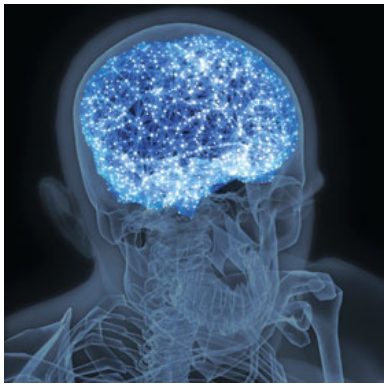
TÉCNICO LISBOA



# Sensorimotor and Neural Networks

for Visual Stimuli Prediction

**Ricardo Manuel Raposo dos Santos**

Thesis to obtain the Master of Science Degree in

## Electrical and Computer Engineering

Supervisor(s): Professor Alexandre José Malheiro Bernardino

### Examination Committee

Chairperson: Professor João Fernando Cardoso Silva Sequeira
Supervisor: Professor Alexandre José Malheiro Bernardino
Member of the Committee: Professor Luis Henrique Martins Borges de Almeida

**October 2014**

# Acknowledgments

Firstly, I would like to express my deep gratitude to Professor Alexandre Bernardino for proposing this thesis, for his advices and support since I first met him as teacher in the course of Modelling and Simulation, and for his patience and assisting through the development of this work.

A special thanks to Ricardo Ferreira and Ângelo Cardoso for sharing their time, experience and ideas, making this thesis richer and enjoyable.

To Fundação para a Ciência e Tecnologia (FCT) and project BIOMORPH for financial support.

To my colleagues for their companionship and fun relaxing lunches.

I would also like to thank my parents Carla e Franciso, my brother Rodrigo and my girlfriend Cristina who unconditionally supported and cheered me, not only in this work but also through the course.

# Resumo

Os seres humanos e outros animais desenvolveram sistemas visuais com retinas organizadas de maneiras bastante distintas, todas elas diferentes das câmaras convencionais de hoje em dia. A aprendizagem automática é normalmente utilizada no processamento de imagens e reconhecimento e/ou identificação de padrões, objectos e faces, mas frequentemente só considerando dados visuais. Esta tese foca-se numa arquitectura inspirada na biologia, denominada Sensorimotor Network, capaz de co-desenvolver ambas estruturas sensorial e motora directamente de dados adquiridos por um agente interagindo com o seu ambiente. Estas estruturas conduzem a um modelo capaz de prever eficientemente estímulos visuais baseados na percepção sensorial e acções realizadas por um agente. O sistema visual sensoriomotor proposto é treinado num ambiente estático e comparado com feedforward Neural Networks comuns (lineares e não lineares), mostrando melhores capacidades preditivas e custos computacionais que as últimas. Motivado pela diversidade organizações de retinas existentes na Natureza, as estruturas sensoriomotoras resultante são interpretadas e a sua relação é explicada. A interdependência de características visuomotoras de um agente e o seu meio envolvente tem um profundo impacto na organização das topologias sensoras e motoras. Adicionalmente, a Sensorimotor Network é treinada para ser invariante ao brilho e uma retina é desenvolvida utilizando dados reais de um drone. No fim, as vantagens de co-desenvolver sistemas que considerem ambas informações sensorial e motora deverão ser claras. Ultimamente, um robot equipado com tal potencial poderia gerar consciência das suas capacidades motoras, adaptar-se ao seu ambiente e estar um passo mais perto de ser independente.

**Palavras-chave:** Predição de estímulo, estruturas sensorimotoras, redes neuronais, campos receptores

## Abstract

Humans and other animals developed visual systems with retinas organized in very distinctive ways, all different from nowadays conventional cameras. Machine learning is usually used for image processing and pattern, object or faces recognition and/or identification, but often only considering visual data. This thesis focus on a biological inspired architecture, denoted as Sensorimotor Network, able to co-develop both sensor and motor structures directly from data acquired by an agent interacting with its environment. These structures lead to a model capable of efficiently predict an agent's self-induced visual stimuli based on sensory perception and performed actions. Here the proposed visual sensorimotor system is trained in a static environment and compared with standard feedforward Neural Networks (linear and non-linear) showing better prediction capabilities and computational costs than the latter. Motivated by the diversity of retinal organizations existing in Nature, the resulting sensorimotor structures are interpreted and their relationship is explained. The interdependency of visuomotor characteristics of an agent and its surroundings is proved to have a deep impact on sensor and motor topologies organization. Additionally, Sensorimotor Network is trained to reconstruct contours and a retina is developed using real data acquired using a drone. In the end, the advantages of co-developing systems which consider both sensory and motor information should be clear. Ultimately, a robot equipped with such potential could originate motor self-awareness, adapt to its environment and be a step closer to independence.



**Keywords:** Stimulus prediction, sensorimotor structures, neural networks, receptive fields

# Contents


Acknowledgments . . . iii
Resumo . . . v
Abstract . . . vii
List of Tables . . . xi
List of Figures . . . xiii
Nomenclature . . . 1
Glossary . . . 1

**1 Introduction** **1**
1.1 Related Work . . . 3
1.1.1 Neural Networks . . . 3
1.1.2 Sensorimotor Networks . . . 6
1.2 Contribution . . . 8
1.3 Outline . . . 9

**2 Sensorimotor Prediction Architectures** **10**
2.1 An agent life-long experience . . . 11
2.2 Multilayer Perceptron . . . 12
2.3 Sensorimotor Network . . . 13
2.4 Sensorimotor Neural Network . . . 15
2.4.1 Multiplication Block . . . 15
2.4.2 Model Optimization . . . 16

**3 Experimental Setup** **18**
3.1 Data Set Configuration . . . 18
3.2 Models Configuration . . . 20
3.3 Statistical Comparison . . . 21
3.3.1 Computational Cost . . . 21
3.3.2 Information Criteria . . . 21

**4 Results** **23**
4.1 Sensorimotor Network vs Multilayer Perceptron – Comparison . . . 24

4.1.1 Prediction Error . . . 24
4.1.2 Visual Stimuli Reconstruction . . . 25
4.1.3 Comparison Summary . . . 27
4.2 Sensorimotor Network - Connectivity and Properties . . . 27
4.2.1 Sensor and Motor Structures . . . 28
4.2.2 Predictive Structure . . . 29
4.2.3 Sensor Visual Receptive Fields Influence . . . 31
4.2.4 Environment Influence . . . 32
4.3 Sensorimotor Network - Drone . . . 33

**5 Conclusions** **37**
5.1 Achievements . . . 37
5.2 Future Work . . . 38

**Bibliography** **43**

# List of Tables

1.1 Artificial neuron activation functions . . . . . 4

4.1 Models comparison summary: Experiments ExpXY and ExpRZ . . . . . 27

# List of Figures


1.1 Sketch of a biological neuron . . . 3
1.2 Feedforward Neural Network - MLP . . . 5
1.3 Nature Visual Systems Examples . . . 7
1.4 Model of corollary discharge circuit . . . 7

2.1 Agent experiences acquisition process . . . 12
2.2 Multilayer Perceptron: schematic diagram . . . 13
2.3 Sensorimotor Network: schematic diagram . . . 14
2.4 Multiplication Block . . . 16
2.5 Sensorimotor Network using Neural Network Toolbox: schematic diagram . . . 16

3.1 Motor Actions: Degrees of Freedom . . . 19
3.2 Motor Actions: Discretized action spaces . . . 20

4.1 RMSE per Pixel: Experiments ExpXY and ExpRZ . . . 24
4.2 Visual Stimuli Reconstruction: Experiment ExpXY . . . 26
4.3 Visual Stimuli Reconstruction: Experiment ExpRZ . . . 26
4.4 Sensor and Motor Structures Evolution: Experiment ExpXY . . . 28
4.5 Sensor and Motor Structures Evolution: Experiment ExpRZ . . . 28
4.6 Predictor, Motor and Sensor structures relationship . . . 30
4.7 Predictive Structure: Actions influence on visual stimuli . . . 30
4.8 Sensor Visual Receptive Fields: Influence on Prediction . . . 31
4.9 Environment influence on visual sensor topology . . . 32
4.10 RMSE of different visual sensor topologies . . . 33
4.11 Sensorimotor Real Data: Drone and Environment . . . 33
4.12 Images acquired using a Drone . . . 34
4.13 Drone: Visual stimuli prediction . . . 35
4.14 Drone: Sensors organization and prediction errors . . . 36

# Chapter 1

# Introduction

## Contents


**1.1 Related Work . . . . . 3**

1.1.1 Neural Networks . . . . . 3

1.1.2 Sensorimotor Networks . . . . . 6

**1.2 Contribution . . . . . 8**

**1.3 Outline . . . . . 9**


At the present time, computers are powerful machines capable of solving mathematical problems which a human being cannot, and it is possible to build robots which exalt in manufacturing facilities by cutting, molding, assembling and placing many distinctive parts and materials, in a rapid and precise manner. However, while a computer can easily calculate a series of mathematical operations, it can't recognize and/or identify other humans, objects or patterns as fast and accurately as a human being does. In addition, even by outperforming humans in so many areas, robots are not yet able to efficiently adapt to their surroundings and are only used in structured environments built specifically so they can perform their tasks safely, free from any unpredictability in the robots point of view.

Not only humans, but many organisms can perform numerous tasks according to their physiological morphology and adaptation to the natural environment. Most of them possess a brain or nervous system (more or less complex) which controls and coordinates all voluntary and involuntary actions by transmitting signals between the different parts of their body. This type of coordination arises from neurons or nervous cells organization, which result from organism's experiences throughout its living, by performing actions and gathering sensory feedback. In addition, the way the organism behaves is of course constrained by its physical capabilities, available sensory organs and surroundings.

The importance of computationally co-develop structures of the nervous system is justified by Nature's ability to produce organisms, from the simplest to most complex, which can efficiently perform countless different tasks to ensure their survivability. It would be an outstanding achievement to be able to produce robots which could be self-aware of their constitution (sensors and actuators), adapt and automatically behave and act to accomplish a specific goal within the environment where they are deployed. Achieving such autonomy for a robot begins with deeply understanding its sensor and motor

relationship. Actually, when programming a machine to automatically act upon a specific environment an individual imposes limitations through his knowledge and perception of how the robot should execute its actions. Having the robot freely and successfully adapting to its given sensors and actuators, would create a much more efficient information management (only meaningful stimuli would be considered) and consequently allow easier decision making.

At the same time as Nature shows countless examples of successful organisms which can prosper even in the most harsh conditions all thanks to their structural, behaviour and physiological adaptation; in our daily lives and in our own environment we can automatically and effortlessly perform tasks like obstacle avoidance, people recognition and object manipulation. Rats, for instance, use touch (whiskers) for object recognition [2]; bats use their auditory capabilities for echolocation (hearing self-produced ultrasound waves) [25] and anteaters use their excellent sense of smell for foraging, feeding and defense. All these skills are easily fulfilled thanks to the brain's ability to process many different inputs which trigger tightly related ideas and memories. Yet, without really understanding which steps a brain takes to reach these results, giving a robot such efficient abilities using preplanned sequential routines is still not possible.

Nowadays, most of robots are equipped with 2D or 3D cameras, which give them the ability to access the most important sense for the majority of diurnal animal, vision. These cameras are capable of acquiring different types of images but robots still lack the ability of autonomously process the given sensory information to decide whether and how to act. In Nature, visual systems can be dramatically different from species to species. This variety is essential to understand their own behaviour and how their actions affect the surroundings. Less complex animals tend to develop more specifically adapted visual sensorimotor systems, while those with larger nervous systems possess structures capable of supporting multiple usage strategies. For instance, in the one hand a small crustacean Copepod *Copilia* has a visual system composed by two sensors with 7 photoreceptors each [17]. On the other hand, the retina in each human eye has two types of photoreceptors: about 92 million rods and 4.6 million cones [7]. All in all, each one has developed in order to help or allow efficient recognition of food, threat, mate and shelter, and ultimately survive given the available resources.

Within the diversity of existing visual sensorimotor systems, all share efficiency in stimuli processing. This processing capabilities are at some extended consequence of predictive characteristics of organisms nervous systems, where sensory and motor information are correlated to process future stimuli. The stimuli prediction becomes important to create the ability to distinguish auto-induced stimuli caused by self-motor actions (*reafference*) from external signals with origin in the environment (*exafference*) [29]. The ability to discern between these two origins of sensory input requires a forward model [24] to predict the effect a given movement (action) has on an organism's sensory input, helping it to recognize the consequences of its actions by associating specific stimuli changes to its behaviour. In a more directioned point of view, visual sensorimotor predictions can be observed in rapid motor sequences such as saccadic movements in primates [37], supported by the recording of presaccadic frontal eye field neurons (visual, visuomotor and motor cells). These experiments concluded that after a period where the primates hold their gaze on a single fixation point reappearing from time to time, about 30% of visual

and visuomotor cells presented predictive visual responses. This way it was shown that brain relies on a predictive control strategies, where motor commands are fired and visual receptive fields (future receptive fields) already yield stimuli before the action is performed.

The main goal of this thesis is to train simple visual and motor structures capable of partially reproduce their interrelationship as in the human brain so that they can efficiently predict visual stimuli given sensory and motor feedback. Moreover, the importance of relating stimuli from both sources become explicit. Two possible architectures are trained and compared: a Multilayer Preceptron (MLP) and an adaptive model [30] here called Sensorimotor Network (SNet). The first is trained without any distinction between motor and sensory data, while the second receives both stimuli separately in each dedicated structure, merging the processed information through a predictive layer.

## 1.1 Related Work

### 1.1.1 Neural Networks

Since the $19^{th}$ century scientists have tried to deeply understand the way the human brain works, and nowadays there is already one approach in the field of machine learning which is inspired in the biological neural networks of the brain and their structural and functional aspects, allowing a computer or a robot to learn by example, the Artificial Neural Networks (ANNs).

This model mimics the neuronal interconnections which make it suitable to solve highly parallel problems such as object or patterns recognition, data classification, image compression, stock market prediction and some applications in medicine, security and loans.

In 1890, William James published his first work regarding brain activity patterns, and later, in 1943, the first artificial neuron was created by the neurophysiologist Warren McCulloch and the logician Walter Pits [23]. A biological neuron [21] consists of a cell body containing a nucleus and cytoplasm, dendrites which receive messages from other neurons and an axon which conducts electrical impulses away from the neuron's cell body. These impulses pass through the synapses which are directly connected to other neurons. A neuron representation can be observed in Figure 1.1.

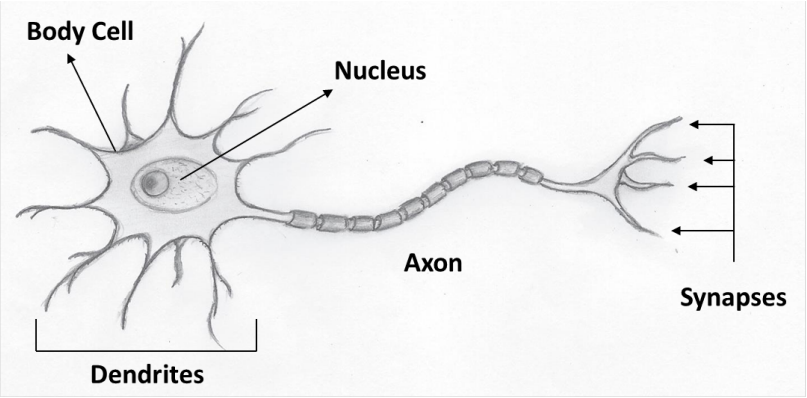



Figure 1.1: Sketch of a biological neuron and its structures.

As for biological nervous systems, neural networks are built from single unit artificial neurons. While a neuron receives information through electrical impulses of others' dendrites, its artificial equivalent has numeric values as inputs. These values are then weighted as synapses modulate the electrical impulses from dendrites. Finally, while in neurons if the signal is strong enough an output is fired, in its model after summing the weighted values an activation function is applied and a numeric output is generated. As presented in [4], the output of the artificial model of the neuron can be mathematically computed as

$$\begin{aligned} y &= f\left(\phi\left(x, w\right)\right) \\ \phi\left(x, w\right) &= w_0 + \sum_{j=1}^{p} w_j x_j \end{aligned} \tag{1.1}$$

which is related with the input values $x = (x_1, ..., x_p)$ and each $x_j$ is fed and then multiplied by respective weight $w_j$. After adding all resulting products the activation function $f$ is applied. This function $f$ can be linear or non-linear, typically one from those presented in Table 1.1.

Properties like synaptic interconnection, activation functions and training or learning are keys for neural networks success. Although, years after McCulloch and Pits proposed the artificial neuron, neural network theory has been somewhat forgotten. In the 80s researchers started to realize its potential when supported by the growing interest in human cognition for Artificial Intelligence applications and the rapid increase of computer processing power. At that time many works based on neural networks were issued with special relevance to Fukushima's work at digit recognition [10] and Sejnowski's work at teaching a network to pronounce English written words [33].

Usually, to solve a problem, a set of instructions is given to a computer based on how the user interprets it and plans its resolution. This means one shall understand how to solve the problem and only then each step can be programmed so that computer reaches the expected solution. Neural networks afford a number of highly interconnected neurons which work in parallel to solve a particular problem. These connections can't be directly programmed for a specific task since they adapt to the given set of examples. The data used to train a neural network (inputs and corresponding outputs) has to be carefully chosen and must represent some kind of dependency or relationship, otherwise time will be unnecessarily lost and the network may not adapt nor organize correctly.

Neurons interconnections are modeled by fully connecting each neural network layer (array of ar-

| | | |
|---|---|---|
| **Linear** | $f(u) = u$ | Output is not affected. |
| **Non-Linear** | $f(u) = sgn(u)$ | Output can be -1 or +1. |
| | $f(u) = sgn(u) + 1/2$ | Output is binary 1 or 0. |
| | $f(u) = (1 + e^{-u})^{-1}$ | Output ranges between 0 and 1 (sigmoidal) |
| | $f(u) = tanh(u)$ | Output ranges between -1 and 1 (hyperbolic tangent) |
| | $f(u) = (u)_+$ | Output is non-negative. |

Table 1.1: Artificial neuron activation functions.

tificial neurons) sequentially, where each neuron from layer *n* receives all outputs from layer *n-1*, and each neuron of layer *n+1* receives all outputs from layer *n*. This kind of neural network is denominated feedforward neural network and when composed by two or more layers it is denominated by Multilayer Perceptron (Figure 1.2). These networks can be trained using an optimization method called backpropagation [14]. This applies another method such as gradient descent which computes the gradient of a cost function (e.g. measuring error between trained and predicted outputs) and iteratively feeds the backpropagation for weight matrices update.

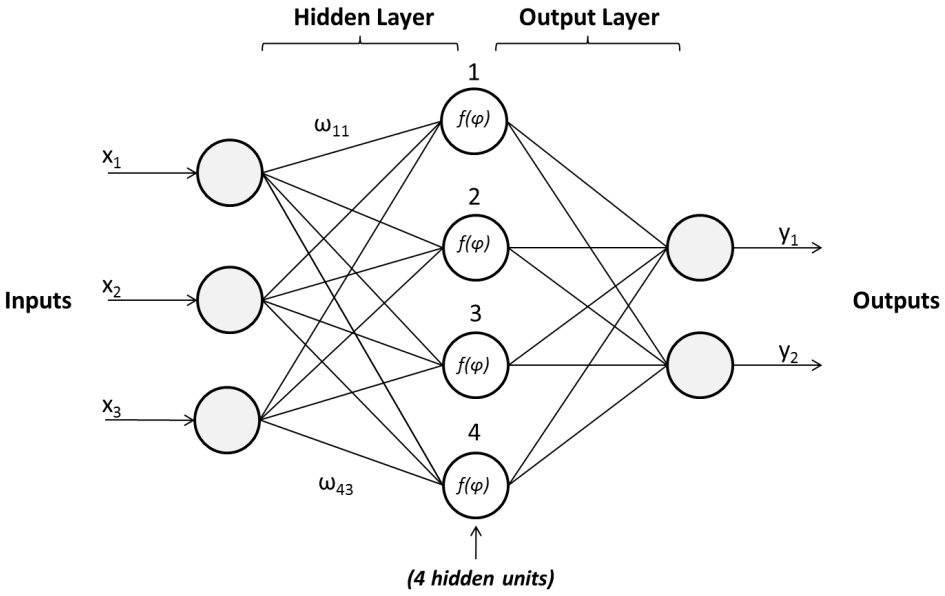



Figure 1.2: Feedforward neural network or multilayer perceptron with a single hidden layer.

As the brain organizes itself by readjusting the many neurons' synaptic connections for better efficiency in information processing, the ANNs weight the connections between all their elements creating a solution which translates relationships that data they were fed with can yield. This type of model is more suitable to solve highly parallel problems such as vision or speech recognition which for instance human brain solves easily due to the ability of process many different inputs which trigger conflicting ideas and memories. ANNs are still not good enough to compete with a brain, however they already give computers the capability of learning by example and are effectively used in applications of object recognition, patterns recognition or data classification. Examples of these applications are:

- Speech generation and recognition
- Automatic recognition of handwritten characters
- Recognition of coins of different denominations
- Identification of cancerous cells [38]
- Recognition of chromosomal abnormalities [20]
- Prediction of financial indices such as currency exchange rates, and others. [36]

### 1.1.2 Sensorimotor Networks

As mentioned before, robotic systems still fail to deduce appropriate actions in a non-artificial environment since it is too complicated for them to create self-awareness of their sensory and motor signals and produce motor actions suited to each situation. According to [30] this fault results from the lack of cognitive skills or, in another point of view, the high complexity of sensor and motor relationship becoming too difficult to be translated by the robot. In this perspective the author favors the concept of simple brains where sensor systems should be adapted to motor systems and vice-versa. Without perception one is left with little criteria to decide which actions to take, while at the same time there is no purpose in having perception if one cannot act on the world. An ideal rational agent [26] always takes the actions which maximizes its performance measure based on its percepts and built-in knowledge. While the sensor should provide stimuli meaningful for the agent's motor capabilities and environment, the motor system should perform actions which lead to significant results for the sensor stimuli. In a sense, the sensor and motor systems should work synergistically, helping each other in the execution of the tasks.

While neural networks are usually used as a tool for pattern recognition, and consequently a possible suitable method for image prediction, when applied to this type of task the processed data focus mainly on image features and visual data, often neglecting the importance of motor data in agent's behaviour. This thesis proposes the recently developed adaptive model of Sensorimotor Network [30] as a path to follow for better image processing and development of retina-like structures. This author's approach considers the interconnection between different areas of the brain (namely the visual and motor areas) and its adaptive properties that optimize the sensor, motor and predictive structures to the agents morphology and environment characteristics, in terms of predictive ability and computational efficiency.

Evidences supporting Sensorimotor Networks exist in biological systems showing that it is not desirable for an animal to have an oversized brain if not needed and nervous systems fight high energy consumption ([18]) by adapting their morphology and physiology. Besides, considering Nature's versatility and the amount of different existing visual sensorimotor systems such as those presented in Figure 1.3, it becomes obvious that allowing a robot to develop a visuomotor system would greatly improve its adaptivity to its physical constraints and the environment it would act upon.

Biologically, a structure called superior colliculus, SC, (denomination of optic tectum for mammals) corresponds to the brain area where visual sensory stimuli is mapped onto motor layers which controls eye and body orientation [34]. A Corollary Discharge Circuit [6] represents a pathway from the SC to frontal eye field (FEF) via the media dorsal nucleous (MD), which conducts motor activations generated in the SC's deep layer to the FEF in a feedforward direction. On FEF both motor signals and visual signals from the main sensory processing stream are integrated, and from shifting visual receptive fields a prediction of future visual stimuli is created.

In Figure 1.4 a simple representation of a visual corollary discharge circuit is presented. The Sensorimotor Network mimics the visuomotor relationship presented in this circuit, where motor neurons code visual saccades (rapid eye movements used for visual fixation) in a retinotopic reference from (1.4d) whose activations are collected from deep motor layers (1.4c) and projected through feedforward connections (1.4b) to sensor area (1.4a).

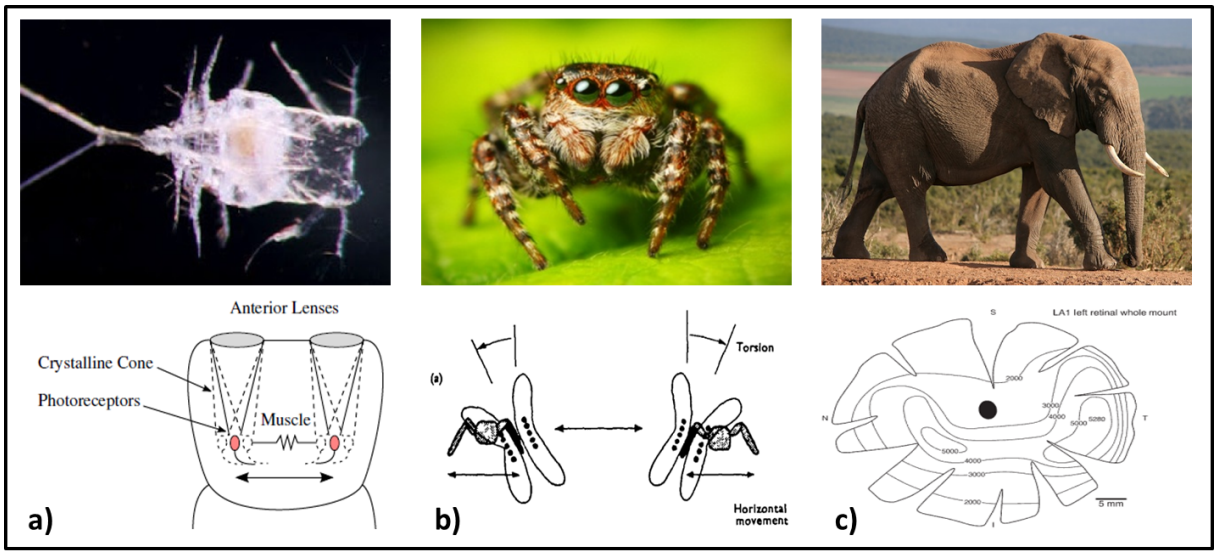



Figure 1.3: Nature visual systems examples: a) Plankton Copilia based on [30], b) Jumping Spider based on [16] and c) African Elephant based on [35]

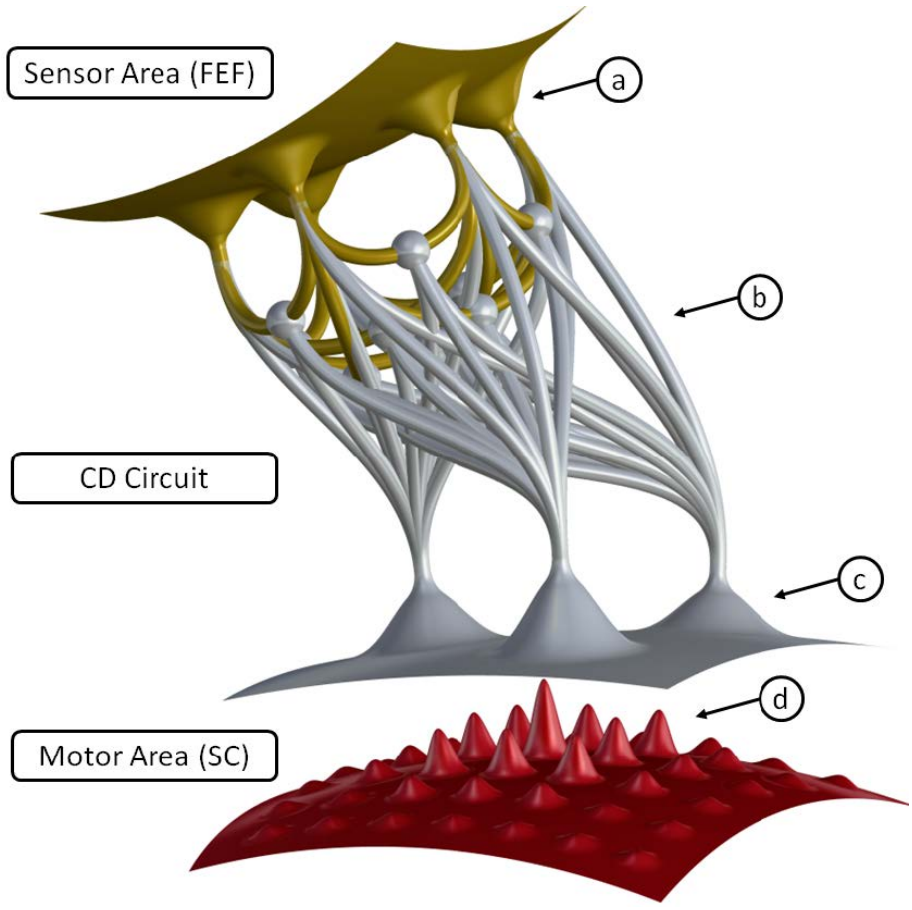



Figure 1.4: Model of a visual corollary discharge circuit. A population of motor neurons coding visual saccades (d) in the motor area of SC. An intermediate layer of CD neurons (c) collects motor activations which follow feedforward connections (b) and are projected in the sensor area (a). The corollary discharge signals modulate the activation of visual receptive fields and their connections such as to predict a future visual stimulus resulting from an activation in (a) [31].

Applying this biological neural evidence allowed the proposed model model to be trained to minimize the error in visual stimuli prediction by developing self-organizing sensorimotor structures of artificial visual systems. The predicted visual stimulus is then obtained based on sensor and motor information perceived by each corresponding layer. An agent's life-long experience and self induced actions create adaptive co-developed structures responsible for its nervous system efficiency.

Motor actions are encoded and mapped in a structure composed by movement fields which react to actions producing similar perception changes. These also influence visual stimulus processing, creating direct relationship between the robot's actions and its perceived visual stimuli.

In the sensory layer through the same developmental process an organization also emerges. This layer is formed by visual receptive fields which gather a set of retina cells covering nearby parts of the visual field and together represent a continuous portion of it.

Following a specific learning process it is possible to minimize the prediction error evaluated by the mean square error between the predicted image and the expected image after a specific motor action. Additionally, even starting from an unknown connectivity, the proposed structures organize themselves considering the recorded visuomotor stimuli and leading to a less costly prediction model. This simultaneous development promotes a coherent representation for similar stimuli (sensory) and actions (motor), which greatly improves the effectiveness of the network.

As shown later on, the developmental process yields better sensory predictions for the effects of actions, when compared to a more naive and straightforward approach which lacks a sensorimotor structure, supporting the importance of coupling sensory and motor information.

## 1.2 Contribution

This thesis contributed to deepen the understanding of co-developing sensory and motor structures, more specifically a visual sensorimotor system which takes robotics a step closer for vision efficiency and autonomy. The emerging structure topologies as well as the prediction capabilities of the adaptive model are compared with a multilayer perceptron on several levels.

With respect to previous work progress was made in the following points:

- Analysis of each resulting sensor, motor and predictive structures after sensorimotor model training and their interdependency.
- Exploration of the adaptive visual sensorimotor approach in distinct scenarios, showing its adaptivity to sensory stimuli perceived on artificial and natural environments. These results are also implemented in a quadricopter drone in an open environment.
- Comparison between the sensorimotor system and a single hidden layer MLP (with linear and non-linear activation functions) which don't distinguish sensory and motor information. This shows the advantages of a proper organisation of the sensorimotor structures following biological principles.
- Deployment of the proposed adaptive model in a modified neural network resorting to algebraic

properties and implemented on the Matlab Neural Network Toolbox. This opening space for further exploratory work on sensorimotor networks using state-of-the-art software simulation tools.

Based on this work, a paper *Sensori-motor Network vs Neural Network in Visual Stimuli Prediction* was submitted and accepted for a poster presentation in ICDL-EPIROB - The Fourth Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, 2014.

## 1.3 Outline

In order to create a system for visual stimuli prediction, in Chapter 2 two possible architectures are presented: a Neural Network and a Sensorimotor Network. The later is also reproduced in a modified neural network from Matlab's Toolbox. In Chapter 3, the data sets and experimental setups used throughout this thesis are explained. In Chapter 4, the standard feedforward neural network approach and the sensorimotor system are compared in terms of prediction error, computational cost and information loss. Additionally, the sensor and motor structures developed in Sensorimotor Networks are studied in terms of visual perception and environment dependency together with an example using real data taken from a quadricopter. In the end, in Chapter 5 the conclusions are stated with this thesis achievements and possible future work.

# Chapter 2

# Sensorimotor Prediction Architectures

## Contents


**2.1 An agent life-long experience** . . . . . **11**

**2.2 Multilayer Perceptron** . . . . . **12**

**2.3 Sensorimotor Network** . . . . . **13**

**2.4 Sensorimotor Neural Network** . . . . . **15**

2.4.1 Multiplication Block . . . . . 15

2.4.2 Model Optimization . . . . . 16


The importance of computationally co-develop structures of the nervous system is justified by the Nature ability to produce organisms, from the simplest to most complex, which can efficiently perform countless different tasks to ensure their survivability. Lately, the relationship between computational science and biology is becoming tighter and tighter [5]. The resulting increment use of biological characteristics and discoveries in neuroscience, have been inspiring fields as machine learning to create innovative approaches for tasks that are known to be hard for a computer to perform (learning to identify a cat by itself [19] or use concepts as perception, memory and judgement [9]).

Following biology, and in view of the complexity of the human brain and its ability to readjust after sensor changes (sight or hearing loss) and/or motor changes (limb loss or partial paralysis), it makes sense to try computationally mimic such a complex characteristic. The nervous system is know for its neuroplasticity, which means it can reorganize its neuronal pathways as a response to internal or external changes [15, 27]. This work focus on a simplistic computational approach modeling the interconnection between the visual and motor areas of the human brain and its adaptive characteristics. Moreover, the brain prediction power based on its organism life experiences and resulting adaptation is computationally tested using the mentioned Multilayer Perceptron and Sensorimotor Networks approaches.

## 2.1 An agent life-long experience

Although an organism is usually equipped with more than one sensory system, like visual, auditory, tactile and olfactory, there are many ways of nervous systems from different species to process these senses (hear different frequencies and/or see different wavelengths, for instance).

Throughout this thesis, a computational agent is presented as an example of an organism equipped with a visual and a motor system. It is considered capable of observing its environment by sensing a continuous light field of intensities *i* which falls on a two dimensional sensory surface. Similarly this agent is able to interact with its environment by activating a particular motor primitive *q* on its action space.

For implementation purposes, the light field (sensor input) is represented as a vector $\mathbf{i}$ of $N_s$ pixels, and the action space (motor input) is represented as a vector $\mathbf{q}$ with $N_m$ elements, where a single non-zero entry represents the activated motor primitive. If the $n^{th}$ index of $\mathbf{q}$ is 1, then the physical action $\mathbf{u}$ coded by that index is performed (e.g. shift left by a certain amount). As an example, for an action space $\mathbf{U}$ composed by $k = 9$ different combinations of horizontal and vertical translational movements,

$$\mathbf{U} = \{(-x,+y),(-x,0),(-x,-y),(0,+y),(0,0),(0,-y),(+x,+y),(+x,0),(+x,-y)\} \tag{2.1}$$

where $x$ and $y$ are constants which are setup dependent and the motor primitive $\mathbf{q}$,

$$\mathbf{q} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \tag{2.2}$$

would represent the action $(-x, 0)$. Even though this represents the performed action, the agent only experiences the effect the action will produce in the visual stimulus, which means it never accesses the effective action $\mathbf{u} = (-x, 0)$. In the same way, we move our head without knowing exactly which and how muscles moved, but we can realize that movement by sensory feedback such as changes in what we are seeing.

During its life-long experience, an organism has its nervous system adapted due to brain plasticity. The synaptic pathways rearrange themselves considering genetic, biological and environmental changes to optimize brain's efficiency. The agent's learning phase occurs during all its experiences. Without any topological assumption of its sensor and motor structures, and regardless its state, the agent interacts with the environment by randomly choosing a motor primitive $\mathbf{q}$ coding the action $\mathbf{u}$ while collecting before and after sensor stimuli ($\mathbf{i}_0$ and $\mathbf{i}_1$).This process is illustrated in Figure 2.1 gathering a data set of triplet experiences ($\mathbf{i}_0$,$\mathbf{i}_1$,$\mathbf{q}$) which are collected in a static forest environment.

During an organism life-span it learns to predict certain patterns from repetition on its action's consequences. Here two possible architectures are trained to model an agent's nervous system capable of using these patterns from performed actions and sensory feedback for visual stimuli prediction. This considered, from an agent interacting with its environment, the trained models are compared regarding 1) their predicting capabilities, i.e. how well they can predict $\mathbf{i}_1$ given $\mathbf{i}_0$ and $\mathbf{q}$; 2) their simplicity, computational cost and information loss, i.e. the number of parameters learned which contribute to prediction and the model ability to adapt to the learning data.

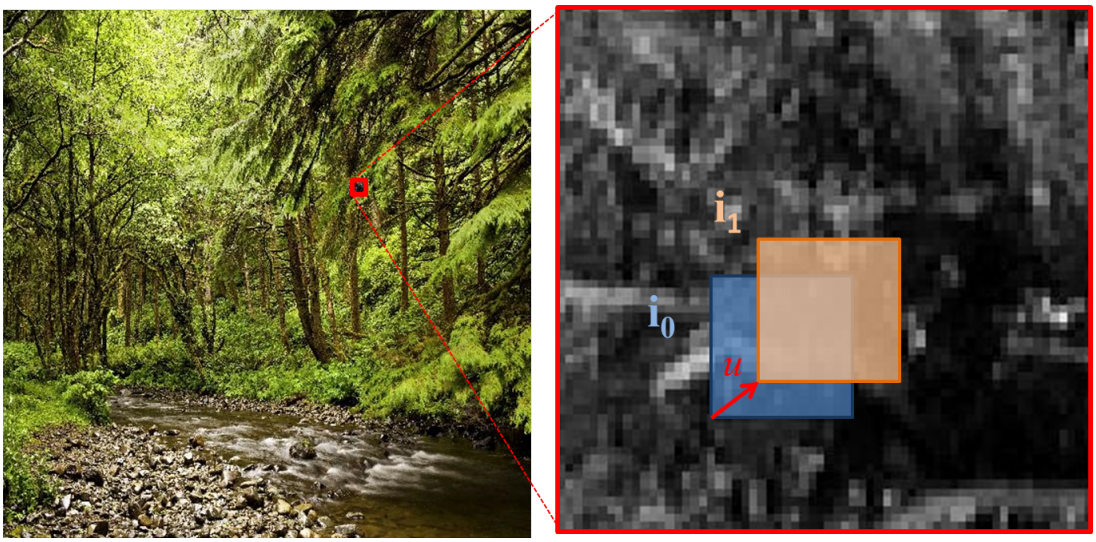



Figure 2.1: Agent experiences acquisition process: a) In the left it is shown the full environment image where the agent performs its exploration. b) In the right it is presented a portion of the environment where the agent is placed to acquire the pre-action visual stimulus $\mathbf{i}_0$ and, after being moved by action $\mathbf{u}$, acquire the post-action image $\mathbf{i}_1$ (Best seen in color).

## 2.2 Multilayer Perceptron

The neural network architecture considered is a Multilayer Perceptron (MLP) with a single hidden layer. Here sensory and motor information are jointly considered and processed by a $n_s$ elements hidden layer emulating neurons where stimuli are encoded. Then each sensor input $\mathbf{i}_0$ is concatenated with an action activation vector $\mathbf{q}$ (working as an action identifier), for each experienced triplet, and is used as input to the network predicting the target image $\mathbf{i}_1$. A set of *k* actions with *L* experiences each is considered. The optimization problem to solve can be written as,

$$
\begin{aligned}
(\mathbf{W}_1^*, \mathbf{W}_2^*) &= \operatorname{argmin} \sum_k \left\| \mathbf{i'}_1^k - \mathbf{i}_1^k \right\|^2, \\
\mathbf{o}^k &= f\left( \mathbf{W_1} \begin{bmatrix} \mathbf{i}_0^k \\ \mathbf{q}^k \\ 1 \end{bmatrix} \right) \\
\mathbf{i'}_1^k &= f\left( \mathbf{W_2} \begin{bmatrix} \mathbf{o}^k \\ 1 \end{bmatrix} \right)
\end{aligned}
\tag{2.3}
$$

and is represented in Figure 2.2. Here, $\mathbf{W_1}$ is an $(N_m + N_s + 1) \times n_s$ matrix, and $\mathbf{W_2}$ is $(n_s + 1) \times N_s$, where each includes a constant bias term. Considering equation 1.1, the activation function $f$ is $f(x) = x$ when linear or $f(x) = tanh(x)$ when non-linear, applied to $\phi_1(\mathbf{i}_0^k, \mathbf{q}^k)$ and $\phi_2(\mathbf{o}_0^k)$. Matrix $\mathbf{W}_1$ is required to force the dimensionality reduction emulating the existence of receptive fields, and $\mathbf{W}_2$ for image prediction and reconstruction.

Although a single hidden layer is used in this architecture, it is important to state this expects to mimic an unique step on image processing in human nervous system, without any consideration regarding the stimuli origin (sensor or motor).

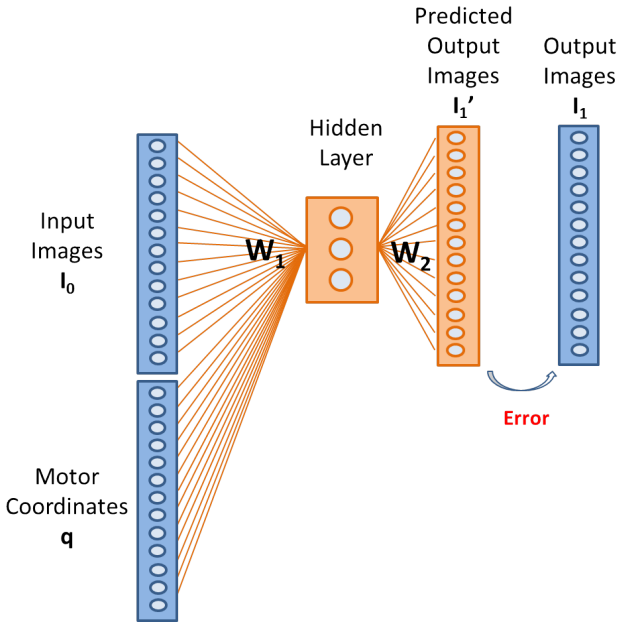



Figure 2.2: Multilayer Perceptron: schematic diagram representing the total data triplets ($\mathbf{I}_0$,$\mathbf{I}_1$,$\mathbf{q}$) used to train the model (blue) and the trained parameters $(\mathbf{W}_1^*, \mathbf{W}_2^*)$ and stimuli prediction $\mathbf{I'}_1$ (orange).

## 2.3 Sensorimotor Network

The here called Sensorimotor Network proposed in [31] explicitly models the existence of a sensory structure equipped with $n_s$ light sensitive receptors (receptive fields) which integrate visual signal $\mathbf{i}$, discretized on $N_s$ pixels, falling on a two dimensional sensory surface, $\phi_s$. The sensory structure is then represented by an $n_s \times N_s$ matrix $\mathbf{S}$ where each row represents each visual receptive field. From the integration of $\mathbf{i}$ by $\mathbf{S}$, the agent has access to a simpler observation $\mathbf{o}$ of the actual visual stimulus received in the sensory area,

$$\mathbf{o} = S\mathbf{i} \tag{2.4}$$

On the motor side a dual structure exists, where a set of $n_m$ discrete motor movement fields integrate the motor signal $\mathbf{q}$ of size $\mathbf{N_m}$ from continous motor space $\phi_m$, providing a dimensional motor action representation space,

$$\mathbf{a} = M^T\mathbf{q} \tag{2.5}$$

where $\mathbf{M}$ is a $N_m \times n_m$ matrix and stands for a topological representation of the motor structure. Each action representation is then fed to a predictive layer, where a predictor $\mathbf{P}_k$ is computed as a linear combination of $n_m$ basis predictors $\mathbf{P}_j$ with linear weights given by the motor movement field activations,

$$\mathbf{P}_k = \sum_j^{n_m} \left(\mathbf{m}_j^T \mathbf{q}^k\right) \mathbf{P}_j \tag{2.6}$$

and $\mathbf{m}_j^T$ represents the $j^{th}$ column of **M** and each motor movement field.

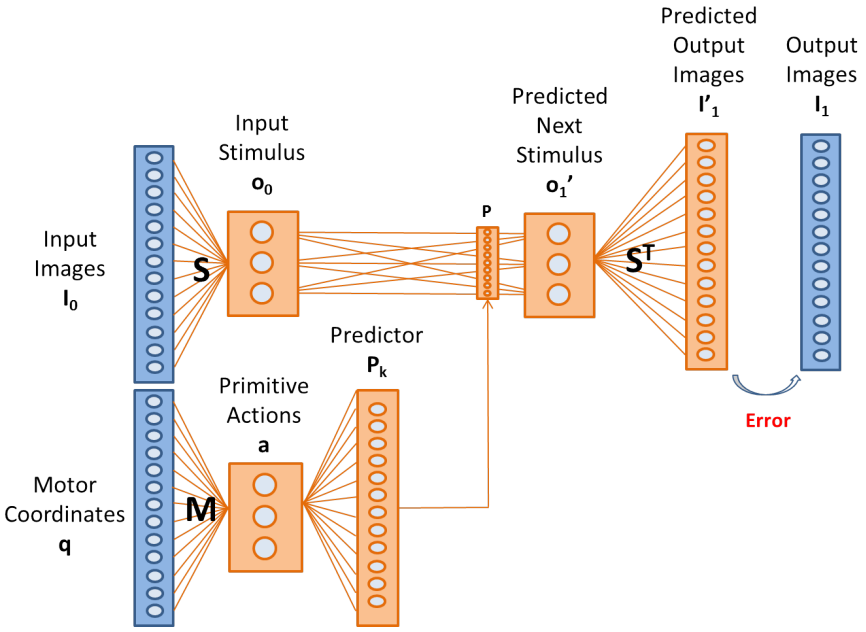



Figure 2.3: Sensorimotor Network: schematic diagram representing data triplets ($\mathbf{I}_0$,$\mathbf{I}_1$,$\mathbf{q}$) used to train the model (blue) and the trained parameters $(\mathbf{P}^*, \mathbf{M}^*, \mathbf{S}^*)$ and stimuli prediction $\mathbf{I'}_1$ (orange).

The full model description is provided in [30] by the resulting optimization problem in equation 2.7. Additionally, in Figure 2.3 a schematic diagram is presented for better understanding of the model structure.

$$
\begin{aligned}
(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*) &= \operatorname{argmin} \sum_k \left\| \mathbf{i'}_1^k - \mathbf{i}_1^k \right\|^2 \\
\mathbf{i'}_1^k &= \mathbf{S}^T \left( \sum_j^{n_m} \left( \mathbf{m}_j^T \mathbf{q}^k \right) \mathbf{P}_j \right) \mathbf{S} \mathbf{i}_0^k \\
\mathbf{S} &\geq \mathbf{0},\ \mathbf{M} \geq \mathbf{0},\ \mathbf{P} \geq \mathbf{0}
\end{aligned}
\tag{2.7}
$$

Unlike the MLP architecture, it should be noticed that the sensor reconstruction model is simplified to be $\mathbf{S}^T$ (instead of independent projection and reconstruction matrices). In [30] the author argues that this simplification is justified by the particular solutions obtained from the model, particularly the fact that the matrix $\mathbf{S}$ will be nearly orthogonal. The Sensorimotor Network approach also considers two important properties: positivity constraints to tackle $\mathbf{P}$ and $\mathbf{M}$ ambiguity and sparsity for computational efficiency.

In a biological point of view and revisiting the Figure 1.4 in Chapter 1, the proposed structures and their connectivity correspond directly to the information processing from the deeper motor layers of the superior colliculus ($\mathbf{M}$), passing by the corollary discharge circuits ($\mathbf{P}$), to the sensory receptive fields on the frontal eye field ($\mathbf{S}$).

## 2.4 Sensorimotor Neural Network

Having in mind the goal of reaching better portability and the application of the proposed sensorimotor model using different optimization methods, this section presents a successful deployment of a co-development model in a modified neural network.

The sesnorimotor network is re-implemented using the Neural Network Toolbox from Matlab, which does not allow to directly produce the optimization function presented in equation 2.7, due to the lack of a multiplication block capable of computing the multiplication which produce the predicted observations,

$$\mathbf{o'}_1^k = \left( \sum_j^{n_m} \left( \mathbf{m}_j^T \mathbf{q}^k \right) \mathbf{P}_j \right) \times \mathbf{Si}_0^k \tag{2.8}$$

To deploy the visual sensorimotor system in a neural network, a product block (smaller neural network with fixed matrix weights) was created. Since dot product and sum functions are available for weight matrices, also resorting to algebra properties and matrix manipulation it is possible to compute a matrix multiplication [22].

### 2.4.1 Multiplication Block

Two matrices are multipliable iff the number of columns of the first, A, is equal to the number of rows of the second, B. Also, the resulting matrix, C, will have the same number of rows as A and the same number of columns as B.

$$\mathrm{C}_{l\times n} = \mathrm{A}_{l\times m} \mathbf{B}_{m\times n} \tag{2.9}$$

In neural networks, each data sample (input and output) is processed as a vector throughout all its layers. In order to produce a multiplication block a small neural network is created. Its inputs are the vectorized $\mathbf{A}$ and $\mathbf{B}$ matrices, and its expected output is the vectorized $\mathbf{C}$. This network is composed by three layers and respective weight matrices $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. From this, and using the available dot product function, the mathematical equivalent of this multiplication block will be

$$vec(\mathbf{C}) = \mathbf{Z}_{(l\times n)\times(l\times m\times n)} \left[ \mathbf{X}_{(l\times m\times n)\times(l\times n)} vec(\mathbf{A}) \cdot \mathbf{Y}_{(l\times m\times n)\times(m\times n)} vec(\mathbf{B}) \right] \tag{2.10}$$

where the weight matrices $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ are fixed, $\mathbf{X}$ and $\mathbf{Y}$ expand $\mathbf{vec}(\mathbf{A})$ and $\mathbf{vec}(\mathbf{B})$, respectively, and $\mathbf{Z}$ rearranges the resulting dot product to obtain $\mathbf{vec}(\mathbf{C})$,

$$\mathbf{X} = diag\left(\mathbf{I}_n\right) \otimes \begin{bmatrix} \mathbf{I}_m \otimes \mathbf{I}_l\left(1,:\right) \\ \mathbf{I}_m \otimes \mathbf{I}_l\left(2,:\right) \\ \vdots \\ \mathbf{I}_m \otimes \mathbf{I}_l\left(l,:\right) \end{bmatrix} \tag{2.11}$$

$$\mathbf{Y} = \mathbf{I}_n \otimes [diag\,(\mathbf{I}_l) \otimes \mathbf{I}_m] \tag{2.12}$$

$$\mathbf{Z} = \mathbf{I}_{l\times n} \otimes diag(\mathbf{I}_m)^T \tag{2.13}$$

where operator $\otimes$ stands for Kronecker operator and $diag$ the diagonal of the matrix.

All this considered, from matrix manipulation and by fixing all weight matrices in the presented simple network, it was possible to create a mathematical equivalent of matrix multiplication and use it as a product block for deployment of sensorimotor system in a neural network on Matlab. In Figure 2.4 it is presented a scheme of this block.

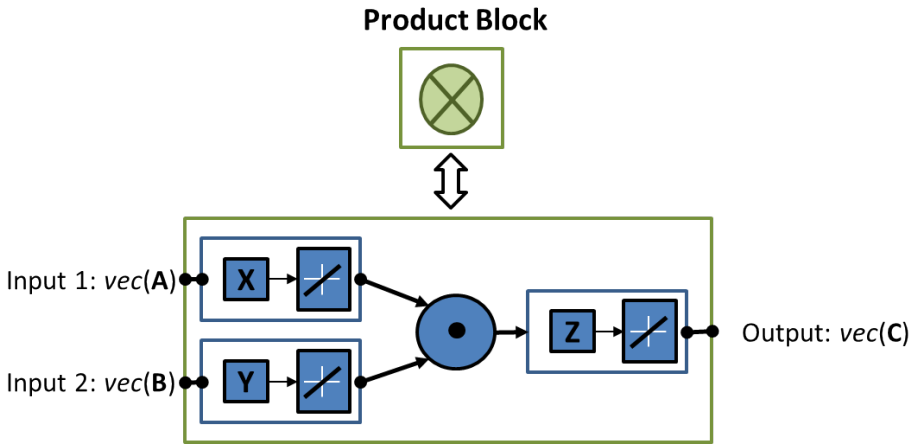



Figure 2.4: Multiplication block using Matlab's Neural Network Toolbox blocks.

### 2.4.2 Model Optimization

In order to apply the cost function given by equation 2.7, it is important to create a network with layers which model both sensor and motor topologies, $\mathbf{S}$ (and its orthogonal $\mathbf{S}^T$) and $\mathbf{M}$, respectively, as well as the predictor matrix, $\mathbf{P}$. Altogether, the developed neural network counts with 4 dynamic layers, whose weights are effectively trained, and 3 static layers from the added multiplication block. Figure 2.5 shows a representative diagram of the sensorimotor system deployed using the Neural Network Toolbox.

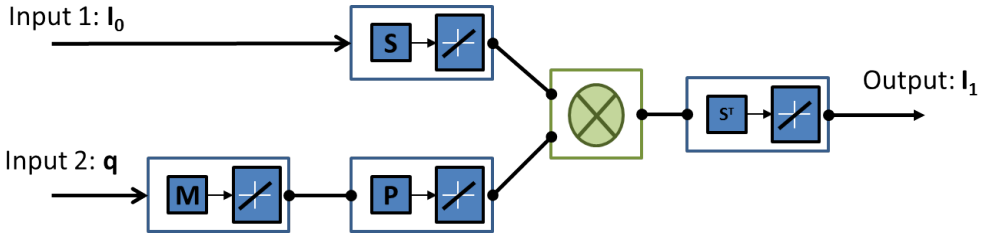



Figure 2.5: Block diagram of Sensorimotor Network approach using Matlab's Neural Network Toolbox. This produces **C**=**B***A** where $\mathbf{A} = \mathbf{o}_0$ (equation 2.5) and $\mathbf{B} = \mathbf{P}_k$ (equation 2.6).

Training the sensorimotor network with the product block previously created, and considering equations 2.11,2.12,2.13, the optimization problem can be mathematically rewritten as,

$$
\begin{aligned}
(\mathbf{S}^*, \mathbf{M}^*, \mathbf{P}^*) = \operatorname{argmin} \quad & \left\| \mathbf{S}^T \left( \mathbf{Z} \left\{ \mathbf{X}' \odot \mathbf{Y}' \right\} \right) - \mathbf{i}_1^k \right\|^2 \\
& \mathbf{X}' = \mathbf{X} vec\left(\mathbf{A}\right) = \mathbf{X} vec\left( \left[ \sum_j^{n_m} \left( \mathbf{m}_j^T \mathbf{q}^k \right) \mathbf{P}_j \right] \right) \\
& \mathbf{Y}' = \mathbf{Y} vec\left( \mathbf{S} \mathbf{i}_0^k \right) \\
& \mathbf{S} \geq \mathbf{0}, \mathbf{M} \geq \mathbf{0}, \mathbf{P} \geq \mathbf{0}
\end{aligned}
\tag{2.14}
$$

During optimization each layer is trained sequentially as in Section 2.3, with exception of $\mathbf{S}$ and $\mathbf{S}^T$ which are trained together. Matlab's Neural Network Toolbox has some limitations when trying to impose some constraints in the weight matrices. Constraints like positivity (negative values being projected to 0) and normalization (applied to $\mathbf{S}$ and $\mathbf{M}$) have to be computed after each neural network training iteration. This sensorimotor neural network is trained using the following pseudo-code, in Algorithm 1.

**Data**: Triplets $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$.
**Result**: Trained model for visual stimuli prediction.
initialization;
**for** *each sequential iteration* **do**
  train **P** = *true*;
  train **M**,$\mathbf{S}_1, \mathbf{S}_2$ = *false* ;
  **for** *each **P** iteration* **do**
    **for** *each k action* **do**
      train Network with **Data**;
      **P** = max(**P**,0);
    **end**
  **end**
  train **M** = *true*;
  train **P**,$\mathbf{S}_1, \mathbf{S}_2$ = *false* ;
  **for** *each **M** iteration* **do**
    **for** *each k action* **do**
      train Network with **Data**;
      **M** = max(**M**,0);
    **end**
  **end**
  **M** = **M**$\backslash norm(\mathbf{M})$ ;
  train $\mathbf{S}_1, \mathbf{S}_2$ = *true* ;
  train **P**,**M** = *false*;
  **for** *each **S** iteration* **do**
    **for** *each k action* **do**
      train Network with **Data**;
      $\mathbf{S}_1 = max(\mathbf{S}_1, 0)$ ;
      $\mathbf{S}_2 = max(\mathbf{S}_2, 0)$ ;
      $\mathbf{S}_1 = mean(\mathbf{S}_1, \mathbf{S}_2^T)$ ;
      $\mathbf{S}_2 = \mathbf{S}_1^T$ ;
    **end**
  **end**
  $\mathbf{S}_1 = \mathbf{S}_1 \backslash norm(\mathbf{S}_1)$ ;
  $\mathbf{S}_2 = \mathbf{S}_2 \backslash norm(\mathbf{S}_2)$ ;
**end**

**Algorithm 1:** Pseudo-Code: Sensorimotor Optimization in Neural Network

# Chapter 3

# Experimental Setup

## Contents


**3.1 Data Set Configuration . . . . . 18**

**3.2 Models Configuration . . . . . 20**

**3.3 Statistical Comparison . . . . . 21**

3.3.1 Computational Cost . . . . . 21

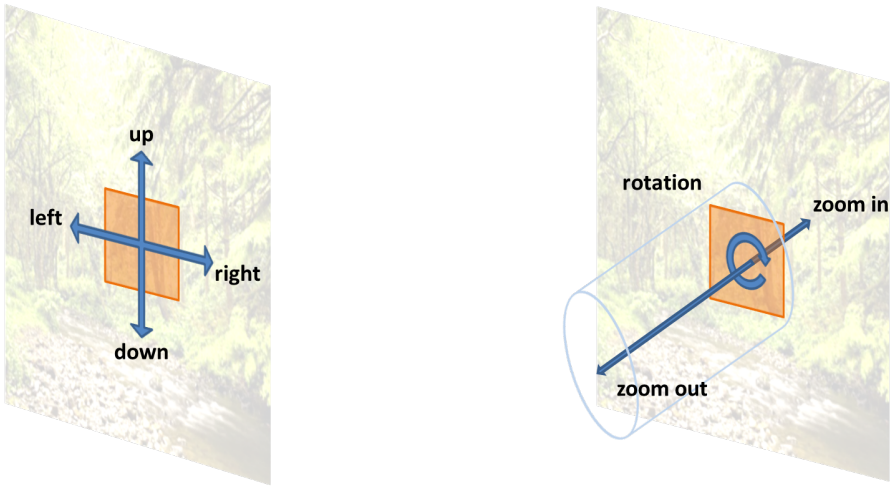
3.3.2 Information Criteria . . . . . 21


In this Section, all the performed tests and experimental apparatus are described. Initially, Sensorimotor Network is compared with Multilayer Perceptron in terms of visual stimuli prediction error, information loss and computational cost. Secondly, the proposed model is studied regarding its structures organization, topology and its dependency with the agent's environment and the way it is perceived. The sensorimotor approach is applied to a neural network using Matlab's Neural Network Toolbox trying to replicate the results from Sensorimotor Network implementation. Lastly, a sensorimotor system was trained using real data acquired by a quadricopter.

## 3.1 Data Set Configuration

In order to compare the proposed biologically inspired sensorimotor architecture, with a feedforward neural network architecture, two experiences were performed.

In the first experiment (ExpXY), the agent performs a set of actions which lead to perceived translational movements in a 2D space. The motor space spans actions leading to translations in the image plane (chosen static environment), whereas in the second experiment (ExpRZ) actions leading to centered rotations and zooms are used. The translational action space mimics an agent moving its sensor parallel to the environment surface or an agent that performs small pan-tilt rotations of the sensor when observing far objects (Figure 3.1(a)). The second set of movements can either be seen as the observations of an agent moving in a tubular structure translating and rotating along its optical axis, or the observations of an agent while actively tracking an object that rotates and changes its distance from the observer (Figure 3.1(b)).




(a) (b)

Figure 3.1: Degrees of freedom for a translational action space (a) and possible movements for a combined action space of rotations and scaling (b).

Each experiment, ExpXY and ExpRZ, was performed by training 10 models using Sensorimotor Network (SNet) and 10 models using the Multilayer Perceptron (MLP). Within the same experiment, each pair of SNet/NNet runs was executed using the same data set, in a total of 10 ActXY data sets for experiment ExpXY and 10 ActRZ data sets for experiment ExpRZ.

Each used data set is composed by three distinctive groups (a training set, a validation set for stopping criteria and a testing set):

- 1) Training Set (100 data samples per action - 8100 triplets).
- 2) Validation Set (50 data samples per action - 4050 triplets).
- 3) Testing Set (100 data samples per action - 8100 triplets).

All ($\mathbf{i}_0$,$\mathbf{i}_1$,$\mathbf{q}$) triplets for each group were uniformly sampled from a discrete set of $N_m = 81$ canonical actions with a total of 250 pairs of visual samples each (pre-action and post-action images). In ExpXY, the set of actions, ActXY, is composed of combined horizontal and vertical pixel translations $\mathbf{u} = \{-4 : 1 : 4\} \times \{-4 : 1 : 4\}$ and in ExpRZ, the set of actions, ActRZ, combine rotations and zoom scale factors transformations $\mathbf{u} = \{-100^\circ : 25^\circ : 100^\circ\} \times \{0.80 : 0.05 : 1.20\}$. Both ActXY and ActRZ have their discretized action space represented in Figure 3.2.

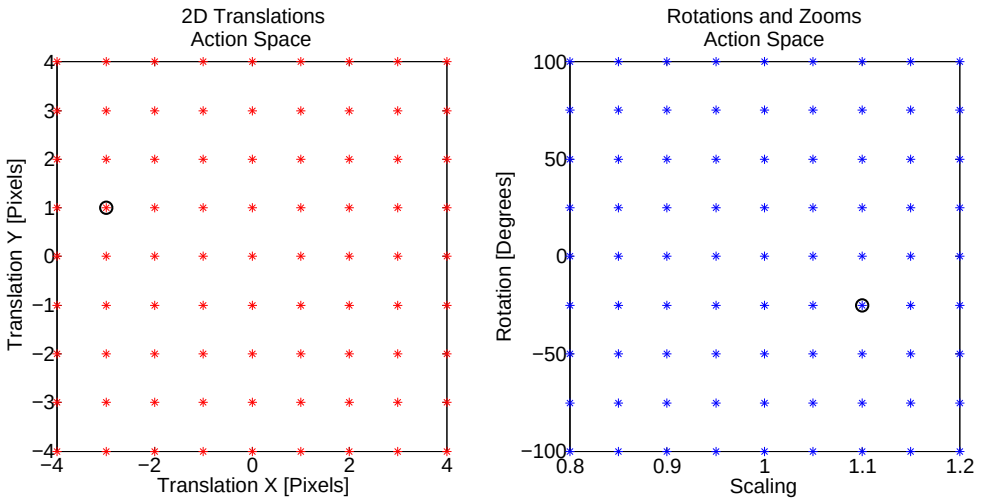



Figure 3.2: Discretized action space ActXY (on the left) and action space ActRZ (on the right).

In order to acquire sensory stimuli, the agent is equipped with a square retina of 15 by 15 pixels ($N_s = 225$) and placed in a chosen environment given by a 2448 by 2448 pixels image of a forest, as presented in Figure 2.1. The exploratory procedure copying the agent's life-long experiences is performed by positioning the agent in a random place in the environment, regardless its state, where an image $i_0$ is sampled. Then, by performing an action resulting in the identifier $q$, a new post-action image $i_1$ is sampled.

## 3.2 Models Configuration

After acquiring its exploration data the agent adapts its visual system in order to minimize the prediction error for visual stimuli prediction. From the Multilayer Perceptron model adaptation emerges from weight matrices $(\mathbf{W}_1, \mathbf{W}_2)$ being updated using a gradient descent backpropagation method [1, 8], while in the Sensorimotor Network adaptation arises from structures $(\mathbf{S}, \mathbf{M}, \mathbf{P})$ organization using a projected gradient descent method [1]. The later performs the sequential optimization of **P**, **M**, **S**, where the input training triplets are considered in batches (action by action) as in [32]. It is important to state that in real organisms adaptation occurs every time an experience is lived. In this case, the first acquisition of the full data sets and later use for learning could represent the brain activity during a night sleep and its adaptation to experiences lived during the day.

The presented models have a big difference in the way they process the acquired training data. While the NNet gathers sensor and motor information by training its parameters using $\mathbf{i}_0$ and $\mathbf{q}$ concatenated as inputs and processed by $\mathbf{W}_1$, the SNet processes both inputs separately with a specified structure each, $\mathbf{S}$ to process image and $\mathbf{M}$ to process encoded actions.

On the one hand, the Sensorimotor Network model is formed by a sensor structure composed by 9 visual receptive fields and a motor structure composed by 9 motor movement fields ($n_s = 9, n_m = 9$). On the other hand, the Multilayer Perceptron is equipped with a hidden layer of 9 neurons. The number of hidden units used (receptor fields, movement fields and neurons) can be chosen taking into account

the resources available in the particular hardware to deploy the system and considering both image and action space size. In order to compare both models in ExpXY and ExpRZ an identical number of $n_s$ and $n_m$ was used, however, if larger images were used as sensory input, one could consider an higher number of visual receptive fields, while a broader action space could consider an higher number of motor movement fields.

## 3.3 Statistical Comparison

After training both networks, they are compared in terms of prediction error (Chapter 4), computational cost (number of parameters used), together with a relative comparison regarding loss of information (information criteria).

### 3.3.1 Computational Cost

Regarding system's computational costs, the number of trained parameters (equations 3.1 and 3.2) and the number of significant parameters (different from zero) are computed.

$$\textit{NNet}_{Parameters} = (N_m + N_s + 1) \times n_s + (n_s + 1) \times N_s \tag{3.1}$$

$$\textit{SNet}_{Parameters} = N_s \times n_s + N_m \times n_m + n_m \times (n_s \times n_s) \tag{3.2}$$

Multilayer Perceptron trains $\mathbf{W}_1$ and $\mathbf{W}_2$ being all different from zero, while Sensorimotor Network trains $\mathbf{S}$, $\mathbf{M}$ and $\mathbf{P}$ sparse matrices. In spite of, training more parameters than NNet if the number of visual receptive fields, $n_s$, and number of movement fields, $n_m$, increase, SNet has an huge advantage by developing sparse structures.

### 3.3.2 Information Criteria

With the goal of performing a relative evaluation of these models performances, two information criterion are computed: Akaike information criterion (AIC) and Bayesian information criterion (BIC) [28]. Using the following approximated functions for high number of data samples,

$$\textit{AIC} = 2k - 2\log(L) \tag{3.3}$$

$$\textit{BIC} = k\log(n) - 2\log(L) \tag{3.4}$$

where $\log$ is the natural logarithm, $k$ is the number of parameters to be estimated and $n$ is the number of data samples (triplets) used for training. In this context, $L$ stands for the considered likelihood function,

$$L = \exp^{-\lambda \text{RMSE}^2} \tag{3.5}$$

with $\lambda = 0.9$ and $\mathbf{RMSE}$ being the root mean square error between the expected post-action images and its predictions.

# Chapter 4

# Results

## Contents


**4.1 Sensorimotor Network vs Multilayer Perceptron – Comparison . . . . . 24**

4.1.1 Prediction Error . . . . . 24

4.1.2 Visual Stimuli Reconstruction . . . . . 25

4.1.3 Comparison Summary . . . . . 27

**4.2 Sensorimotor Network - Connectivity and Properties . . . . . 27**

4.2.1 Sensor and Motor Structures . . . . . 28

4.2.2 Predictive Structure . . . . . 29

4.2.3 Sensor Visual Receptive Fields Influence . . . . . 31

4.2.4 Environment Influence . . . . . 32

**4.3 Sensorimotor Network - Drone . . . . . 33**


In this section the obtained results are shown. Initially are presented the results regarding the reconstruction error, loss of information and computational costs from the optimization of the two models under comparison (Sensorimotor Network *vs* Multilayer Perceptron), using the architectures and experimental setup described in the previous chapter.

After comparison, the structures composing the sensorimotor architecture are studied with respect to their evolution during the learning process, as well as the effect each resulting topology has on the others. Each structure property and interrelationship is described using different examples from previous experiments ExpXY and ExpRZ. In addition, and with the objective of understanding the effect visual perception takes on the sensorimotor system, new tests are performed using different sizes of visual sensors and others using different environments for agent exploration (the action spaces and motor topologies remain the same in both cases). In the end, a Sensorimotor Network is trained using real data acquired using a quadricopter in an open environment.

## 4.1 Sensorimotor Network vs Multilayer Perceptron – Comparison

After training converge in both networks several statistics were computed in order to evaluate and compare their performance. For both ExpXY and ExpRZ experiments, the Multilayer Perceptron was trained using two different activation functions: one linear (*purelin*) and another non-linear (*hyperbolic tangent sigmoid*).

### 4.1.1 Prediction Error

For SNet and MLP, while performing the optimization, the RMSE between the predicted, $\mathbf{I'}_1$, and the expected images, $\mathbf{I}_1$, is computed for the training set and for the validation set, separately,

$$\text{RMSE} = \sqrt{\frac{1}{N_m \times L \times N_s} \sum_{k=1}^{N_m} \sum_{l=1}^{L} \sum_{p=1}^{N_s} \left( \mathbf{i'}^{k}_{1_{(l,p)}} - \mathbf{i}^{k}_{1_{(l,p)}} \right)^2} \tag{4.1}$$

where $L$ stands for the number of samples per action, and the validation set is used as a stopping criterion: the optimization stops when the training error becomes almost constant and the validation error starts to grow. Additionally, for comparison with the expected prediction results, the RMSE was computed for each pixel of the used 15x15 square retina,

$$\text{RMSE}_p = \sqrt{\frac{1}{N_m \times L} \sum_{k=1}^{N_m} \sum_{l=1}^{L} \left( \mathbf{i'}^{k}_{1_{(l,p)}} - \mathbf{i}^{k}_{1_{(l,p)}} \right)^2} \tag{4.2}$$

using the training set and the test set for both experiments. In Figure 4.1 the linear Multilayer Perceptron and the Sensorimotor Network are compared in terms of visual stimuli prediction performance.

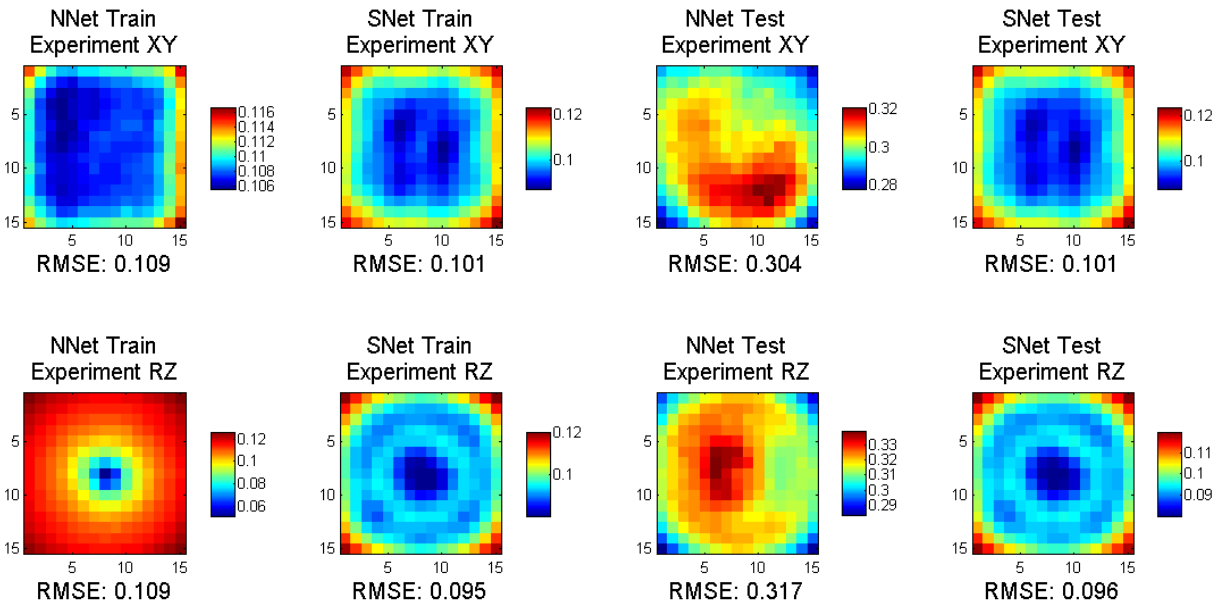



Figure 4.1: Average RMSE per Pixel considering all 10 runs from ExpXY (top) and ExpRZ (bottom) evaluated for training and test sets from ActXY and ActRZ, respectively.

Analysing the previous figure, it can be observed that Sensorimotor Network has better results in visual stimuli prediction then Multilayer Perceptron. On the one hand, using the training data, SNet has relative RMSE decrement of $8\%$ for ExpXY and about $13\%$ for ExpRZ. On the other hand, using the test set, SNet maintains its high prediction results while MLP triples its prediction errors. From these results, it starts to be clear that using a visual sensorimotor system it is possible to create some relationship between motor feedback and its effects on sensory stimuli. On contrary, MLP simply adapts to training data and is not capable of predicting visual stimuli which has never been processed before.

Moreover, as expected, the prediction error is higher in the retina's periphery because when the agent performs movements which result on image translations or zoom out scalings, part of the post-action visual stimulus will be unpredictable (it was out of the pre-action field of view).

In spite of having a similar error pattern in both Multilayer Perceptron and Sensorimotor Network for visual stimuli resulting from performing translational actions, when predicting visual stimuli for rotations and scaling Sensorimotor Network shows greater performance than linear Multilayer Perceptron. Results using a non-linear Multilayer Perceptron are shown in the next subsection.

### 4.1.2 Visual Stimuli Reconstruction

Since the optimization problem in both trained networks has a cost function minimizing the difference between the used output images for training and the target visual stimuli to be perceived by the agent, it is expected that after the cost function reaches a minimum it becomes possible to compute images $\mathbf{I}_\mathbf{1}'$ which somehow are similar to those used as output $\mathbf{I_1}$ during training by applying the same input data $(\mathbf{I_0}, \mathbf{q})$. Computing an image $\mathbf{i}_\mathbf{1}'$ corresponds to reconstructing what the agent expects to see after performing a specific action.

Considering the visual stimuli reconstruction (Figures 4.2 and 4.3) it can be stated that even with a similar RMSE, the SNet model presents more coherent predictions than linear or non-linear MLP models when using training data inputs. Here, the non-linear Multilayer Perceptron was used trying to fight the radial pattern presented by linear Multilayer Perceptron on its reconstructions, but the obtained prediction error increased. While Multilayer Perceptron computes an average of intensities from the training data, the Sensorimotor approach can use sensor and motor information in a more efficient way, creating a relation action-consequence where by activating a motor movement field it is know how it will directly affect visual stimuli (see Section 4.2).

From SNet predicted visual stimuli, it can be observed that visual receptive fields emerge. These visual receptive fields are activated with light intensities equivalent to those which were expected from training data, however almost as a subsampling equivalent from pixels to visual receptive fields. This said, it may be valid to affirm that an higher number of receptive fields could lead to a not so coarse representation of the predicted stimuli.

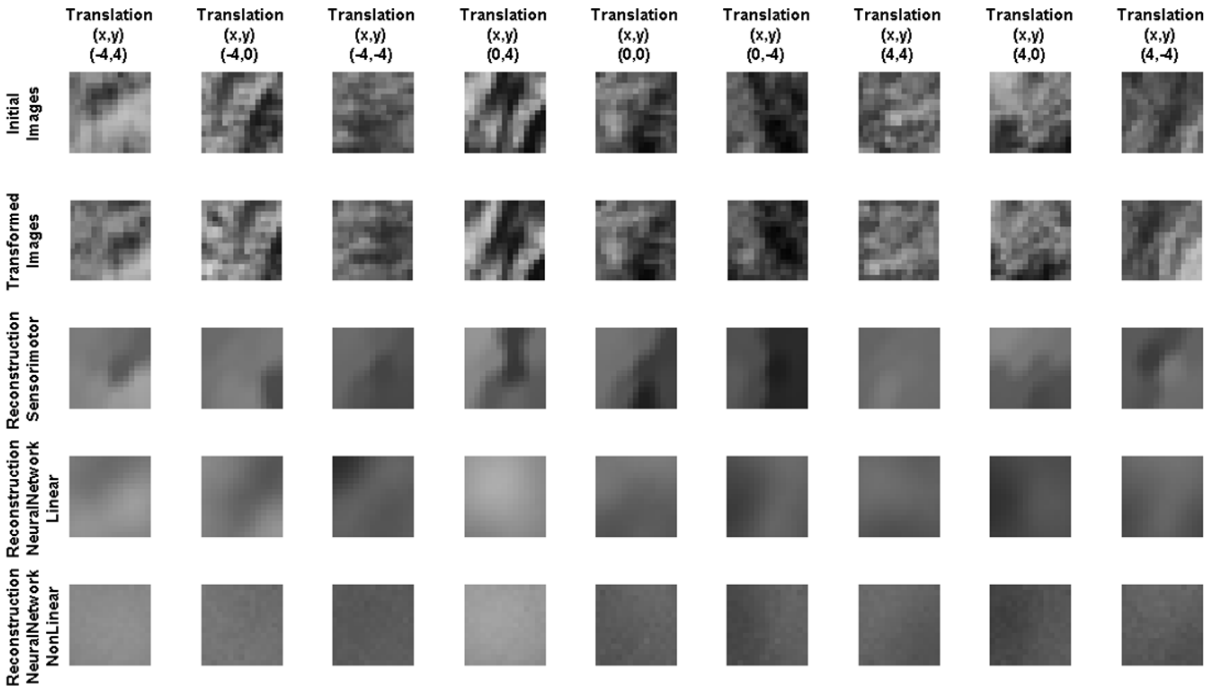



Figure 4.2: Visual stimuli reconstruction for different action examples on ExpXY, using Sensorimotor Network, linear and non-linear Multilayer Perceptron. The obtained RMSE for the presented models were: 0,1003 (SNet) <0,1091 (linear MLP) <0,1166 (non-linear MLP).

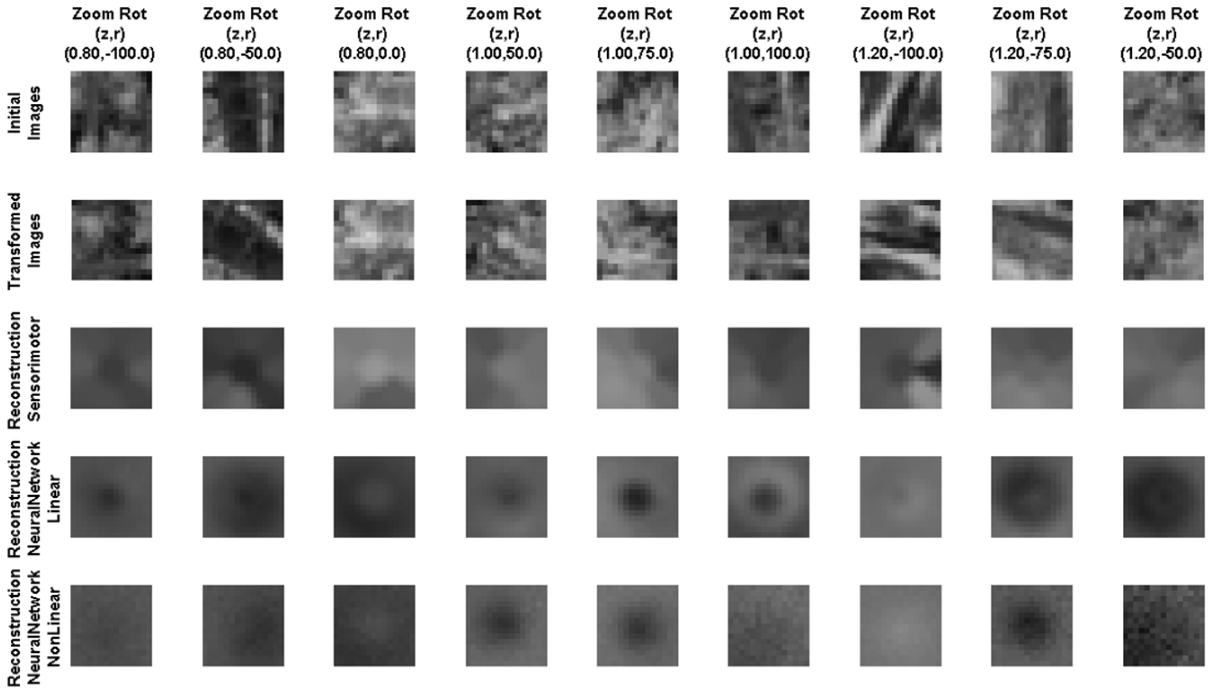



Figure 4.3: Visual stimuli reconstruction for different action examples on ExpRZ, using Sensorimotor Network, linear and non-linear Multilayer Perceptron. The obtained RMSE for the presented models were: 0,0944 (SNet) <0.1097 (linear MLP) <0,1145 (non-linear MLP).

### 4.1.3 Comparison Summary

By computing the reconstruction error on both experiments (Section 4.1), the number of parameters and the relative information loss criterion (Section 3.3), it was possible to successfully compare the Sensorimotor Network *vs* Multilayer Perceptron performance on visual stimuli prediction. Tables 4.1 summarize the obtained results.

| **ExpXY** | SNet | Linear MLP | MLP/SNet | Non-Linear MLP | MLP/SNet |
|---|---|---|---|---|---|
| All Parameters | 3483 | 5013 | 1,44 | 5013 | 1,44 |
| Parameters $\neq 0$ | 1140 | 5013 | 4,40 | 5013 | 4,76 |
| Parameters $\geq 10^{-3}$ | 803 | 4910 | 6,11 | 4992 | 6,22 |
| RMSE | 0.1004 | 0.1087 | 1,08 | 0.1241 | 1,24 |
| AIC | 2.654 | 10.457 | 3,94 | 10.026 | 3,78 |
| BIC | 10.628 | 45.546 | 4,29 | 45.115 | 4,24 |

| **ExpRZ** | SNet | Linear MLP | MLP/SNet | Non-Linear MLP | MLP/SNet |
|---|---|---|---|---|---|
| All Parameters | 3483 | 5013 | 1,44 | 5013 | 1,44 |
| Parameters $\neq 0$ | 1053 | 5013 | 4,76 | 5013 | 4,76 |
| Parameters $\geq 10^{-3}$ | 743 | 4925 | 6,63 | 4993 | 6,72 |
| RMSE | 0.0955 | 0.1100 | 1,15 | 0.1233 | 1,29 |
| AIC | 2.442 | 10.467 | 4,29 | 10.026 | 4,11 |
| BIC | 9.817 | 45.556 | 4,64 | 45.115 | 4,60 |

Table 4.1: Comparison between SNet, linear MLP and non-linear MLP in experiments ExpXY and ExpRZ. The presented values result from the average of all 10 runs in each trained model and experiments.

As observed sensorimotor approach uses over 4 times less significant parameters than the common multi-layer perceptron, produces 5 to 10% less reconstruction error and has less loss of information, with better relative results in experiment ExpRZ.

## 4.2 Sensorimotor Network - Connectivity and Properties

As mentioned before, the Sensorimotor Network model is based on visuomotor feedback interaction. As an organism can't decide how or whether to act without any sensory stimulus, it also makes no sense for an organism to develop a sensory structure whose information gives no meaningful perception of its actions consequences and its environment.

During a sensorimotor system training, a sensor structure $\mathrm{S}$ and a motor structure $\mathrm{M}$ organize themselves from random initial topologies by adapting to visual stimuli and motor activations. The visuomotor interactions are given by a third predictive structure which associates a specific motor activity to changes in the visual stimuli. In the next subsections some examples regarding sensor and motor structures evolution are presented and in addition the predictor role is explained.

### 4.2.1 Sensor and Motor Structures

Every performed experiments and runs using the SNet result on organized sensorimotor structures which correspond to a local minimum to the optimization problem of minimizing the prediction error of visual stimuli given an image and an action to be performed.

As in Nature, visual systems presented in individuals of the same species (apart from diseases or accidents) develop the same type of structures, however this structures are interconnected using different neural pathways due to neuroplasticity response to individual behaviour and experiences. Since every data set used random samples from the environment and consequently slightly different sensory inputs, it is expected that each experiment developed similar structures. In Figures 4.4 and 4.5 an example of visual sensory structure $\mathbf{S}$ and motor structure $\mathbf{M}$ adaptation per experiment is shown.

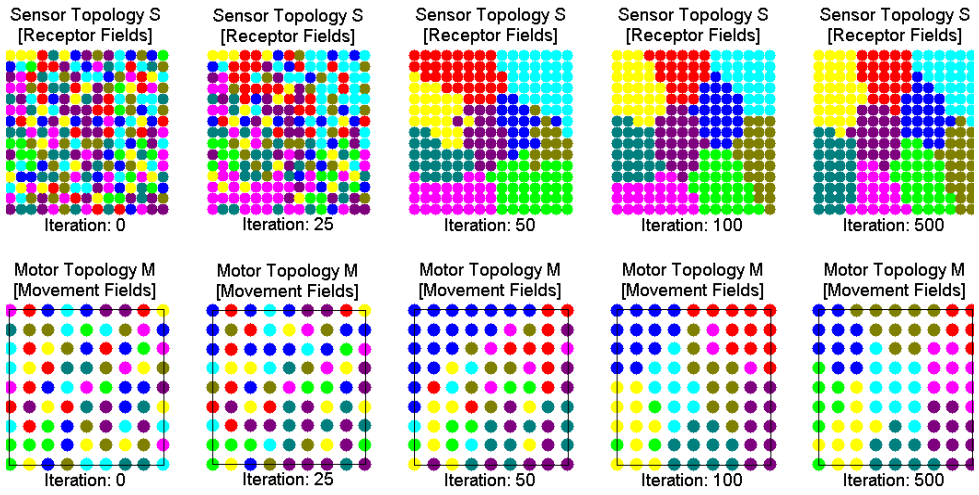



Figure 4.4: Sensor receptive fields (top) and motor movement fields (bottom) evolution from an experiment ExpXY example run training Sensorimotor Network. Horizontal and vertical axis stand for horizontal and vertical translations, respectively, in motor space.

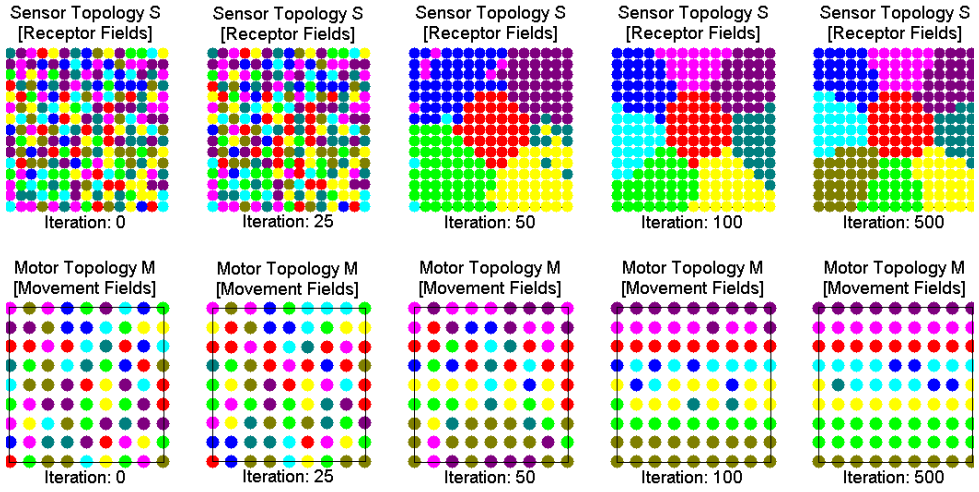



Figure 4.5: Sensor receptive fields (top) and motor movement fields (bottom) evolution from an experiment ExpRZ example run training Sensorimotor Network. Horizontal and vertical axis stand for scalings and rotations, respectively, in motor space.

Throughout sensorimotor model learning process, from a random initial topology, these structures develop very distinctive organizations which yield properties from the training data, creating some sort of relational memory. An agent can recognize the effect of a performed action in its perceived visual stimuli, and from two consecutive visual stimuli it can coarsely identify which action it took.

From the presented results, a direct adaptation by the motor structure to the received stimuli can be observed. Remembering the used ActRZ data set, where rotations and scalings were combined, it can be concluded that the chosen rotations produce more changes in visual stimuli than zooms. Each movement field maps a rotation angle, with exception on center movement fields where rotation is null, and scaling gains significance.

Although sensor and motor structures present different topologies when the agent's action space varies from translations to rotations and scaling, also within the same experiment some variations can be observed and the sensorimotor model can converge, for instance, to configurations which only make use of 8 out of the 9 available visual receptive fields. These solutions present equally good results and an example of this behaviour can be seen in Subsection 4.2.3.

### 4.2.2 Predictive Structure

As in a nervous system a sensory area and motor area of the brain are intimately related, but the solo existence of both sensor and motor structures is not enough. As in a corollary discharge circuit leads motor information from the superior colliculus deep motor layers to the visual receptors in the frontal eye field, the sensor and motor structures, **S** and **M**, from SNet must be connected using a predictive structure, **P**. This will create a dependency where a specific group of motor actions will cause a similar effect in the visual perception (movement receptive fields to visual receptive fields), but also an effect observed between two consecutive visual stimuli can lead to a coarse identification of which kind of action caused it (visual receptive fields to movement receptive fields).

In Figure 4.6 simple and clear examples of visual prediction capabilities of SNet are presented. Taking the presented translational example the prediction of visual stimuli can be explained taking the following steps:

1. An action primitive $\mathbf{q}$, encoding action $\mathbf{u} = (+4, +4)$, for instance, is received by $\mathbf{M}$ and associated to movement field 7 activation.

2. A visual stimulus $\mathbf{i_0}$ is received by $\mathbf{S}$ and is mapped into visual receptive fields intensity activations generating an observation $\mathbf{o_0}$.

3. Each movement field is associated with a predictor $\mathbf{P_k}$ which translates the motor influence on the existing visual receptive fields.

4. The predictor weights connections between the receptive fields by identifying areas of observation $\mathbf{o_0}$ which will move from a receptive field (transmitter) to another (receiver).

5. By moving information within the visual receptive fields (arrows), a new observation $\mathbf{o_1}$ is generated and the visual stimuli $\mathbf{i_1}$ is reconstructed.

Yet, in Figure 4.6, visual receptive fields and motor movement fields are divided using Voronoi diagrams between each field centroid. Moreover, the prediction links are only considered if their weights are above 0.25. In Figure 4.7, examples of movement prediction in ExpXY and ExpRZ are shown.

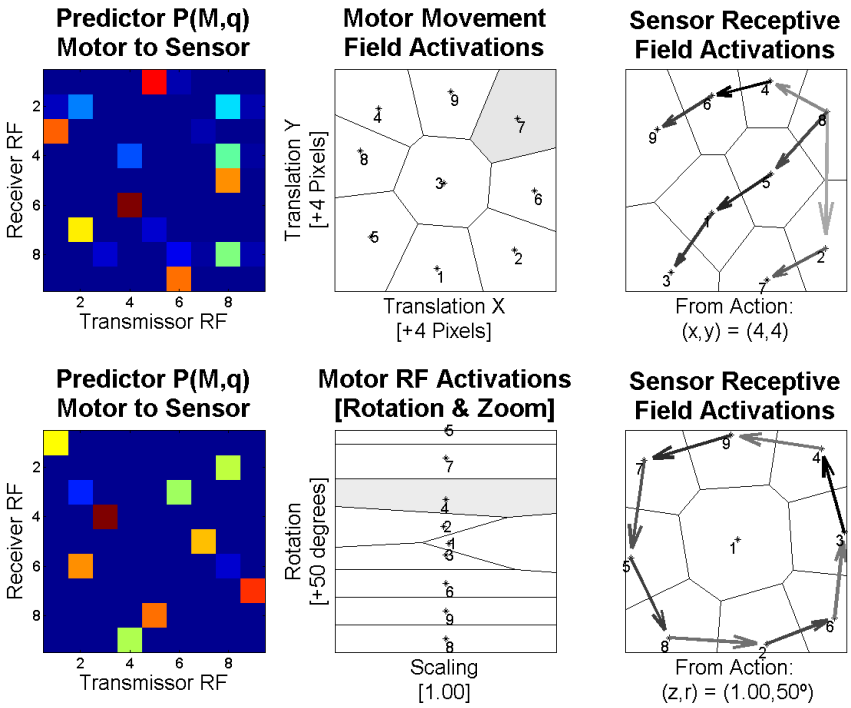



Figure 4.6: Predictor, Motor and Sensor structures relationship example from ExpXY and ExpRZ for action $u = (+4, +4)$ (top) and action $u = (1, 50°)$ (bottom), respectively. Sensor receptive field connections are represented by arrows with intensity proportional to the corresponding prediction matrix entry.

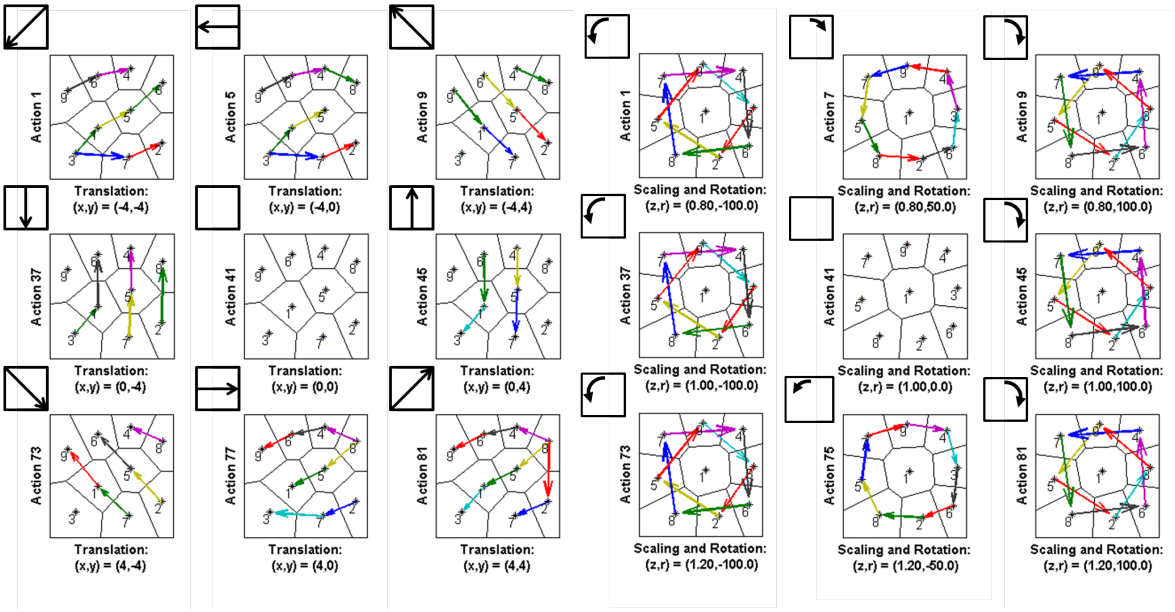



Figure 4.7: Predictive structure influences on sensory receptive fields after receiving a motor activation from action space ActXY (left) and action space ActRZ (Right). On each example, in the top left corner the arrow in a square represents the direction and/or amplitude of the sensor translation or rotation with respect to environment when an action is performed.

### 4.2.3 Sensor Visual Receptive Fields Influence

Looking for the direct influence of sensor structure complexity on the stimuli prediction, some tests were made using experiment ExpXY data set. Three different models were trained, all in the same conditions and using the same action space ActXY, but with different number of available visual receptive fields in the sensory structure (9, 16 and 25). In Figure 4.8 it can be observed the organized sensor topologies from the three different sensor complexity models. Besides, the reconstruction error, $\mathbf{RMSE}$, was computed using a test set from the same experiment ExpXY, and a reconstruction example is shown. For the same action and pre-action image, the visual stimuli prediction is computed and compared with the actual observed post-action visual stimulus.

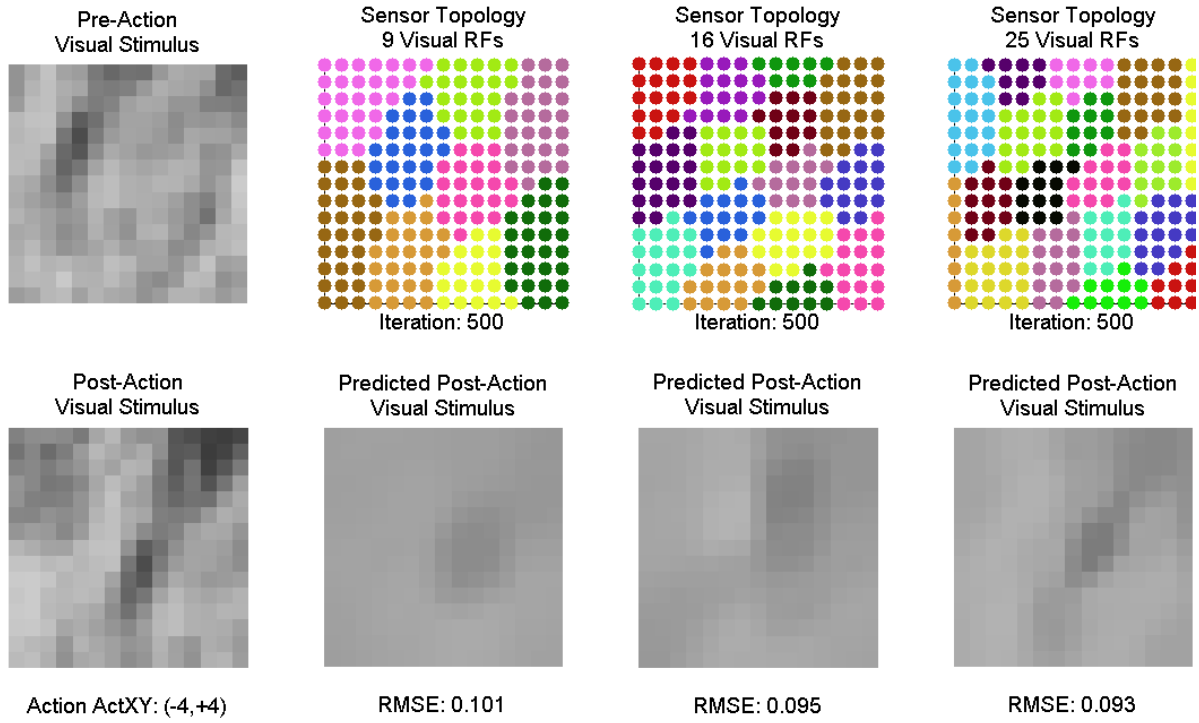



Figure 4.8: Sensor Visual Receptive Fields: Influence on prediction error and quality on image reconstruction.

As expected and observable, the quality of the reconstruction improves with the number of visual receptive fields. Although the prediction error decreases with the number of available receptive fields, this also presents an increasing relative number of empty receptive fields. In the case where 9 sensor receptive fields were considered, the model used all of them. However, when reaching higher number of visual receptive fields, some become unnecessary for the adapted model (1 out of 16 RFs and 7 out of 25 RFs). All in all, a trade-off can be found in increasing the sensor complexity: on the one hand the prediction error decreases, but the amount of receptive fields consuming computational power, without any advantage for the prediction, increases.

### 4.2.4 Environment Influence

As proved in many works [12, 16, 35] the eyes, retinas and/or visual systems evolved in many species in very distinctive manners, but all highly efficient when vision appears as the most important sense for the organism. Three main characteristics can be enumerated which directly influence their structures: organism's nervous system, organism's motor capabilities and organism's perception of the environment.

Here the environment influence on the sensory structure within the sensorimotor system is tested. Using the ActXY and ActRZ data sets and the same sensorimotor configuration as in ExpXY and ExpRZ, four different environments were used training: 3 artifical (vertical stripes, diagonal stripes and dots) and 1 natural (textured picture of dry soil). In Figure 4.9 there are represented the resulting sensor organizations, $\mathbf{S}$, for the 4 environments and the number of iterations until convergence.

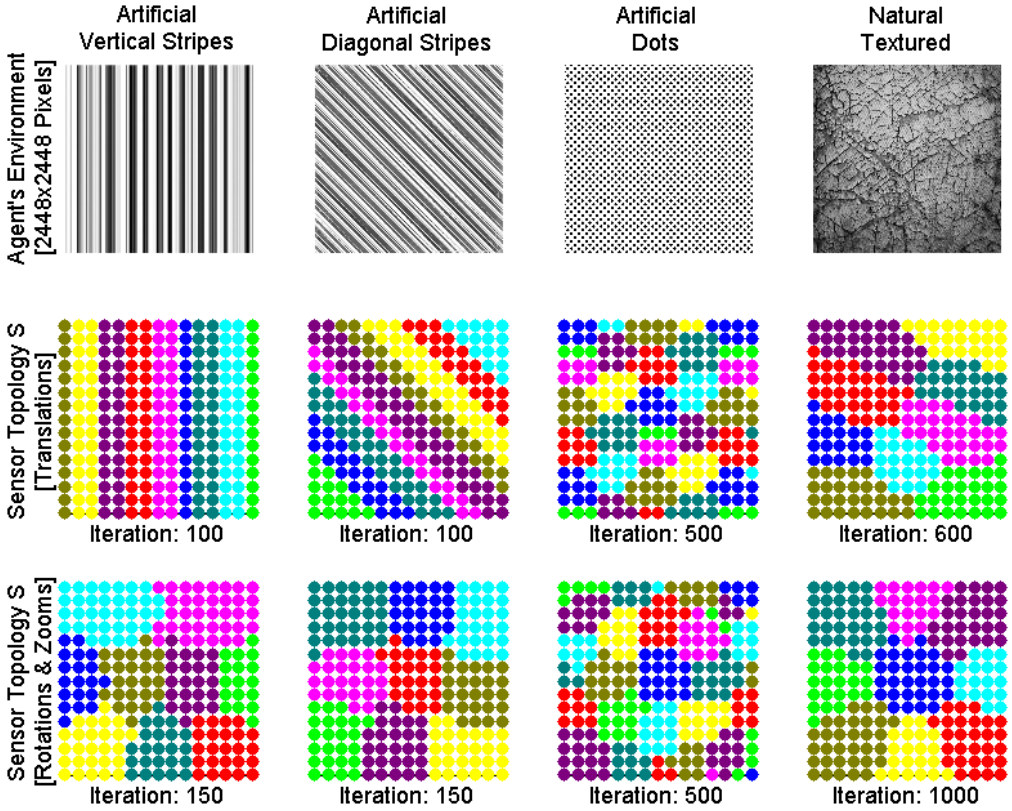



Figure 4.9: Environment influence on visual sensor topology. Sequence of visual sensor topologies resulting from training the sensorimotor system using action spaces, ActXY and ActRZ, and four different environments: three artificial environments (vertical stripes, diagonal stripes and dots) and one natural environment (textured picture of dry soil).

From the presented results it can be concluded that the sensor structure organization depends on the environment. Considering the question made in [3] and the tested sensorimotor system, it can be hypothesized that a retina does acquire knowledge, in its organization, about the natural scenes (environment). However, it is shown that the way the agent perceives its environment is the key factor for the resulting visual sensor topology. Even with very different topologies between environments, it can be observed that only by changing the set of movements the agent can perform, the way the same environment is perceived also changes.

Taking as an example, the artificial environment composed by vertical stripes, if an agent performs only translational movements parallel to the environment, the unique type of stimuli the agent will know corresponds to vertical stripes, then the most efficient retina it could develop should be one which translates the possible changes in the perceived stimuli (horizontal movement of the vertical stripes).

From another point of view, if the agent is only able to perform rotational and scaling movements, then the visual stimuli can change from vertical stripes to diagonal or even horizontal stripes. With such a variation of stimuli, it is expected that the retina topology should be different.

In Figure 4.10, the prediction error between the trained sensory topologies can be compared. As expected, and taking into account the small number of visual receptive fields, more complex artificial images lead to worse predictions. Furthermore, the Sensorimotor Model shows better performances in grayscale images than in black and white ones (less natural).

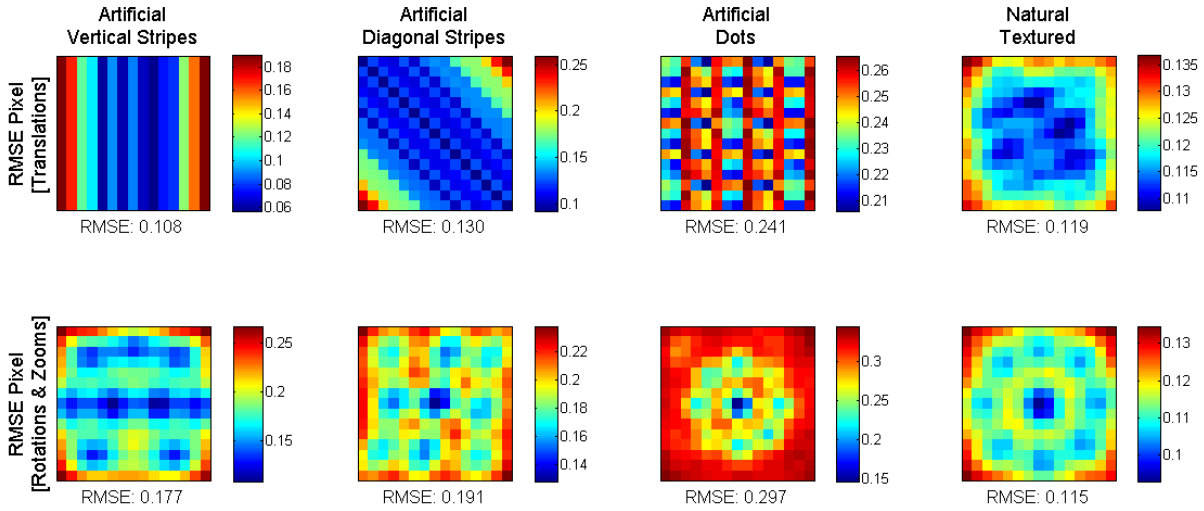



Figure 4.10: RMSE of different visual sensor topologies.

## 4.3 Sensorimotor Network - Drone

After studying the structure organization of the sensorimotor system and its adaptability to artificial and natural static environments, a retina was trained using real data acquired by a quadricopter Drone in an open environment.

Figure 4.11: Sensorimotor Real Data: Parrot AR.Drone2.0 and its flight exploratory path in Monsanto park, Lisbon.

A Parrot AR.Drone2.0 was used to acquire images from a natural environment in Monsanto park in Lisbon. This drone is equipped with a fixed front HD camera which during the experiment was always pointing to its movement direction. During the flight a video was recorded at a rate of 30 frames per second, together with drone position variations ($\Delta x$,$\Delta y$) from GPS, orientation variations ($\Delta\theta$) and absolute altitude (which corresponds to the state of the drone). For the sake of simplicity and since state is not explicitly modeled in this work, data from drone taking off or landing was removed and the altitude was admitted as constant.

The data acquisition (image and actions) was performed while the drone followed a pre-planned trajectory, on constant altitude, where it had to pass over some locations defined by GPS coordinates using its inner flight planner set through QRGround Flight Control. Examples of acquired images, and respective bilinear subsampling to 15x15 pixels images for training, can be seen in Figure 4.12.

Figure 4.12: Examples of images acquired using the Drone in Monsanto, Lisboa, and respective subsampling images used for training.

The full data set recorded has 8340 samples, but with a rate of 30 recorded samples per second, the variation between a pre-action and a post-action image was practically unnoticeable. This considered, the training samples were cut to 556 with a time difference between two consecutive images of 0.5 seconds (2 samples per second). The retina was trained using 556 data triplets, $(\mathbf{i}_0, \mathbf{i}_1, \mathbf{q})$, with 95 different action identifiers.

Differently from the direct application of the Sensorimotor Network used in [30] where the action space discretizes a two dimensional motor space, in this experiment a motor space with 4 degrees of

freedom is considered. Each degree of freedom was separately quantized in 4 bins, using k-means clustering algorithm [13]). These were concatenated and then, to each unique combination of the concatenated vectors a specific action identifier $\mathbf{q}$ is assigned.

In Figure 4.13 three examples of visual stimuli prediction are shown, using two different complexities of sensor structure: one with 9 visual receptive fields and another with 16.

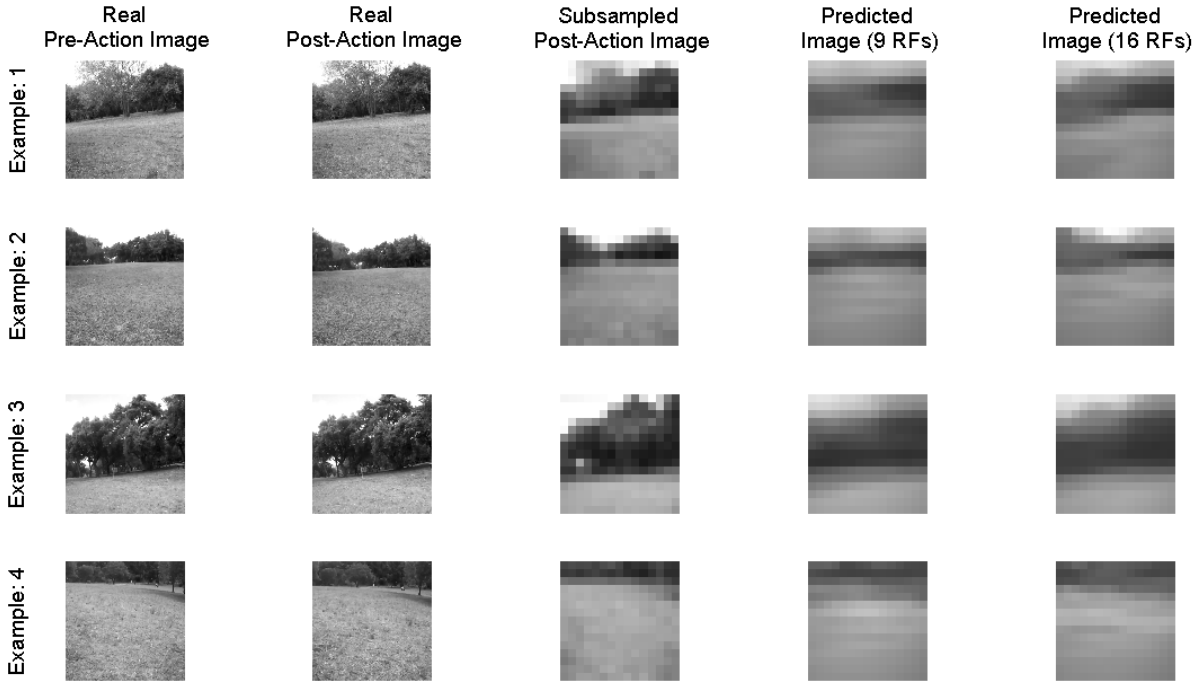



Figure 4.13: Visual stimuli prediction using two different Sensor complexities (9 and 16 visual receptive fields).

As observable, and expected from previous results, the reconstruction is slightly better using the more complex retina. Above, in Figure 4.14 it is show both sensor organization topologies and respective RMSE. The area with lower error corresponds to ground which occupies the bottom half of the field of view with some deviations. During its flight, the ground suffers some vertical movements (bigger and more horizontal receptive fields). Looking at the top half of the drone's field of view, it can be seen that a greater variability exists, originating a denser distribution of visual receptive fields.

The tested sensorimotor network is used with structures complex enough to successfully demonstrate its applicability and great predictive skills. However, if this model is to be used in a certain task, it can be required that training images become larger and the number of visual receptive fields and/or motor movement fields increase considerably. This would need a bigger time for training the model.

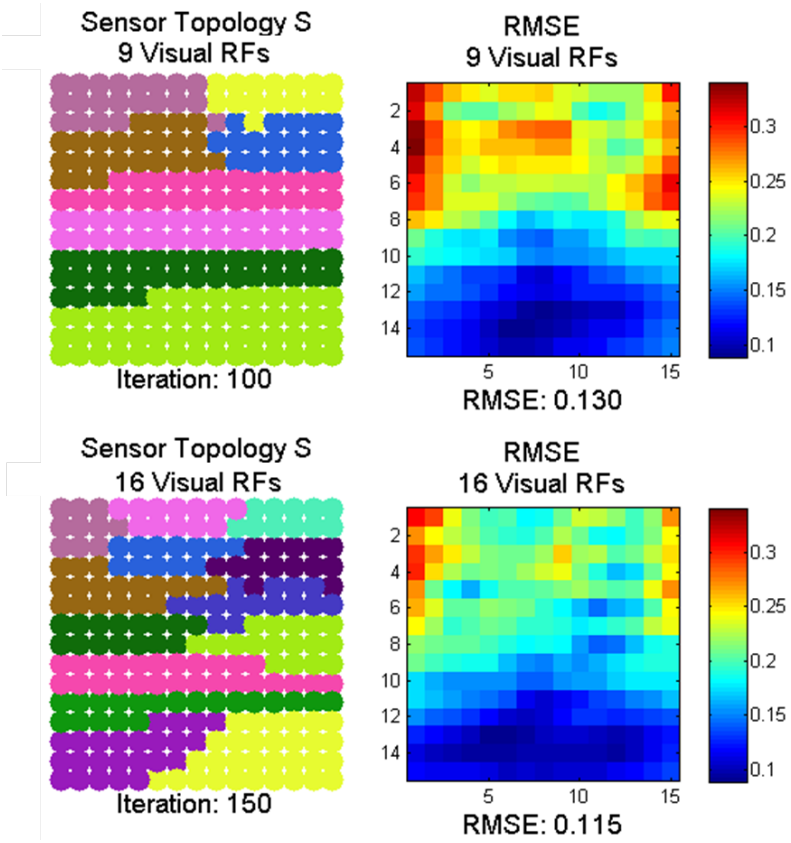



Figure 4.14: Sensors organization topologies after training and respective prediction error, RMSE.

# Chapter 5

# Conclusions

## Contents


5.1 Achievements . . . . . 37

5.2 Future Work . . . . . 38


Nature shows us countless organisms all with the ability to adapt and evolve in a dynamic environment. In robotics, as in many other engineering fields, there are numerous problems where Nature is often the best role model to solve them.

As organisms thrive in their habitats, robots are now to be developed so they can be deployed in a unpredictable environment and adapt to successfully perform their tasks, instead of following highly strict routines in an artificially controlled environment. With this work, it was possible to verify the adaptation of an agent with two different structures (visual and motor) to its capabilities, resources and surroundings.

Furthermore, the Sensorimotor Network approach presents more meaningful structures after optimization (biologically comparing) than the weights which result from Multilayer Perceptron optimization.

## 5.1 Achievements

In this work, the proposed method [30] was successfully applied for post-action images reconstruction and significant reduction of the number of parameters needed to predict visual stimuli caused by self-induced actions by drawing inspiration from biological systems.

The development of visual receptive fields taking into account the changes induced by motor actions allows a good adaptability of the organism to the environment and thus a cheaper way for an agent to process and predict visual stimuli. A specialized network architecture like the described SNet is advantageous for predicting the interactions between a sensory and a motor system, as well as obtaining more reliable predictions of what an agent is expecting to see after moving.

This tight relationship between perception and actions is key for guiding the development of sensory and motor systems which will support successful acting upon the environment. As demonstrated, the developed structures evolve in a manner that an action-consequence kind of organizational memory emerges from the experiences of the agents and its adaption.

The comparison performed in this work between standard Multilayer Perceptrons and Sensorimotor Networks, suggests that the latter might prove useful in bringing us a step closer to biological performance.

It is also shown that the obtained structures using the SNet are very dependant of the agent's visual resources (size of the retina and number of receptive fields) and the way it perceives its environment. On one hand, if the environment and size of the retina is such that the agent perceives its images mainly as textures, then the receptive fields will organize themselves uniformly. On the other hand, if the perceived images represent the same directional pattern, then the visual structure will evolve with receptive fields distributed in areas representing that same pattern.

Throughout the development of this thesis and from the presented results, it can be concluded that at some extent biological characteristics are starting to be computationally implemented and the statement from R. L. Gregory in [11] is gradually becoming addressed.

> *'In the human being we see preserved almost all the stages in the developments of vision from the simples reflex (closing of the eyes on sudden change of illumination), to pattern recognition, and identification of objects from unusual points of view, with prediction of the immediate future based in the past. Such feats cannot be simulated with even the most advanced computers.'*
>
> **- R. L. Gregory, 1967**

## 5.2 Future Work

Considering that it was possible to deploy a sensorimotor structure using modified neural networks, it could be important to follow the path of developing such system using a more state-of-the-art machine learning method such as Deep Learning which allows sequential training of many layers. Adding complexity and increasing the number of interconnected sensor/motor layers as presented in [31] would be the direction to take since as shown in biology the human brain, for instance, processes image at many levels of abstraction, allowing richer representations for complex tasks.

Another element which would increase the applicability of the presented work is the notion of state to support planning tasks. Applying two consecutive sensorimotor networks where the first would consider the state variables of the robot and its motor limitations to predict achievable actions, and the second would have these actions mapped as motor input and visual data as sensory input.

Developing such a system could generate two possible adaptation methods for a robot designed to act on a specific environment:

- If the type of sensory input and motor configuration of the robot and its future environment are already known, then the system could be trained offline, and later applied in the robot to use its predictive skills to perform some tasks. In this case, computational costs for using a sensorimotor model would be lowered and even a small robot could access to the pre-trained visuomotor potentials.

- If nothing is known about the robot, applying a non-trained sensorimotor system would allow it to adapt to its unknown dynamic environment and motor capabilities by exploring and gathering information from different motor actions and sensory consequences. Still, it would lack purpose or motivation to act or even decide what actions to take, concepts which jump of this thesis scope.

Finally, such a model doesn't force the sensory input data to be visual, so trying to train these models with auditory input would also be viable. In the end, depending on the type of used sensory data, tasks like anomaly detection or tracking (for vision) and speaker recognition or locations (for audio) would be more easily implemented. Using vision, detection task could be performed by evaluating the difference between the expected future visual stimuli after an action and the observed stimuli (a big difference would correspond to an abnormality detection). Moreover, tracking could be achieved by forcing the robot to maintain a specific visual stimulus in the center region of its retina (center visual receptive fields), meaning that if the target moves in a way it is perceived in the left visual receptive fields, the robot could activate the motor command which compensates that movement.

# Bibliography


[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[2] Ehud Ahissar and David Kleinfeld. Closed-loop neuronal computations: focus on vibrissa somatosensation in rat. *Cerebral Cortex*, 13(1):53–62, 2003.

[3] Joseph J Atick and A Norman Redlich. What does the retina know about natural scenes? *Neural computation*, 4(2):196–210, 1992.

[4] Bing Cheng and D Michael Titterington. Neural networks: A review from a statistical perspective. *Statistical science*, pages 2–30, 1994.

[5] D. Cox and T. Dean. Neural Networks and Neuroscience-Inspired Computer Vision. *Current Biology*, page (accepted), 2014.

[6] Trinity B Crapse and Marc A Sommer. Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, 9(8):587–600, 2008.

[7] Christine A Curcio, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson. Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4):497–523, 1990.

[8] Emile Fiesler and Russell Beale. *Handbook of neural computation*. Oxford University Press, 1996.

[9] XiaoLan Fu, LianHong Cai, Ye Liu, Jia Jia, WenFeng Chen, Zhang Yi, GuoZhen Zhao, YongJin Liu, and ChangXu Wu. A computational cognition model of perception, memory, and judgment. *Science China Information Sciences*, 57(3):1–15, 2014.

[10] Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2):119–130, 1988.

[11] RL Gregory. Origin of eyes and brains. *Nature*, 213(5074):369–372, 1967.

[12] RL Gregory, HELEN E Ross, and N Moray. The curious eye of copilia. *Nature*, 201(4925):1166–1168, 1964.

[13] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.

[14] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593–605. IEEE, 1989.

[15] Andrew J King, Mary E Hutchings, David R Moore, and Colin Blakemore. Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. 1988.

[16] MF Land. Movements of the retinae of jumping spiders (salticidae: Dendryphantinae) in response to visual stimuli. *Journal of experimental biology*, 51(2):471–493, 1969.

[17] Michael F Land and Russell D Fernald. The evolution of eyes. *Annual review of neuroscience*, 15(1):1–29, 1992.

[18] Simon B Laughlin, Rob R de Ruyter van Steveninck, and John C Anderson. The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41, 1998.

[19] Quoc V Le. Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8595–8598. IEEE, 2013.

[20] Boaz Lerner. Toward a completely automatic neural-network-based human chromosome analysis. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 28(4):544–552, 1998.

[21] Irwin B Levitan and Leonard K Kaczmarek. *The neuron: cell and molecular biology*. Oxford University Press, 2002.

[22] H Lütkepohl. w handbook of matrices. 1996.

[23] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[24] RC Miall and Daniel M Wolpert. Forward models for physiological motor control. *Neural networks*, 9(8):1265–1279, 1996.

[25] Cynthia F Moss and Shiva R Sinha. Neurobiology of echolocation in bats. *Current opinion in neurobiology*, 13(6):751–758, 2003.

[26] P Russel Norvig and S Artificial Intelligence. A modern approach, 2002.

[27] Raphael Pinaud, Liisa A Tremere, and Peter De Weerd. *Plasticity in the visual system: from genes to circuits*. Springer, 2006.

[28] David Posada and Thomas R Buckley. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5):793–808, 2004.

[29] JFA Poulet and B Hedwig. A corollary discharge mechanism modulates central auditory processing in singing crickets. *Journal of neurophysiology*, 89(3):1528–1540, 2003.

[30] Jonas Ruesch. *A Computational Approach on the Co-Development of Visual Sensorimotor Structures*. PhD thesis, Instituto Superior Técnico, 2014.

[31] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. Predicting visual stimuli from self-induced actions: an adaptive model of a corollary discharge circuit. *Autonomous Mental Development, IEEE Transactions on*, 4(4):290–304, 2012.

[32] Jonas Ruesch, Ricardo Ferreira, and Alexandre Bernardino. An approach toward self-organization of artificial visual sensorimotor structures. In *Biologically Inspired Cognitive Architectures 2012*, pages 273–282. Springer, 2013.

[33] Terrence J Sejnowski and Charles R Rosenberg. Parallel networks that learn to pronounce english text. *Complex systems*, 1(1):145–168, 1987.

[34] Marc A Sommer and Robert H Wurtz. Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. *Journal of Neurophysiology*, 83(4):1979–2001, 2000.

[35] Jonathan Stone and Paul Halasz. Topography of the retina in the elephant loxodonta africana. *Brain, behavior and evolution*, 34(2):84–95, 1989.

[36] Robert R Trippi and Efraim Turban. *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance*. McGraw-Hill, Inc., 1992.

[37] Marc M Umeno and Michael E Goldberg. Spatial processing in the monkey frontal eye field. i. predictive visual responses. *Journal of Neurophysiology*, 78(3):1373–1383, 1997.

[38] Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, and Shi-Fu Chen. Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine*, 24(1):25–36, 2002.