a tilt variation of - 22 degrees		Objects in right profile subjected to a 10% Gaussian noise rate		Objects in right profile subjected to a 10% size increase rate		Objects in right profile subjected to a 10% size decrease rate		
	Wcat	NT <sub>P</sub>	Wcat	NT <sub>P</sub>	Wcat	NT <sub>P</sub>	Wcat	NT <sub>P</sub>	Wcat	NT <sub>P</sub>	
Apple6	A	1	A	1	A	1	A	1	A	1	
Apple7	A	1	A	1	A	1	A	1	A	1	
Apple8	A	1	A	1	A	1	A	1	A	1	
Apple9	A	1	A	1	A	1	A	1	A	1	
Apple10	A	1	A	0.97	A	1	A	1	P	0.82	
Car3	Ca	1	Ca	0.88	Ca	1	Ca	1	Ca	1	
Car6	Ca	0.99	Ca	0.86	Ca	0.99	Ca	1	Ca	1	
Car11	Ca	1	Ca	0.91	Ca	1	Ca	1	Ca	1	
Car12	Ca	1	Ca	0.87	Ca	1	Ca	1	Ca	1	
Car14	Ca	1	Ca	0.93	Ca	1	Ca	1	Ca	1	
Cow3	Co	1	Co	1	Co	1	Co	1	Co	1	
Cow4	Co	1	Co	1	Co	1	Co	1	Co	1	
Cow8	Co	0.95	Co	1	Co	0.95	H	0.92	H	0.96	
Cow9	Co	1	Co	1	Co	1	Co	1	Co	1	
Cow10	Co	1	Co	1	Co	1	Co	1	Co	1	
Cup2	C	1	C	1	C	1	C	0.96	C	1	
Cup5	C	1	C	1	C	1	C	1	C	1	
Cup6	C	1	C	1	C	1	C	1	C	1	
Cup7	C	1	C	1	C	1	C	1	C	1	
Cup8	C	1	C	1	C	1	C	1	C	1	
Dog3	D	1	D	1	D	1	D	1	D	1	
Dog4	H	1	H	1	H	1	H	1	H	1	
Dog6	D	1	D	1	D	1	D	1	D	1	
Dog7	D	1	D	0.98	D	1	D	1	H	0.93	
Dog9	H	0.99	H	0.97	H	0.98	H	0.97	H	1	
Horse4	H	0.99	H	1	H	0.98	H	1	H	0.97	
Horse6	H	1	H	1	H	1	H	1	H	1	
Horse7	H	0.97	H	0.99	H	0.97	H	0.96	H	0.97	
Horse8	H	1	H	1	H	1	H	1	H	1	
Horse9	H	1	H	1	H	1	H	1	H	1	
Pear2	P	1	P	1	P	1	P	1	P	1	
Pear6	P	1	P	1	P	1	P	1	P	1	
Pear7	P	1	P	1	P	1	P	1	P	1	
Pear8	P	1	P	1	P	1	P	1	P	1	
Pear10	P	1	P	1	P	1	P	1	P	1	
Tomato1	T	1	T	1	T	1	T	1	Co	0.91	
Tomato3	T	1	T	1	T	1	T	1	T	0.95	
Tomato4	T	1	T	1	T	1	T	1	T	1	
Tomato5	T	1	T	1	T	1	T	1	T	1	
Tomato7	T	1	T	1	T	1	T	1	C	0.86	
CSR:	<b>GA:</b>	CSR:	<b>GA:</b>	CSR:	<b>GA:</b>	CSR:	<b>GA:</b>	CSR:	<b>GA:</b>	CSR:	<b>GA:</b>
	95%	≈ 0.997	95%	0.984	95%	≈0.997	92.5%	≈0.995	82.5%	≈0.984	

**Table 6.4.14** - Summary of the main categorization results regarding the tests of rotation, GN and scale invariance.

<b>Test</b>	<b>Categorization success rate</b>	<b>Global average of the necessary threshold percentage (GA)</b>
(a) Reference: testing objects images in right profile	95%	≈ 0.997
(b) Rotation invariance test (-22°, tilt)	95%	0.984
(c) Gaussian noise test (+10%)	95%	≈0.997
(d) Scale invariance test (+10%)	92.5%	≈0.995
(e) Scale invariance test (-10%)	82.5%	≈0.984

# Bibliography

## B

Bar, M., 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15(4), 600–609.

Bar, M., Kassam, K., Ghuman, A., Boshyan, J., Schmid, A., Dale, A., Hämäläinen, M. S., Marinkovic, K., Schacter, D., Rosen, B., Halgren, E., 2006. Top-down facilitation of visual recognition. *Proc. 1st Int. Visual Informatics Conf. Bridging research and practice* 103 (2), 449–454.

Bebis, G.; Egbert, D.; Shah, M. (2003) Review of Computer Vision Education. *IEEE Trans. on Education*, February 2003, 46(1), pp. 2-21.

Belongie, S.; Malik, J.; Puzicha, J. (2001) Matching shapes. *Proc. 8th IEEE Int. Conf. on Computer Vision*, 1: 454 - 461.

Biederman, I. (1995) “Visual object recognition” in Chapter 4 of: *An invitation to cognitive science*, vol. 2, *Visual Cognition*, (S. F. Kosslyn and D. N. Osherson, eds.), pp. 121-165, MIT press, 1995.

Bornstein, M.H.; Arterberry, M.E.; Mash, C. (2010) Infant object categorization transcends diverse object–context relations. *Infant Behavior and Development*, February 2010, 33(1), pp.7–15, Elsevier.

Brady, T.F.; Konkle, T.; Alvarez, G.A.; Oliva, A. (2008) Visual long-term memory has a massive storage capacity for object details. *Proc. of the National Academy of Sciences of the United States of America*; editor: D. Purves; September 2008; 105 (38), 14325–14329.

Brown, R. (1958) How shall a thing be called? *Psychological Review*, 65(1):14–21, January 1958.

Buckley, M.J.; Sigala, N. (2010) Is Top-Down Control from Prefrontal Cortex Necessary for Visual Categorization? *Neuron* - 27 May 2010, 66(4), pp. 471-473, Elsevier.

## C

Chelazzi, L.; Miller, E.K.; Duncan, J.; Desimone, R. (2001) Responses of Neurons in Macaque Area V4 During Memory-guided Visual Search. *Cerebral Cortex (Oxford Journals)*; 11 (8), pp.761-772.

Chen, S.; Akselrod, P.; Zhao, B.; Carrasco, J.A.P.; Linares-Barranco, B.; Culurciello, E. (2012) Efficient Feedforward Categorization of Objects and Human Postures with Address-

Event Image Sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, February 2012, 34 (2), 302-314.

## D

Deco, G.; Rolls, E.T. (2004) A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, March 2004, 44(6), 621-642, Elsevier.

Duchenne, O.; Joulin, A.; Ponce, J. (2011) A Graph-Matching Kernel for Object Categorization. *Proc. IEEE Int. Conf. Computer Vision*, pp.1792-1799.

Duda, R.O.; Hart, P.E.; Stork, D.G. (2001) *Pattern Classification*, John Wiley & Sons, New York, 2nd Edition.

## E

Elazary, L.; Itti, L. (2008). Interesting objects are visually salient. *J. of Vision*, March 2008, 8(3), article 3, pp.1–15.

ETH-80 (2013), <http://www.d2.mpi-inf.mpg.de/Datasets/ETH80>, accessed in 21/03/2013.

Evans, C. (2009) Notes on the OpenSURF Library. CSTR-09-001, University of Bristol, January 2009. Available at <http://www.chrisevansdev.com>.

## F

Farivar, R. (2009) Dorsal-ventral integration in object recognition. *Brain Research Reviews*, October 2009, 61(2), pp.144-153, Elsevier.

Fei-Fei, L.; Fergus, R.; Perona, P. (2004) Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *CVPRW '04: Computer Vision and Pattern Recognition Workshop*, 2004. Pp. 178.

Fisiologia ocular (2013), <http://sentidos5espsmm.blogspot.com/2008/01/fisiologia-ocular-parte-2.html>, accessed in 21/03/2013.

## G

Galleguillos, C.; Belongie, S. (2010) Context based object categorization: A critical survey. *Computer Vision and Image Understanding*, June 2010, 114(6), 712–722, Elsevier Science Inc. New York, NY, USA.

Gegenfurtner, K.R.; Kiper, D.C.; Levitt, J.B. (1997) Functional Properties of Neurons in Macaque Area V3. *Journal of Neurophysiology*, April 1997, 77 (4), pp. 1906-1923.

Goldstone, R.L.; Lippa, Y.; Shiffrin, R.M. (2001) Altering object representations through category learning. *Cognition* 78:27–43.

Grigorescu, C.; Petkov, N.; Westenberg, M.A. (2003) Contour Detection Based on Nonclassical Receptive Field Inhibition. *Proc. IEEE Trans. on Image Processing*, July 2003, 12(7), pp.729-739.

Grossberg, S.; Srinivasan, K.; Yazdanbakhsh, A. (2011) On the road to invariant object recognition: How cortical area V2 transforms absolute to relative disparity during 3D vision. *Neural Networks*, September 2011, 24(7), 686-92, Elsevier.

## H

Han, X.-H.; Chen, Y.-W.; Ruan, X. (2011) Canonical correlation analysis of local feature set for view-based object recognition. 2011 Proc. 18th IEEE Int. Conf. on Image Processing, pp.3601–3604.

Hubel, D.H. (1995) *Eye, Brain and Vision* (Scientific American Library, No 22), published by W. H. Freeman; available in: <http://hubel.med.harvard.edu/index.html>.

## J

Jiang, X.; Bradley, E.; Rini, R.A.; Zeffiro, T.; VanMeter, J.; Riesenhuber, M. (2007a) Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron* - 15 March 2007, 53(6), pp.891-903, Elsevier.

Jiang, Y.-G.; Ngo, C.-W.; Yang, J. (2007b) Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval. *Proc. of the 6th ACM international conference on Image and video retrieval*, pp. 494-501, ACM New York, NY, USA.

## K

Kim, E.; Medioni, G. (2011) Scalable Object Classification in Range Images. *Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 65–72.

Kinnunen, T.; Kamarainen, J.-K.; Lensu, L.; Kälviäinen, H. (2010) Unsupervised visual object categorisation via self-organisation. *Proc. 20th Int. Conf. on Pattern Recognition*, pp. 440–443.

Kourtzi, Z.; Connor, C.E. (2011) Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. *Annual Review of Neuroscience*, July 2011, 34:45–67.



## L

Lam, R.; du Buf, J.M.H. (2011) Retrieval of 3D Polygonal Objects Based on Multiresolution Signatures. Lecture Notes in Computer Science - Advances in Visual Computing: 7th International Symposium, ISVC 2011, Las Vegas, NV, USA, September 26-28, 2011. Proc., Part II; 6939:136-147; editors: G. Bebis, R. Boyle, B. Parvin, D. Koracin, S. Wang, K. Kyungnam, B. Benes, K. Moreland, C. Borst, S. DiVerdi, C. Yi-Jen, J. Ming; Springer-Verlag GmbH Berlin Heidelberg, Springer Berlin Heidelberg.

Lee, T.S. (1996) Image representation using 2D Gabor wavelets. IEEE Trans. on Pattern Analysis and Machine Intelligence, October 1996, 18(10), pp. 959-971.

Leibe, B., Schiele, B. (2003) Analyzing Appearance and Contour Based Methods for Object Categorization. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, II-409-15, vol. 2.

Leibe, B.; Schiele, B. (2004) Scale-Invariant Object Categorization Using a Scale-Adaptive Mean-Shift Search. Lecture Notes in Computer Science - Pattern Recognition: Proc. 26th DAGM Symposium, Springer LNCS 3175, pp.145-153.

Logothetis, N.K.; Pauls, J.; Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current Biology, May 1995, 5 (5), pp. 552–563, Elsevier.

Lowe, D.G. (1999) Object recognition from local scale-invariant features. Proc. 7th IEEE Inter. Conf. on Computer Vision, 1999; 2:1150–1157.

Lowe, D.G. (2001) Local Feature View Clustering for 3D Object Recognition. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 1:682-688.

## M

Martins, J.A.; Rodrigues, J.; du Buf, J.M.H. (2009). Focus of Attention and Region Segregation by Low-level Geometry. Proc. of the 4th Int. Conf. on Computer Vision Theory and Applications, Lisboa, Portugal, 2:267-272.

Miller, E. (2000). The prefrontal cortex and cognitive control. Nature Rev.Neuroscience, 1(1), pp. 59-65.

Mundhenk, T.N.; Navalpakkam, V.; Makaliwe, H.; Vasudevan, S.; Itti, L. (2004) Biologically inspired feature based categorization of objects. SPIE Proc. - Human Vision and Electronic Imaging IX (HVEI04), San Jose, CA; editors: B.E. Rogowitz, T.N. Pappas; 5292:330-341.

Mutch, J.; Lowe, D.G. (2008) Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields. Int. J. of Computer Vision, 80(1), pp.45-57.

## N

Nene, S.A.; Nayar, S.K.; Murase, H. (1996) Columbia Object Image Library (COIL-100). Technical Report CUCS-006-96, February 1996.

## O

Olshausen, B.A.; Field, D.J. (2005). How Close Are We to Understanding V1? *Neural Computation - MIT Press Journals*, August 2005, 17 (8), pp. 1665-1699.

## P

Palmer, S.E. (1975) The effects of contextual scenes on the identification of objects, *Memory & Cognition*, September 1975, 3(5):519-526, Springer-Verlag.

Passalis, G.; Kakadiaris, I.A.; Theoharis, T. (2007) Intra-class Retrieval of Nonrigid 3D Objects: Application to Face Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, February 2007, 29(2), 218-229.

Petre, R.-D.; Zaharia, T. (2011) Still Image Object Categorization Using 2D Object Models. *Proc. 5th IEEE Int. Conf. on Semantic Computing (ICSC)*, pp.419-423.

Pinz, A. (2006) Object Categorization. *Foundations and Trends in Computer Graphics and Vision*. September 2006, 1(4), 2005, pp.255-353, Now Publishers Inc. Hanover, MA, USA.

## Q

Qiu, F. T.; von der Heydt, R. (2005). Figure and Ground in the Visual Cortex: V2 Combines Stereoscopic Cues with Gestalt Rules. *Neuron*, July 2005, 47 (1), pp.155-166, Elsevier.

## R

Rensink, R.A. (2000) The Dynamic Representation of Scenes. *Visual Cognition*, 7 (1-3), 17–42.

Rodrigues, J.; du Buf, J.M.H. (2006) Multi-scale keypoints in V1 and beyond: Object segregation, scale selection, saliency maps and face detection. *BioSystems*, October–December 2006, 86(1-3), 75-90; *Brain, Vision and Artificial Intelligence (BVAi) 2005: Papers from a Symposium on Brain Basics and Natural Vision*; Elsevier.

Rodrigues, J.; du Buf, J.M.H. (2009a) A cortical framework for invariant object categorization and recognition. *Cognitive Processing*, August 2009, 10(3), pp. 243-261, Springer-Verlag.

Rodrigues, J.; du Buf, J.M.H. (2009b) Multi-scale lines and edges in V1 and beyond: Brightness, object categorization and recognition, and consciousness. *BioSystems*, March 2009, 95(3), pp. 206-226, Elsevier.

Rodrigues, J.; Lam, R.; du Buf, H. (2012) Cortical 3D Face and Object Recognition Using 2D Projections. *International Journal of Creative Interfaces and Computer Graphics*, January 2012, 3(1), pp. 45-62, IGI Publishing Hershey, PA, USA.

Rodrigues, J.M.F. (2007) Integrated multi-scale architecture of the cortex with application to computer vision. Ph.D. Thesis in Electronic and Computer Engineering, Major in Computer Science. Faculty of Sciences and Technology, University of the Algarve, Faro.

Rodrigues, J.M.F.; du Buf, J.M.H. (2011) A cortical framework for scene categorization. *Proc. of VISAPP 2011 - International Conference on Computer Vision Theory and Applications*, pp. 364-371.

Rodrigues, J.M.F.; Lam, R.; du Buf, J.M.H. (2011) Cortical 3D Face Recognition Framework. *SIACG 2011 - Proc. V Ibero-American Symposium in Computer Graphics*, Faro, Portugal (pp.81-87).

Rosch, E.; Mervis, C.B.; Gray, W.D.; Johnson, D.M.; Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439, July 1976, Elsevier Inc., Elsevier B.V.

Ruesch, J.; Lopes, M.; Bernardino, A.; Hörnstein, J.; Santos-Victor, J.; Pfeifer, R. (2008) Multimodal Saliency-Based Bottom-Up Attention: A Framework for the Humanoid Robot iCub. *Proc. IEEE Int. Conf. on Robotics and Automation*, pp.962–967.

## S

Saleiro, M.A.N. (2011) *Visão Activa para Robô Cognitivo*. MSc Thesis in Electrical and Electronics Engineering. Instituto Superior de Engenharia, Universidade do Algarve, Faro.

Savarese, S.; Fei-Fei, L. (2010) Multi-View Object Categorization and Pose Estimation. In *Studies in Computational Intelligence - Computer Vision: Detection, Recognition and Reconstruction*, 285:205-231; editors: R. Cipolla, S. Battiato, G.M. Farinella; Springer-Verlag Berlin Heidelberg, Springer Berlin Heidelberg.

Schiele, B.; Crowley, J.L. (2000) Recognition without Correspondence using Multidimensional Receptive Field Histograms. *Int. J. of Computer Vision (IJCV)*, 36(1):31–50, January 2000, Kluwer Academic Publishers.

Serre, T.; Oliva, A.; Poggio, T. (2007) A feedforward architecture accounts for rapid categorization. *Proc. of the National Academy of Sciences of the United States of America*, 104(15), pp. 6424–6429.

Shilane, P.; Min, P.; Kazhdan, M.; Funkhouser, T. (2004) The Princeton Shape Benchmark. Proc. Int. Conf. on Shape Modeling and Applications, Genova, Italy.

Sousa, R.J.R. (2009) Correlatos neuronais no reconhecimento de emoções faciais em humanos. MSc Thesis in Electrical and Electronics Engineering. Instituto Superior de Engenharia, Universidade do Algarve.

Stöttinger, J.; Hanbury, A.; Sebe, N.; Gevers, T. (2012) Sparse Color Interest Points for Image Retrieval and Object Categorization. IEEE Trans. on Image Processing, May 2012, 21(5), pp.2681-2692.

Su, T.-M.; Lin, C.-C.; Lin, P.-C.; Hu, J.-S. (2006) Shape Memorization and Recognition of 3D Objects Using a Similarity-Based Aspect-Graph Approach. Proc. IEEE Int. Conf. on Systems, Man and Cybernetics, 6:4920-4925.

Swain, M.J.; Ballard, D.H. (1991) Color indexing. International Journal of Computer Vision (IJCV), 7(1):11–32, November 1991, Kluwer Academic Publishers.

## T

Tangelder, J.W.H.; Velkamp, R.C. (2008) A survey of content based 3D shape retrieval methods. Multimedia Tools and Applications, September 2008, 39(3), pp.441-471, Springer US.

Troncoso, X.G.; Macknik, S.L.; Martinez-Conde, S. (2011) Vision's First Steps: Anatomy, Physiology, and Perception in the Retina, Lateral Geniculate Nucleus, and Early Visual Cortical Areas. Visual Prosthetics - Physiology, Bioengineering, Rehabilitation; pp 23-57; editor: G. Dagnelie; Springer Science+Business Media, LLC; Springer US.

Turk, M.; Pentland, A. (1991) Eigenfaces for Recognition. J. of Cognitive Neuroscience, 3(1):71–86, 1991.

## V

VanRullen, R.; Thorpe, S. J. (2001) “Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects”, Perception, 30(6), pp.655-668, 2001.

## W

Weldon, T. P.; Higgins, W. E.; Dunn, D. F. (1996) Gabor filter design for multiple texture segmentation. In Optical Engineering (SPIE Journal), 35 (10), pp.2852-2863.

Wu, L.; Hu, Y.; Li, M.; Yu, N.; Hua, X.-S. (2009) Scale-Invariant Visual Language Modeling for Object Categorization. IEEE Trans. on Multimedia, February 2009, 11(2), 286-294.

## Z

Zaharia, T.; Prêteux, F.(2004) 3D versus 2D/3D Shape Descriptors: A Comparative study, in Proc. Conf. on Image Processing: Algorithms and Systems, Vol. 2004 , Toulouse, France, January 2004.