BIOLOGICALLY-PLAUSIBLE OBJECT RECOGNITION USING SPIKING NEURONS

December 2, 2014

By
Qian Liu
School of Computer Science

Contents

List of Tables

List of Figures

Abstract

The brain recognises huge amount of objects rapidly and effortlessly even in cluttered and natural scenes. While the major stumbling crux of the computer object recognition systems lies in the invariance problem.

To explore the invariant object recognition of the brain in a biologically plausible way is the right place to tackle the computational difficulty; in turn it also contribute to the understanding of biological visual processing by means of mimicking the neural activities in the visual systems.

As a first step in the exploration, a recognition system for dynamic hand postures is developed on the bespoken platform. It enables the study of the methods used in the ventral visual pathway of the primate brain. Inspired by the behaviours of the primary visual cortex, Convolutional Neural Networks (CNNs) are modelled using both linear perceptrons and spiking Leaky Integrate-and-Fire (LIF) neurons.

The future work is proposed to built an object recognition system with position, scale and view invariance by mimicking the neurons along the ventral stream.

Chapter 1

Introduction

Patterns or objects in two-dimensional images can be described with four properties [7]: position, geometry (i.e. size, area and shape), colour/texture, and trajectory. Appearance-based methods are the most direct approach to performing pattern recognition where the test image is compared with a set of templates to find the best match for an individual or combination of properties. However, the 2D projection of an object changes under different conditions including illumination, viewing angles, relative positions and distance, making it virtually impossible to represent all appearances of an object. To improve reliability, robustness and classification efficiency, approaches such as edge matching [8], divide-and-conquer [9], gradient matching [10] and feature based methods [11, 12] are used. Finding a proper feature for a specific object still remains an open question and there is no process as general, accurate, or energy-efficient as that provided by the brain. It is not a new idea to turn to nature for inspiration. Riesenhuber et al. [13], for instance, presented a biologically-inspired model based on the organisation of the visual cortex which has the ability to represent relative position- and scale-invariant features. Integrating a rich set of visual features became possible using a feed-forward hierarchical pathway.

1.1 What Is Object Recognition?

The definition of object recognition is well accepted [4] as the ability to assign labels to particular objects, ranging from precise labels (‘identification’) to course labels (‘categorisation’). It involves the ability to accomplish the tasks under the various identity preserving transformations such as object position, scale, viewing angel, background clutter and etc.

The brain can accurately recognise and categorise objects remarkably quickly, e.g. the recognition time in monkeys takes less than 200 ms [14] and the images are presented sequentially in spikes less than 100 ms [15]. This research focuses on this rapid and highly accurate object recognition, ‘core recognition’, which is defined in [16].

1.2 Why Is It Important?

The brain recognises huge amount of objects rapidly and effortlessly even in cluttered and natural scenes. While the major stumbling crux of the computer object recognition systems lies in the invariance problem. Each encounter of an object on the retina is completely unique, because of the illumination (lighting conditions), position (projection locations on the retina), scale (distances and sizes), pose (viewing angles), and clutter (visual contexts) variabilities. In addition, a difficult specificity-invariance trade-off occurs in the categorisation tasks, since the recognition should be able to discriminate different object classes (intraclass variability) while at the same time tolerant to image transformations.

To explore the invariant object recognition of the brain in a biologically plausible way is the right place to tackle the computational difficulty; in turn it also contribute to the understanding of biological visual processing by means of mimicking the neural activities in the visual systems. Moreover, the energy-efficient manner will help in building object recognition systems, i.e. posture recognition, as human-machine interfaces in portable devices.

1.3 How to Mimic The Brain?

To explore how brain may recognise objects, we have employed a biologically-inspired DVS silicon retina [17], a good example of low-cost visual processing due to its event-driven and redundancy-reducing style of computation; and a SpiNNaker system [18], which is a massive parallel computing platform aimed at real-time simulation of SNNs. With this neuromorphic hardware system we have the ability to explore visual processing by mimicking the functions of different layers along the visual pathway.

Building a real-time recognition system for dynamic hand postures is a first step of exploring visual processing in a biological fashion and is also a validation of the neuromorphic platform. To keep the task simple at first, the postures are of similar size and the goal is to recognise the shape of a hand with moving positions.

1.4 Report Structure

In Chapter 2, this report starts from the biology aspects of object recognition: the task is mainly processed along the ventral visual pathway with the untangling object representations at different level in the hierarchical abstractions. This is followed by the introduction of spiking neural networks to illustrate the abilities of neural modelling from single neurons to populations. Besides, the learning algorithms are also included together with successful recognition/classification tasks where they are applied. The final part in this chapter describes the details of the hardware neuromorphic system, including the silicon retina and the SpiNNaker platform.

In terms of the preliminary work, the convolutional neural network models exploiting V1-like neurons are defined and tested on Matlab, and the model structures and experimental results are stated in Chapter 3. In Chapter 4, the rate-based models are converted into spiking neurons, and real-time live recognition and recorded data experiments are carried out.

Finally, the contribution of this work is summarised and the future directions are provided in Chapter 5.

Chapter 2

Background

This chapter provides the readers with detailed biology background of object recognition in the brain and the fundamental principles of neural modelling using spiking neural networks. This is followed by an introduction to the neuromorphic hardware platform specialised for neural simulations of visual processing.

2.1 How the Brain Represent Objects?

The central visual system consists of several cortical areas responsible for visual processing, which are placed in a hierarchical pattern according to the anatomical experiments [19]. There are two basic streams locating in the visual area: a dorsal and a ventral pathway (Figure 2.1).

They differ in behavioural patterns according to the observation from brain lesions [20], and also in functions where the dorsal pathway targets on the 'where' tasks and the ventral on the 'what'. The ventral visual stream holds the critical circuits for object recognition and stimulus identification, whereas the dorsal pathway pathway contributes to the processing of the spatial location of the stimulus [20, 21]. Another definition of the difference between these two pathways is a 'perception/action' dichotomy: the ventral ('perception') stream perceives the world by means of object recognition and memory, while the dorsal ('action') stream provides real-time visual guidance for motor actions such as eye movements and grasping objects [22].

This research mainly focuses on the ventral visual pathway, since it dominates the object recognition among the cortical areas. Thus, the dorsal pathway will beyond the scope of the this study.

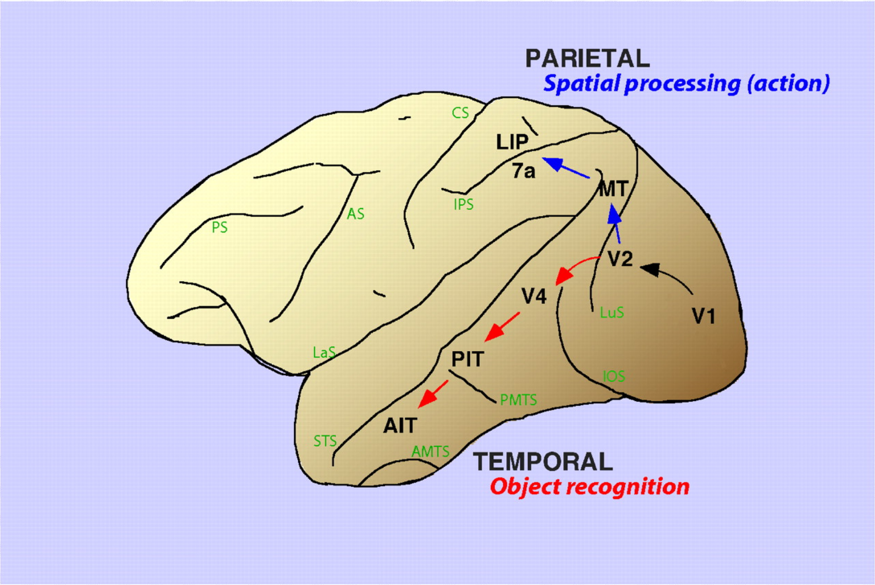


Figure 2.1: The dorsal and ventral pathway in the brain [1]. The dorsal stream (blue) arrives to the parietal lobe, whereas the ventral pathway (red) reaches the inferotemporal (IT) cortex in the temporal lobe.

2.1.1 Cortical Areas in The Ventral Visual Pathway

The ventral visual pathway starts from the primary visual cortex V1 in the occipital cortex through areas such as V2 and V4 to the Inferotemporal (IT) cortex. These cortex areas are divided based on the anatomical experiments and retinotopic maps. Accordingly, the IT complex is commonly parsed into sub areas such as TEO and TE [23, 24] or posterior IT (PIT), central IT (CIT), and anterior IT (AIT) [19].

Primary Visual Cortex:V1

As the simplest and earliest cortical area in the ventral stream, the primary visual cortex V1 is the best-studied since the well-known discovery of the orientation selectivity by Hubel and Wiesel [25] in 1958. The retinotopic map is well-defined to transform spatial information from retinal image to V1 [26]. In human and animals with a fovea in the retina, the central 10 degrees of the visual field occupies roughly half of V1. This distorted retinotopic map in V1 is a phenomenon known as cortical magnification.

In the spatial domain, V1 neurons are tuned to Gabor-like transforms applied to their small local receptive field. The retinotopic and orientation map on the surface of V1 of a tree shrew is shown in Figure 2.2. A black bar presented in the retinal image evokes a response in the corresponding grid square of V1 (6°, 2°) depending on the

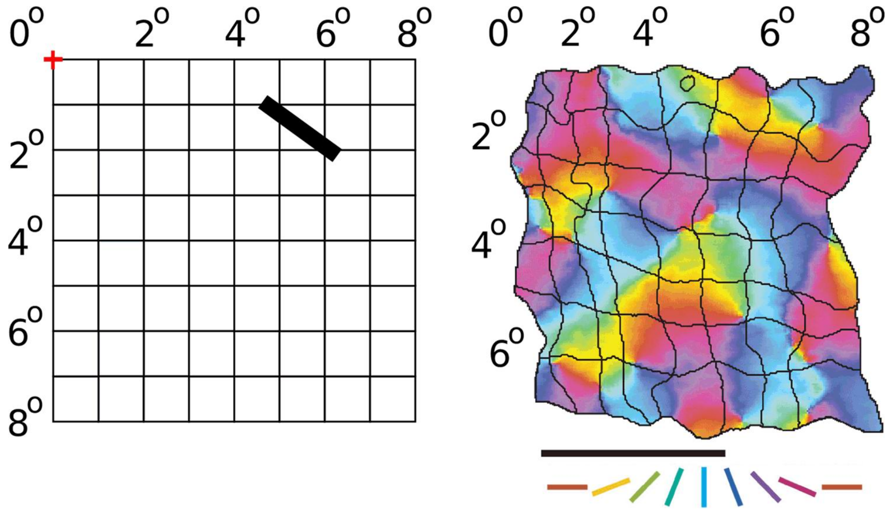


Figure 2.2: The retinotopic and orientation map on the surface of V1 of a tree shrew [2]. The visual field (left) with a fixation point marked as a red cross on the up-left corner can be divided into a regular grid. Each square represents a 1° × 1° area of visual space. In cortical area V1 of mammals, neurons are arranged into a retinotopic map (right) responding to the retinal visual space. As an example, the retinotopic map shows the orientation preference of the V1 neurons of a tree shrew for an 8° × 7° area of visual space (adapted from [3]; scale bar is 1 mm).

orientation of the stimulus. The coloured map on the right represents the preferred orientation of neurons in each location. Thus the black bar shown at left will activate V1 neurons coloured in purple in the specific square. In theory, these Gabor-like filters together can carry out neuronal processing of spatial frequency, orientation, motion, direction, speed, and many other spatiotemporal features. Similar maps could be plotted for this same area showing preference for other visual features.

Prestriate Cortex: V2

Visual area V2, also called prestriate cortex [27], is the second major area located in the occipital lobe of the primate brain, and the first region within the visual association area [28]. It receives strong feed-forward connections from V1 and has many properties in common with V1.

The responses of V2 neurons are tuned to simple shapes such as orientation and sinusoidal gratings. Moreover, V2 neurons are able to represent variety of higher order shapes that are based on contours (e.g., angles and curves with multiple orientations at different subregions within a single receptive field) or grating patterns [29]. The responses of many V2 neurons are also modulated for complex properties: orientation of illusory contours [30], binocular disparity [31], and whether the stimulus is part of the figure or the ground [32].

Visual Area V4

Area V4 is the third cortical area in the ventral stream, receiving strong feedforward input from V2 and transmitting to the PIT. It also receives direct inputs from V1 which are mostly generated in the visual central space. V4 is the first area in the ventral stream to show strong attentional modulation. Most studies indicate that selective attention can change firing rates in V4 by about 20% [33]. This discovery found by Moran and Desimone [34] was the first report of attention effects anywhere in the visual cortex.

Although V4 is mainly modulated for colour recognition, it is also tuned for orientation and spatial frequency similar to V1. Comparing to V1, V4 responds to more complex object features with intermediate complexity but is not tuned for complex objects as areas in the inferotemporal cortex are [35].

Inferotemporal Cortex: IT

Inferotemporal Cortex is only found in the temporal lobe in primates including humans. It is tuned to a range of object features complexity starting with simpler patterns in the PIT/TEO [36]; And the complexity increases along the ventral stream towards AIT/TE where objects are represented and recognised [37]. The high-order complex features includes the combinations of colour or texture with complicated shapes [36], and body parts such as faces and hands [38]. The distinguishing features of the IT cortex is that the neuronal responses are position and size invariant [39, 40], and also invariant to changes in luminance, texture, and relative motion [41, 42]. It is wide-accepted that the identity-preserving transformation invariance makes IT ideal for representing objects despite changes in the surrounding environment and retinal image.

In the next section, this report will explore the detailed mechanism of object representation in this cortical area.

2.1.2 Object Representation in IT

The neuronal representation in the cortical area of IT is considered to be the spatio-temporal pattern of spikes. The spiking activities of single neurons and populations are thought to hold the key to encode visual information. In Section 2.2, the report will introduce single neuron models and spike coding mechanisms in computational spiking neural networks.

Single neurons

Most studies have investigated the neural activities in the IT by means of firing rate or spike count. A typical histogram, Figure 2.3(A) [43], shows the spike count of a single neuron in time bins of 25 ms for a duration of 300 ms in total right after the presentation of a visual image. The highlighted time window, the so-called 'decoding' window, is adjusted to the latency of the conductance along the ventral stream. The spike count of the 'decoding' window is well modulated for object identity, position or size [44, 45], see example in Figure 2.3(B) where the left shows the spiking activities for clean figures and the right for natural scenes. The neural responses were sorted from high to low with the corresponding figures presented, where the red point

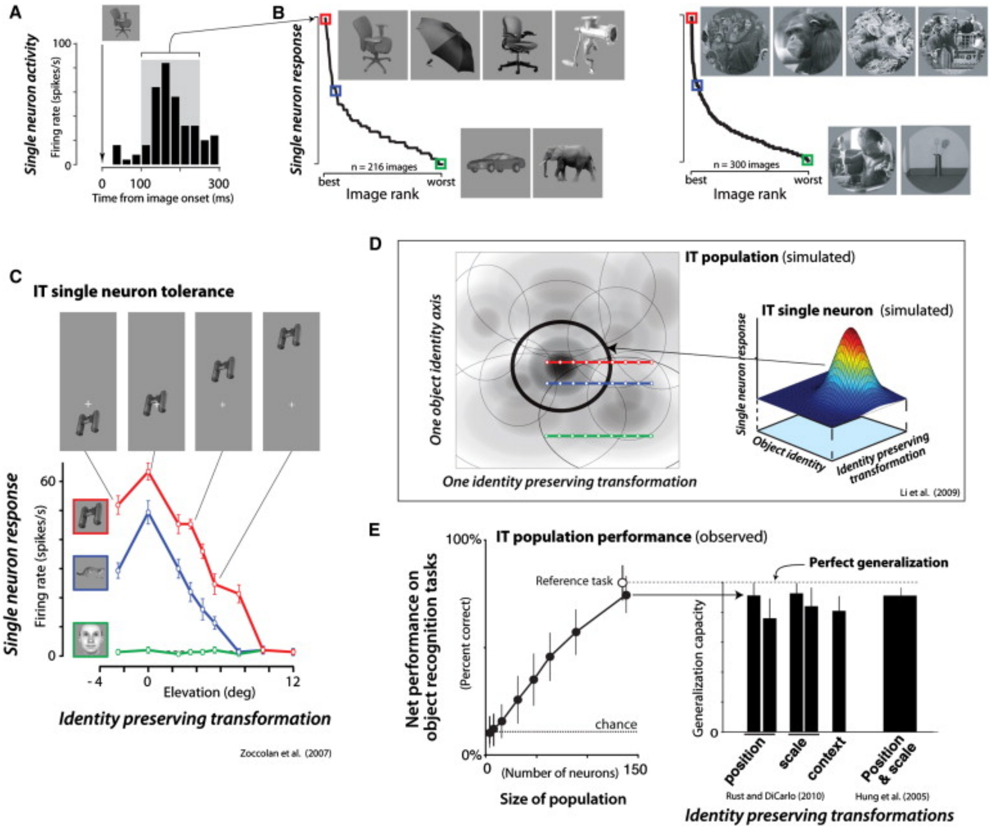


Figure 2.3: IT single-neuron properties and their relationship to population performance [4].

indicated the highest respond while the green the lowest and the blue the medium. Another example in Figure 2.3(C) shows the responses of an example IT neuron obtained by varying the position (elevation) of three objects with high (red), medium (blue), and low (green) activities. The object identity preference was maintained in the entire test range regardless of the position changes. These tuning curves are similar to the well-understood firing rate modulation in visual area V1 on the bar orientation.

Understanding IT single neuron responses has proven to be extremely challenging and even predicting the responses of an IT neuron to a new image remains impossible. Nevertheless, IT neurons are activated by complex combinations of visual features and that they are often able to maintain their relative object preference over small to moderate changes in object position, scale, pose [46], illumination [47] and clutter [48].

Although IT neurons are commonly described as narrowly selective object identifier, neurophysiological studies have shown a diverse selectivity of single neurons [44]. Most IT neurons are broadly tuned and the typical IT neuron responds to many different images and objects [43], also see Figure 2.3(B). As illustrated in Figure 2.3(D), a single neuron (right) is modulated to both object identities and variables of identity-preserving transformations. To explain the plot in 2.3(C), position is the variable here; thus the tuning curve for different identities on each position can be described as a slice in the 3-D plot which is Gaussian-like. As a result, the rank order of the three objects remains the same due to the Gaussian-like curve stays similar. If a population of such IT neurons tiles with the overlapping fashion, see left panel of Figure 2.3(D), a more accurate recognition result containing the transformation parameter can be carried out with population coding.

Population of neurons

Spike timing variability in the ms resolution of spikes is consistent with a Poisson-like stochastic spike generation process. The underlying output rate of IT neurons is determined by each particular image. Despite the timing variability, the brain can reliably recognise the presented object by integrating the neural responses across IT population [49]. However, it still remains unclear whether the spike timing variability brings down the encoding/decoding accuracy or if it contributes to the population tuning for useful informations [50].

Although the first stage of the ventral stream, V1, is reasonably well studied, the visual processing in higher stages especially in V4 and IT remains poorly understood.

Nevertheless, as stated above IT is the main part of ventral stream to recognise and categorise the objects in real-time and is tolerant to identity-preserving transformations. Specifically, simple linear classifier built on the output rates of randomly selected population with only a few hundred neurons reveals a high-level of object recognition performance [51]; and the simple weighted summation explains a wide range of invariant object recognition behaviour sufficiently [52].

Figure 2.3(E) shows the direct tests of measuring the cross-validated population performance on categorisation tasks using linear classifiers. The recognition performance approaches ceiling level with only a few hundred neurons (left panel), and the same population shows a good generalization across moderate changes in position, scale, and context.

Decoding Window Matters

The output spiking pattern of the ventral visual stream are well described by a firing rate code where the decoding window size is 50 ms [51]. Thus the visual representation in IT is usually found in the first 50 ms of neuronal response, although different time epochs relative to stimulus onset may encode different types of visual information [53] (see Figure 2.3(A), an appropriate decoding window can be 100150 ms after image onset).

In sum, the output of the ventral stream is reflexively expressed in neuronal firing rates across a short interval of 50 ms and is an explicit object representation; and the rapid production of this representation is consistent with a largely feed-forward, non-linear processing of the visual input [4], which is described in the following section.

2.1.3 Hierarchical Feed-forward Organisation

Figure 2.4(A) illustrates the ventral stream cortical area locations in the macaque monkey brain, and the flow of visual information from the retina. The corresponding hierarchical organisation is showed in Figure 2.4(B). Each area is plotted with the size proportional to its cortical surface size. Approximate total number of neurons of both hemispheres is shown in the corner of the cortical areas. The approximate number of projections is written above each block. In addition, the colour dedicates to processing the central 10° of the visual field. At last, approximate median response latency is listed on the right.

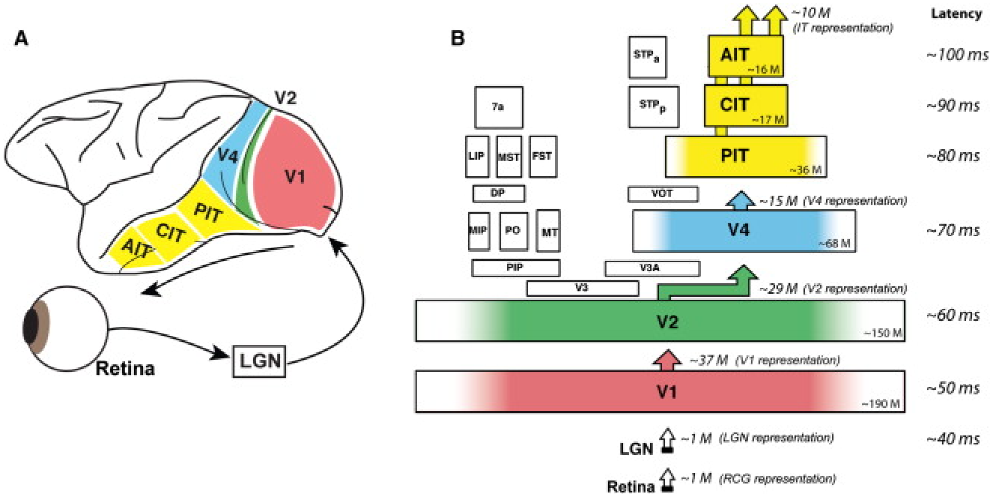


Figure 2.4: The ventral visual pathway and its hierarchical organization [4].

Latency

Each cortical area along the ventral stream contributes a conductance latency of about 10 ms of the visual information [54]. Thus, just around 100 ms after images appeared in front of the retina, a first wave of object identity neuronal activity is present throughout much of IT (e.g., Figure 2.3(A)).

Neurons and Connections

Because retinal and LGN receptive fields are point-wise spatial sensors, the object visual information conveyed to V1 are nearly as raw as the pixel representation (1 million pixels). As V1 carries out its visual process, the total object representation increases approximately 30-times [55] because of its non-linear filtering. This dimensionality expansion results in an overcomplete population re-representation [56] in which the object representation vectors have more dimensions than the LGN input. As a result, simulations show that a V1-like representation is clearly better than RGN-like/pixel-based representation, but still far below human performance for real-world recognition problems [16].

The output projections of each area decreases from V2 (about 29 million) to AIT which represents the object with 10 million dimensions. At the same time, the receptive field size of neurons increases to complete the object representation with a whole image and to perform invariant recognition.

Tuned Features and Receptive Fields

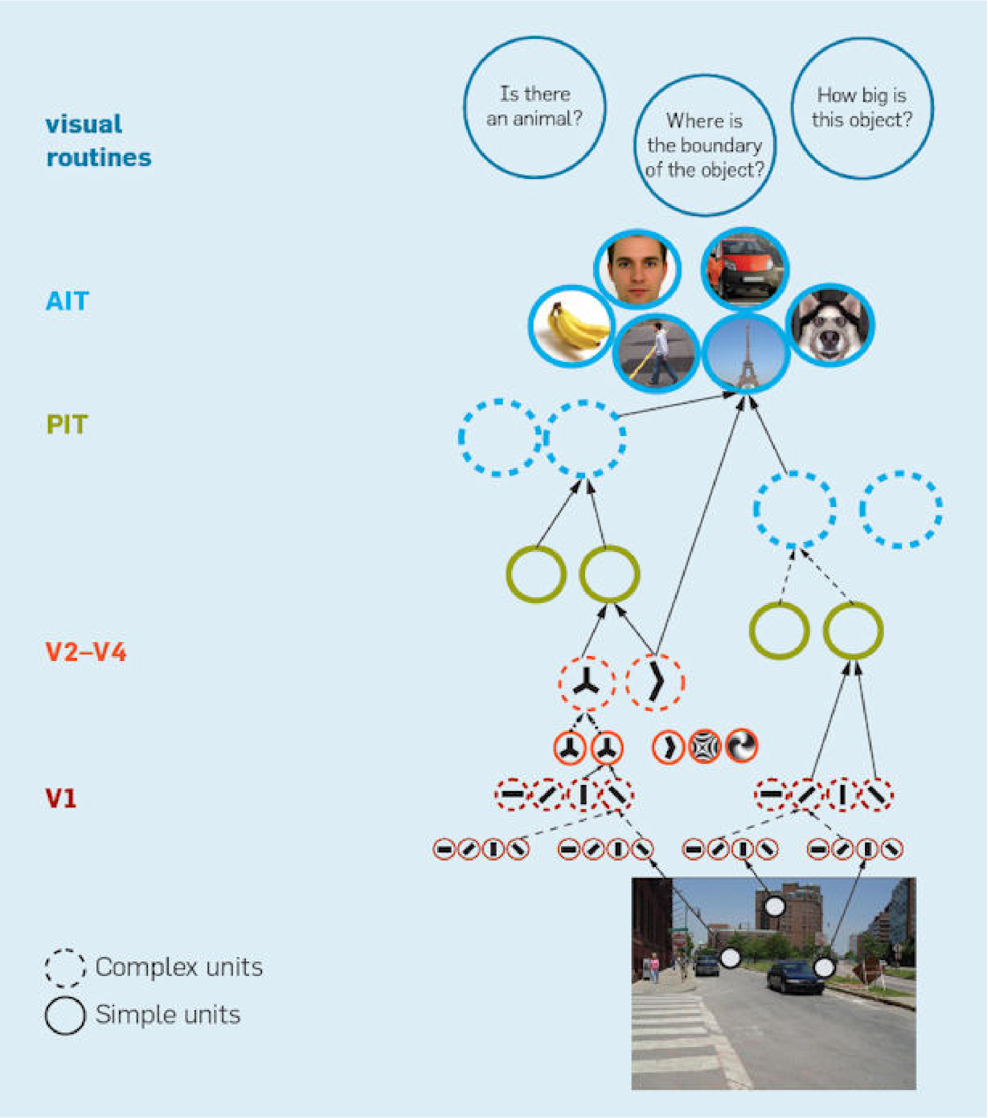


Figure 2.5: The hierarchical ventral stream and the corresponding tuned features for each layer [5].

As the visual information conducts along the ventral stream, neurons become selective for stimuli that are increasingly complex from simple oriented bars and edges in early visual area V1 to moderately complex features in intermediate areas: V2, V4 and PIT to complex objects and faces in AIT, see Figure 2.5. Along with this growing complexity of the preferred stimulus, the invariance properties of neurons also increase. Neurons become more and more tolerant with respect to the exact position and scale of the stimulus within their receptive fields. As a result, the receptive field size of neurons increases from about one degree or less in V1 to several degrees in IT, see Figure 2.6.

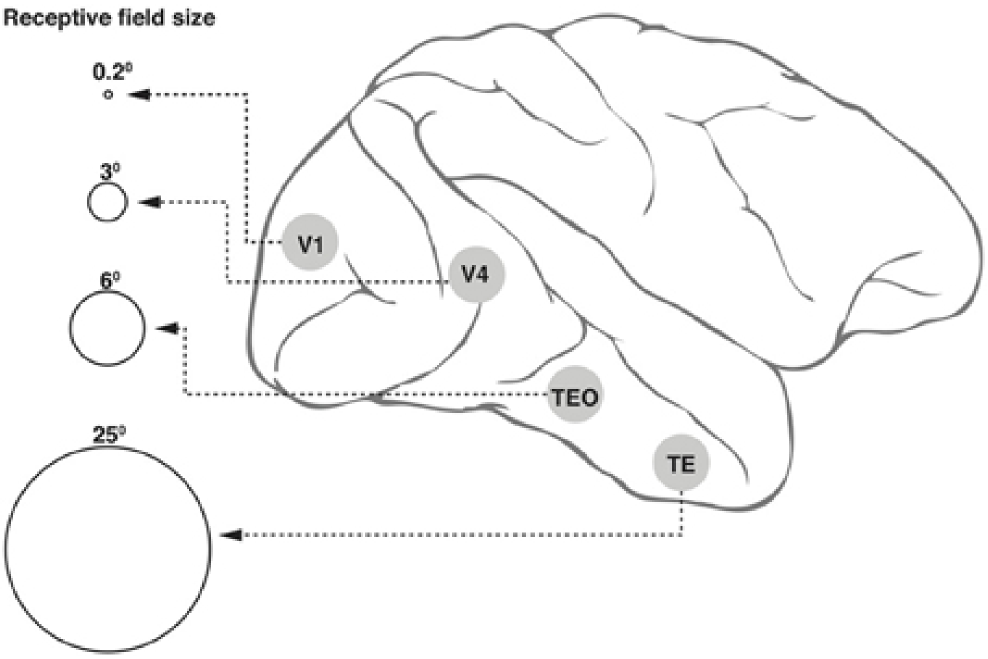


Figure 2.6: Receptive field (RF) sizes along the ventral cortical stream in the primate. While the degree of complexity of processing may increase, the RF size at any one eccentricity also increases dramatically along the various cortical areas from V1 into the temporal pole. The circles shown in the figure are not drawn to scale, but the numbers above the circles indicate approximate relative sizes of the RF diameters. [6].

2.2 Spiking Neural Network

The so-called third generation of neural networks [57] introduces a different set of functions and parameters to model neurons; these both model biological neurons more precisely [58] and increase the computational power of networks of neurons if compared to classical sigmoidal units. Such networks rely on the propagation of an all-or-none signal, the action potential, which asynchronously carries information to its connected units by means of its timing.

2.2.1 Neuronal Model

The level of biological detail of such models varies greatly but many models build on the ‘leaky integrate and fire’ (LIF) model. Spikes arriving at a LIF neuron cause a temporary flow of current into or out of the neuron, modelling the behaviour of synapses in biological neurons. The LIF neuron integrates this current over time, accumulating charge which gradually leaks away. If the charge in the neuron reaches a certain threshold, the neuron produces a spike and its charge is reset.

2.2.2 Learning

One of the key parameters of a neural network is the amount of influence each incoming spike has on a neuron. Typically, this influence is modelled by assigning a ‘weight’ to each synapse which scales the impact of a spike arriving via that synapse. Models of many types of learning revolve around modelling changes in weights over long periods of time observed within the brain. The exact rules by which these weights are adjusted is the subject of much active research though most promising approaches attempt to learn from the relative timing [59] or rate [60] of spikes arriving at a neuron. As well as adjusting weights, some learning rules can also form entirely new connections between previously unconnected neurons [61].

2.2.3 Successful Applications

Numerous applications using SNN-based vision processing have been successfully carried out in the past. A dual-layer SNN has been trained using Spike Time Dependent Plasticity (STDP) and employed for character recognition [62]. Lee et al. [63] have implemented direction selective filters in real time using spiking neurons, considered as a convolution layer in the model of a so called CNN [64]. Different features, such as Gabor filter features (scale, orientation and frequency) and shape can be modelled as layers of feature maps. The similar behaviours have been found in the primary visual cortex (V1) in the visual pathway [65] as the foundation for higher level visual process e.g. object recognition. Rank order coding, as an alternative to conventional rate-based coding, treats the first spike as the most important and has been successfully applied to an orientation detection training process [66]. Nengo [67] is a graphical and scripting based software package for simulating large-scale neural systems and has been used to build the world’s largest functional brain model, Spaun [68]. An FPGA implementation of a Nengo model for digit recognition has been reported [69]. Deep Belief

Networks (DBNs), the 4th generation of artificial neural network, have shown great success in solving classification problems. Recent study [70] in this area has mapped an offline-trained DBN onto an efficient event-driven spiking neural network for digit recognition tasks with resounding success.

2.3 Platforms

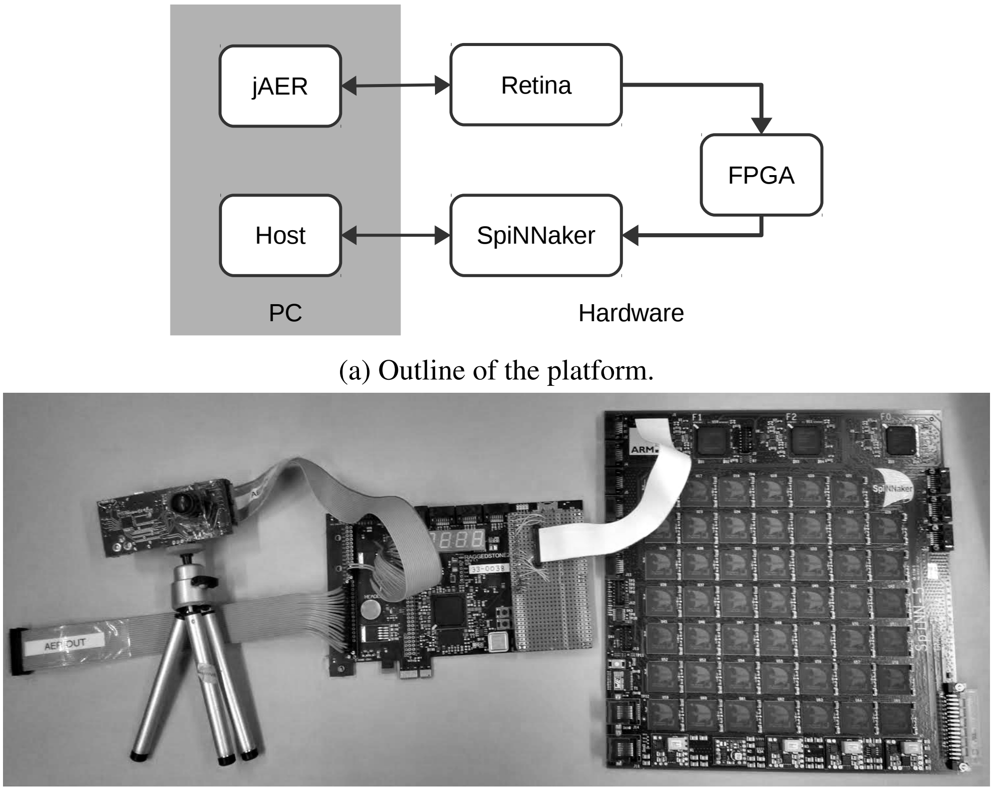
The outline of the platform is illustrated in Figure 2.7a, where the hardware system is configured, controlled and monitored by the PC. The jAER [71] event-based processing software on the PC configures the retina and displays the output spikes through a USB link. The host communicates to the SpiNNaker board via Ethernet to set up its runtime parameters and to download the neural network model off-line. It visualises [72] the spiking activity of the network in real-time. The photograph of the hardware platform, Figure 2.7b, shows that the silicon retina connects to the SpiNNaker 48-node system via a Spartan-6 FPGA board [73].

2.3.1 Vision Processing Front-ends

The visual input is captured by a DVS silicon retina, which is quite different from conventional video cameras. Each pixel generates spikes when its change in brightness reaches a defined threshold. Thus, instead of buffering video into frames, the activity of pixels is sent out and processed continuously with time. The communication bandwidth is therefore optimised by sending activity only, which is encoded as pixel events using Address-Event Representation (AER [74]) protocol. The level of activity depends on the contrast change; pixels generate spikes faster and more frequently when they are subject to more active change. The sensor is capable of capturing very fast moving objects (e.g., up to 10 K rotations per second), which is equivalent to 100 K conventional frames per second [17].

2.3.2 SNNs Back-ends

The SpiNNaker project's architecture mimics the human brain's biological structure and functionality. This offers the possibility of utilizing massive parallelism and redundancy, as the brain, to provide resilience in an environment of unreliability and failure of individual components.



(a) Outline of the platform.

(b) Picture of the hardware platform. From left to right: a silicon retina, a FPGA board, and a 48-node SpiNNaker system.

Figure 2.7: System overview of the dynamic hand posture recognition platform.

In the human brain, communication between its computing elements, or neurons, is achieved by the transmission of electrical 'spikes' along connecting axons. The biological processing of the neuron can be modelled by a digital processor and the axon connectivity can be represented by messages, or information packets, transmitted between a large number of processors which emulate the parallel operation of the billions of neurons comprising the brain.

The engineering of the SpiNNaker concept is illustrated in Figure 2.8 where the hierarchy of components can be identified. Each element of the toroidal interconnection mesh is a multi-core processor known as the 'SpiNNaker Chip' comprising 18 processing cores. Each core is a complete processing sub-system with local memory. It is connected to its local peers via a Network-on-Chip (NoC) which provides high bandwidth on-chip communication and to other SpiNNaker chips via links between them. In this way massive parallelism extending to thousands or millions of processors is possible.

The '103 machine' is the name given to the 48-node board which we use for the

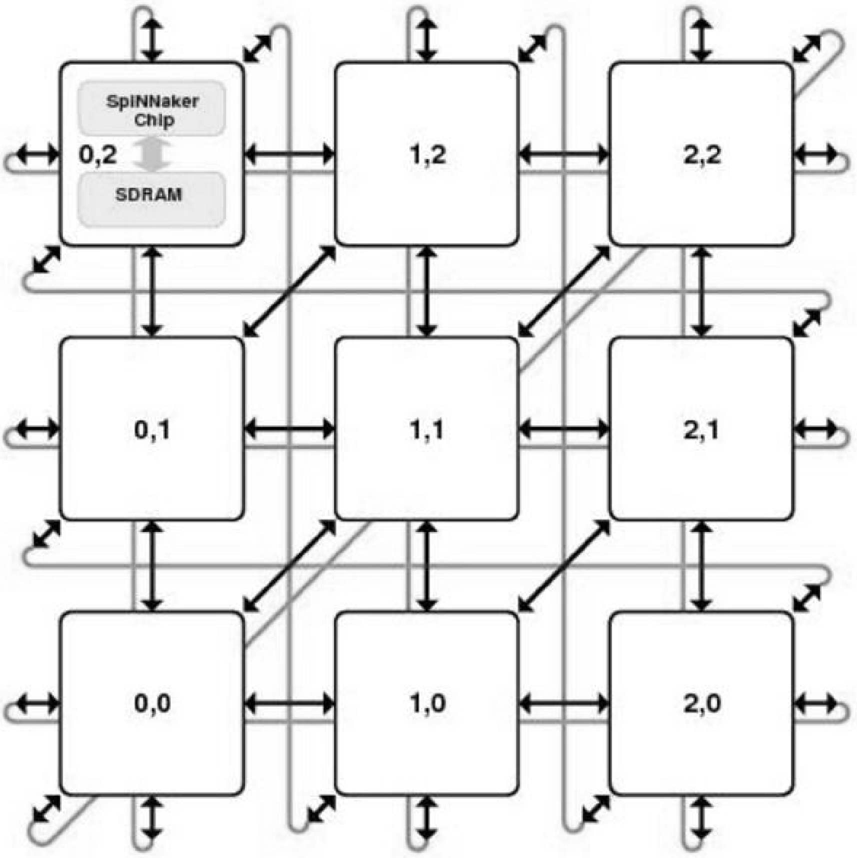


Figure 2.8: SpiNNaker system diagram. Each element represents one chip with local memory. Every chip connects to its neighbours through the six bi-directional on-board links.

hand posture recognition system, see Figure **??**. It has 864 ARM processor cores, typically deployed as 768 application, 48 monitor and 48 spare cores. The boards can be connected together to form larger systems using high-speed serial interfaces.

2.3.3 SpiNNaker distinguishing features

Spikes from the silicon retina are injected directly into SpiNNaker via a SPARTAN-6 FPGA board that translates them into a SpiNNaker compatible AER format [75].

From a neural modelling point of view, interfacing the silicon retina is performed using pyNN [76]. The retina is configured as a spike source population that resides on a virtual SpiNNaker chip, to which an AER sensor's spikes are directed, thus abstracting away the hardware details from the user[73]. Besides the retina, we have successfully connected an AER based silicon cochlea [77] to SpiNNaker for a sound localisation task [78], see Figure 2.9.

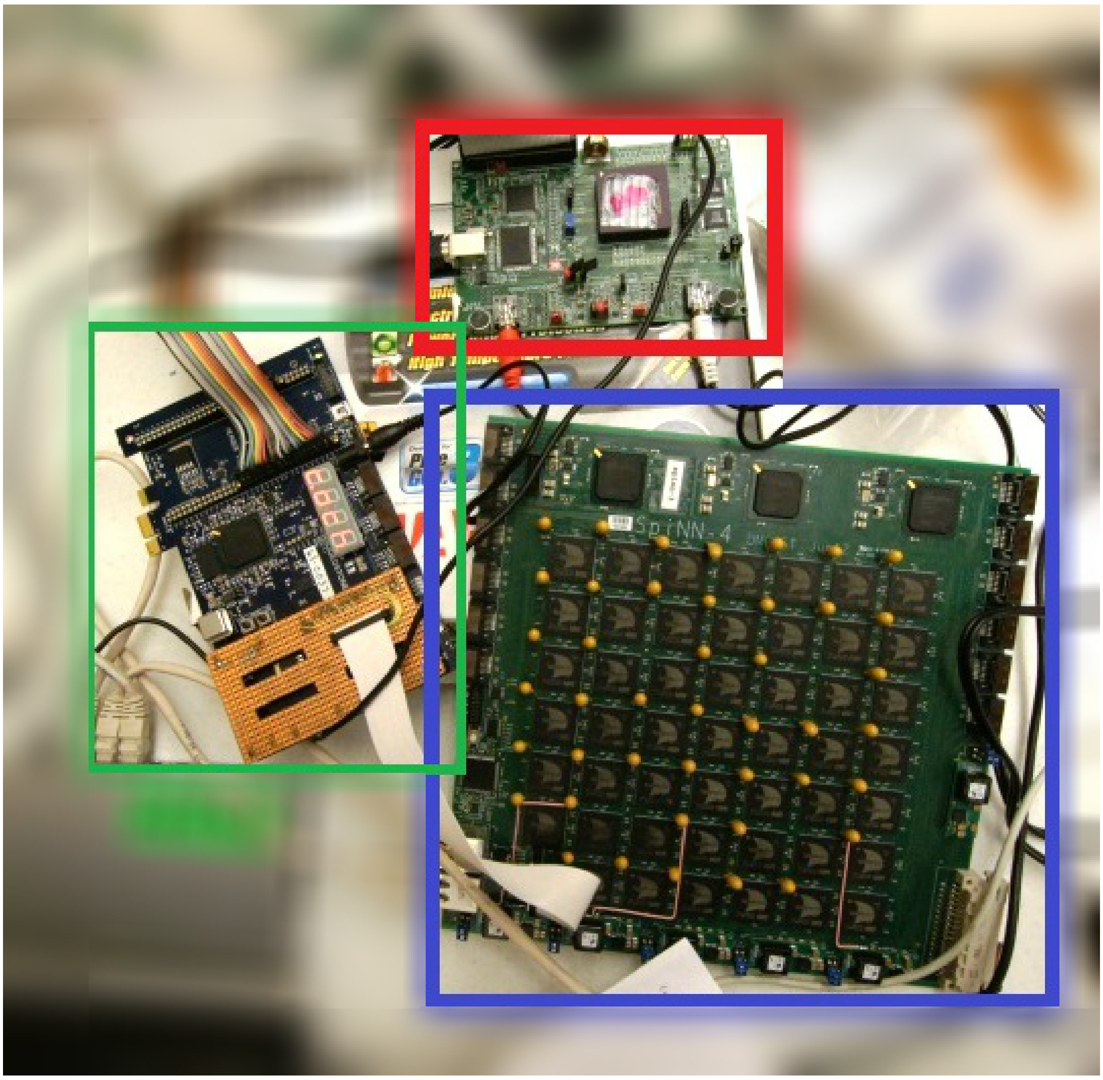


Figure 2.9: Neuromorphic platform for sound localisation: a silicon cochlea connects to a 48-node SpiNNaker board via a FPGA.

Chapter 3

Convolutional Neural Networks

The convolutional network is well-known as an example of a biologically-inspired model. Figure 3.1 shows a typical convolutional connection between two layers of neurons. The repeated convolutional kernels are overlapped in the receptive fields of the input neurons.

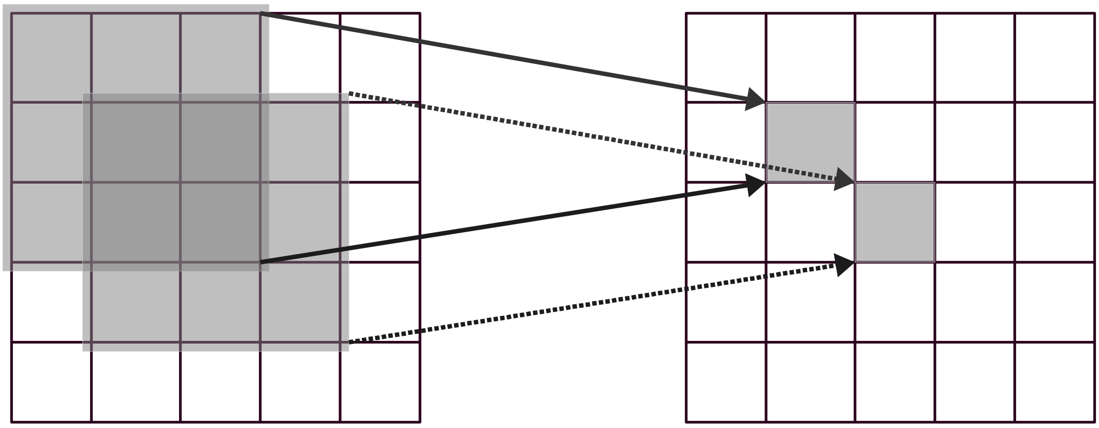


Figure 3.1: Each individual neuron in the convolution layer (right matrix) connects to its receptive field using the same kernel. The value of the kernel is represented by the synaptic weights between the connected neurons.

3.1 Model Description

There are two CNNs proposed to accomplish the dynamic hand posture recognition task. A straight forward method of template matching is employed at first, followed by a network of multi-layer perceptrons (MLP) trained to improve the recognition performance.

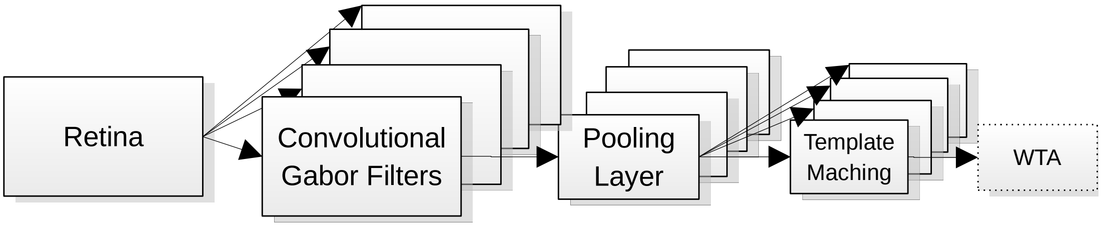


Figure 3.2: Model 1. The retina input is convolved with Gabor filters in the second layer, and then shrinks the sizes in the pooling layer. The templates are considered as convolution kernels in the last layer. The WTA circuit can be used as an option to show the template matching result more clearly.

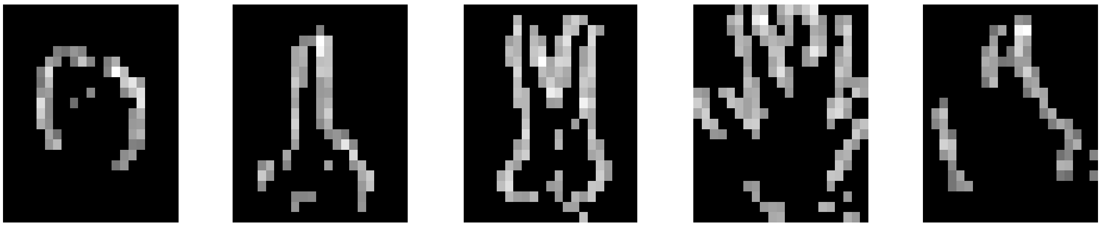
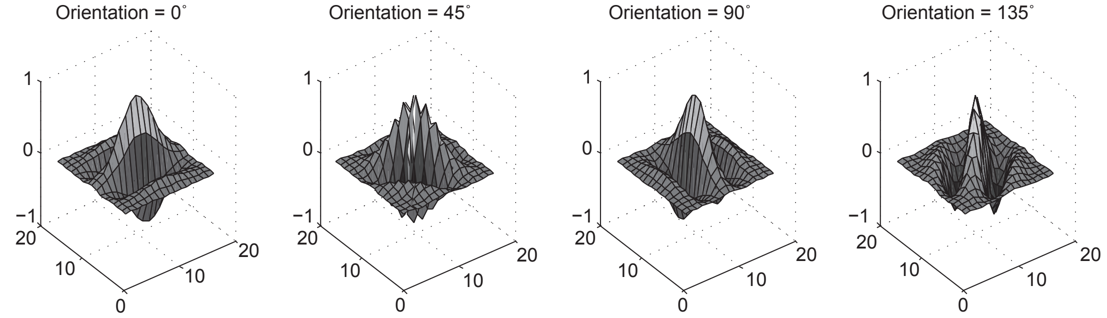


Figure 3.3: Templates of the five postures: ‘Fist’,‘Index Finger’, ‘Victory Sign’, ‘Full Hand’ and ‘Thumb up’.

Model 1: Template Matching. Shown in Figure 3.2 the first layer is the retina input, followed by the convolutional layer, where the kernels are Gabor filters responding to edges of four orientations. The third layer is the pooling layer where the size of the populations shrinks. This down-sampling enables robust classification due to its tolerance to variations in the precise shape of the input. The fourth layer is another convolution layer where the output from the pooling layer is convolved with the templates. The optional layer of Winner-Take-All (WTA) neurons enables a clearer classification result due to the inhibition between the neurons. In the Matlab simulation, the retina input spikes are buffered into 30 ms frames, and the neurons are simple linear perceptrons. The templates are chosen by sampling the output of the pooling layer when given some reference stimulus, see Figure 3.3.

The Gabor filter is well-known as a linear filter for edge detection in image processing. A Gabor filter is a 2D convolution of a Gaussian kernel function and a sinusoidal


Orientation = 0°

Orientation = 45°

Orientation = 90°

Orientation = 135°

Figure 3.4: Real parts of the Gabor filters orienting four directions.

plane wave; see Equation 3.1.

$$
\begin{aligned}
&\text{RealParts} = \exp\left(\frac{-x'^2+y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}\right)\\
&\text{ImaginaryParts} = \exp\left(\frac{-x'^2+y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}\right)\\
&\text{where :}\\
&x' = x\cos(\theta)+y\sin(\theta)\\
&y' = -x\sin(\theta)+y\cos(\theta)
\end{aligned}
\tag{3.1}
$$

θ represents the orientation of the filter, λ is the wavelength of the sine wave, and σ is the standard deviation of the Gaussian envelope. The frequency and orientation features are similar to the responses of V1 neurons in the human visual system. Only the real parts of the Gabor filters (see Figure 3.4) are used as the convolutional kernels to configure the weights between the input layer and the Gabor filter layer.

The output score of a convolution is determined by the matching degree between the input and the kernel. Regarding the template matching layer, each neuron in a population responds to how closely its receptive field matches the specific template. The position of moving gesture is also naturally encoded in the address of template matching neuron. Thus, there are five populations of template matching neurons, one for each hand posture listed.

Model 2: Trained MLP. Inspired by the research of Lecun [79], we designed a combined network model with MLP and CNN (Figure 3.5). The first three layers are

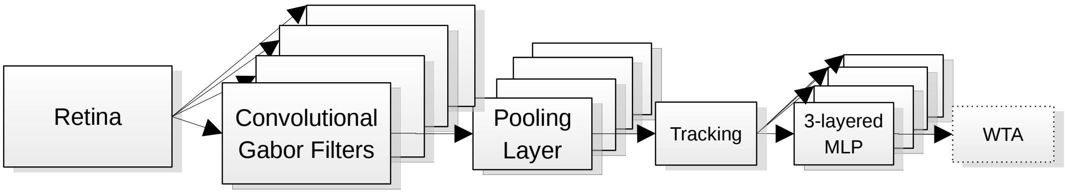


Figure 3.5: Model 2. The retina input convolves with Gabor filters in the second layer, and then shrinks the sizes in the pooling layer. The following tracking layer finds the most active area of some fixed size, moves the posture to the centre and pushes the image to the trained MLP. The winner-take-all (WTA) layer can be used as an option to show the template matching result more clearly.

exactly the same as the previous model. The training images for the 3-layered MLP are of same size and the posture is centred in the images. Therefore, a tracking layer plays an important role to find the most active region and forward the centred image to the next layer.

3.2 Experimental Set-up

In order to evaluate the cost and performance trade-offs in optimizing the number of neural components, both the convolutional models described above are tested at different scales. Five videos of every posture are captured from the silicon retina in AER format, all of similar size and moving clock-wise in front of the retina. The videos are cut into frames (30 ms per frame) and presented to the convolutional networks. The configurations of the networks are listed in Table 3.1. The integration layer is not necessary in a convolutional network, but is used here to decrease the number of synaptic connections.

3.3 Experimental Results

In Figure 3.6 the first two plots refer to Model 1, using template matching. Each colour represents one of the recognition populations. Each point in the plot is the highest neuronal response in the recognition population during the time of one frame (30 ms). The neuronal response, ‘the spiking rate’, is normalised to [-1, 1]. It can be seen that the higher resolution input makes the boundaries between the classes clearer. On the other hand, recognition only happens when the test image and template are similar enough. The templates are only selected from the frames where the gestures

Table 3.1: Sizes of the convolutional neural networks.

(a) Model 1: Template matching

	Full Resolution **128 × 128**		**Sub-sampled Resolution** **32 × 32**	
	Population Size	Connections per Neuron	Population Size	Connections per Neuron
Retinal Input	128 × 128	1	32 × 32	4 × 4
Gabor Filter	112×112×4	17 × 17	28×28×4	5 × 5
Pooling Layer	36×36×4	5 × 5	null	null
Integration Layer	36 × 36	4	28 × 28	4
Template Matching	16×16×5	21 × 21	14×14×5	15 × 15
Total	74,320	15,216,512	5,925	318,420

(b) Model 2: Trained MLP

	Full Resolution **128 × 128**		**Sub-sampled Resolution** **32 × 32**	
	Population Size	Connections per Neuron	Population Size	Connections per Neuron
Tracked Input	21 × 21	null	15 × 15	null
Hidden Layer	10	21×21×10	10	15×15×10
Recognition Layer	5	5×10	5	5×10
Total	456	4,460	240	2,300

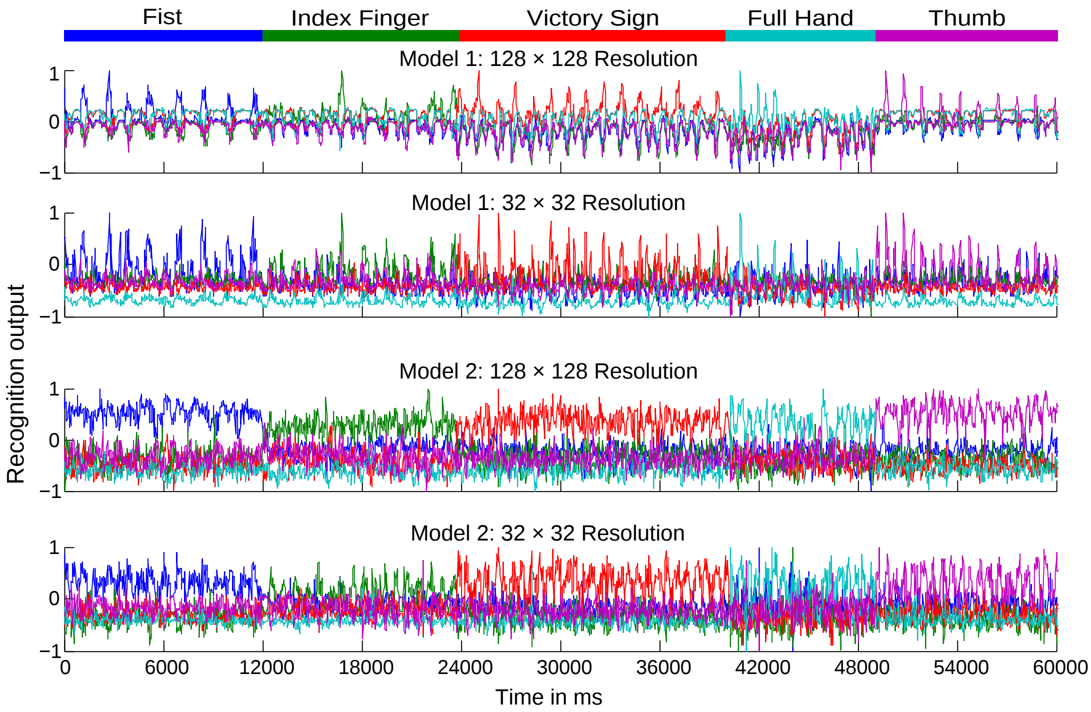


Figure 3.6: Neural responses with time of four experiments to the same recorded moving postures. The recognition output is normalised to [-1, 1]. Every point represents the highest response in a specific population (different colour) for a 30 ms frame. The 1st plot refers to Model 1 with the full input resolution, and the 2nd plot Model 1 with the sub-sampled input resolution; and the 3rd and fourth plots both refer to Model 2, and with high and low input resolution respectively.

are moving towards the right, and the gestures are moving clockwise in the videos, thus, all the peaks in plot 1 correspond with moments when the gesture moves towards right. It is notable that the higher resolution causes the recogniser to be more sensitive to the differences between the test data and the template, while the smaller neural network can recognize more generalized patterns. Therefore, a threshold is required to differentiate between data that is close enough and that which is not. Since the gestures are moving in four different directions during the clockwise movement, a rejection rate (i.e. none of the template is matched) of 75% is to be expected.

The latter two plots of Figure 3.6 refer to Model 2. The three-layer MLP network significantly improves the recognition rate and can generalise the pattern. There is no rejection rate for Model 2, since the MLP is trained with all the moving directions of the postures.

Detailed results are listed in Table 3.2. The correct recognition rate is calculated

Table 3.2: Recognition results using linear perceptrons in %

		Model 1		**Model 2**	
		High Resolution	Low Resolution	High Resolution	Low Resolution
Fist	Correct	99.11	99.23	96.24	84.21
(399 Frames)	Reject	71.93	67.42	Null	Null
Index Finger	Correct	92.98	80.00	94.39	71.69
(392 Frames)	Reject	70.92	75.77	Null	Null
Victory Sign	Correct	96.56	93.07	95.64	87.66
(551 Frames)	Reject	73.68	81.67	Null	Null
Full Hand	Correct	95.65	72.41	93.52	72.01
(293 Frames)	Reject	92.15	90.10	Null	Null
Thumb up	Correct	89.61	84.44	96.68	74.68
(391 Frames)	Reject	80.31	76.98	Null	Null
Average	Correct	94.78	85.83	95.29	78.05
	Reject	77.80	78.39	Null	Null

from the non-rejected frames. The lower resolution of the 32×32 retina input is adequate (85.83%) for this gesture recognition task. The smaller network uses only 1/10th the number of neurons and 1/50th the number of synaptic connections compared with the full resolution network, while the recognition rate drops only around by 9.0% with Model 1 and 17.2% with Model 2.

Chapter 4

Recognition on SpiNNaker

4.1 Moving from Perceptrons to Spiking Neurons

It remains a challenge to transform traditional artificial neural networks into spiking ones. There are attempts [80] [81] to estimate the output firing rate of the LIF neurons (Equation 4.1) under certain conditions.

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\frac{V(t) - V_{rest}}{\tau_m} + \frac{I(t)}{C_m} \tag{4.1}$$

The membrane potential V changes in response to input current I, starting at the resting membrane potential V_{rest}, where the membrane time constant is $\tau_m = R_m C_m$, R_m is the membrane resistance and C_m is the membrane capacitance.

Given a constant current injection I, the response function, i.e. firing rate, of the LIF neuron is

$$\lambda_{out} = \left[t_{ref} - \tau_m \ln \left(1 - \frac{V_{th} - V_{rest}}{IR_m} \right) \right]^{-1} \tag{4.2}$$

when $IR_m > V_{th} - V_{rest}$, otherwise the membrane potential cannot reach the threshold V_{th} and the output firing rate is zero. The absolute refractory period t_{ref} is included, where all input during this period is invalid. In a more realistic scenario, the post-synaptic potentials (PSPs) are triggered by the spikes generated from the neuron's pre-synaptic neurons other than a constant current. Assume that the synaptic inputs are Poisson spike trains, the membrane potential of the LIF neuron is considered as a diffusion process. Equation 4.1 can be modelled as a stochastic differential equation referring to Ornstein-Uhlenbeck process,

$$\tau_m \frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\left[V(t) - V_{rest}\right] + \mu + \sigma\sqrt{2\tau_m}\xi(t) \tag{4.3}$$

where

$$\begin{aligned} \mu &= \tau_m(\mathbf{w_E} \cdot \lambda_\mathbf{E} - \mathbf{w_I} \cdot \lambda_\mathbf{I}) \\ \sigma^2 &= \tfrac{\tau_m}{2}\left(\mathbf{w_E^2} \cdot \lambda_\mathbf{E} + \mathbf{w_I^2} \cdot \lambda_\mathbf{I}\right) \end{aligned} \tag{4.4}$$

are the conditional mean and variance of the membrane potential. The delta-correlated process $\xi(t)$ is Gaussian white noise with zero mean, $\mathbf{w_E}$ and $\mathbf{w_I}$ stand for the weight vectors of the excitatory and the inhibitory synapses, and λ represents the vector of the input firing rate. The response function of the LIF neuron with Poisson input spike trains is given by the Siegert function [82],

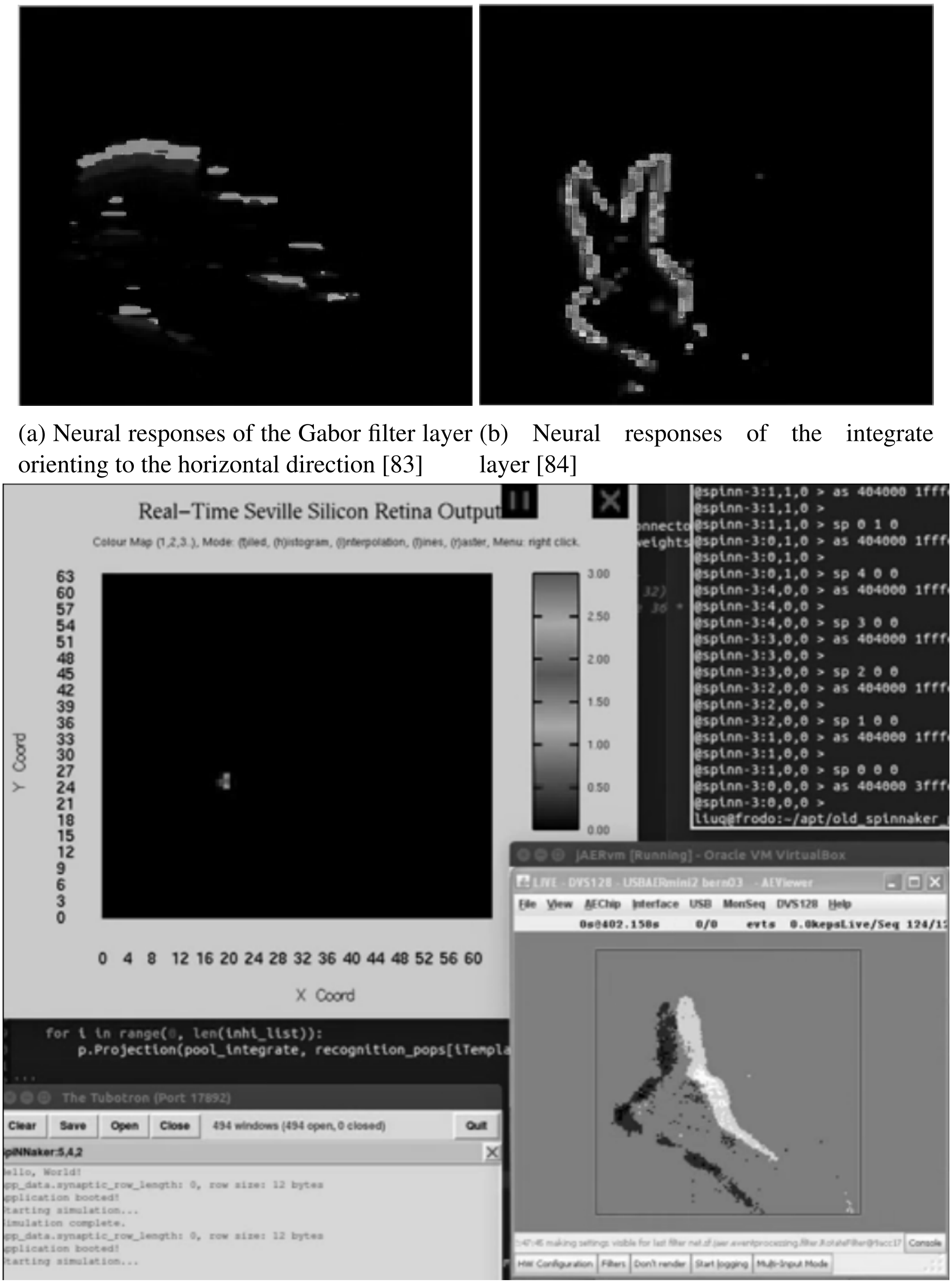
$$\lambda_{out} = \left(\tau_{ref} + \frac{\tau_Q}{\sigma_Q}\sqrt{\frac{\pi}{2}}\int_{V_{rest}}^{V_{th}} du \exp\left(\frac{u-\mu_Q}{\sqrt{2}\sigma_Q}\right)^2 \left[1 + erf\left(\frac{u-\mu_Q}{\sqrt{2}\sigma_Q}\right)\right]\right)^{-1} \tag{4.5}$$

where τ_Q, μ_Q, σ_Q are identical to τ_m, μ, σ in Equation 4.4, and erf is the error function.

Still there are some limitations on the response function. For the diffusion process, only small amplitude (weight) of the PostSynaptic Potentials (PSPs) generated by a large amount of input spikes (high spiking rate) work under this circumstance; plus, the delta function is required, i.e. the synaptic time constant is considered to be zero. Thus only a rough approximation of the output spike rate has been determined. Secondly, given different input spike rate to each pre-synaptic neurons, the parameters of the LIF neuron and the output spiking rate, how to tune every single corresponding synaptic weight remains a difficult task.

4.2 Live Recognition

We implemented the prototype of the dynamic posture recognition system on SpiNNaker using LIF neurons. The input retina layer consists of 128×128 neurons; each Gabor filter has 112×112 valid neurons, since the kernel size is 17×17; each pooling layer is as big as 36×36, convolving with five template kernels (21×21); thus, the recognition populations are 16×16 neurons each. Altogether $74,320$ neurons and



(a) Neural responses of the Gabor filter layer orienting to the horizontal direction [83]

(b) Neural responses of the integrate layer [84]

Real−Time Seville Silicon Retina Output
Colour Map (1,2,3..), Mode: (t)iled, (h)istogram, (i)nterpolation, (l)ines, (r)aster, Menu: right click.
Y Coord
X Coord
63 60 57 54 51 48 45 42 39 36 33 30 27 24 21 18 15 12 9 6 3 0
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60
3.00 2.50 2.00 1.50 1.00 0.50 0.00
JAERvm [Running] - Oracle VM VirtualBox
LIVE - DVS128 - USBAERmini2 bern03 - AEViewer
File View AEChip Interface USB MonSeq DVS128 Help
0s@402.158s 0/0 evts 0.0kepsLive/Seq 124/1
Console
HW Configuration Filters Don't render Start logging Multi-Input Mode
The Tubotron (Port 17892)
Clear Save Open Close 494 windows (494 open, 0 closed) Quit
piNNaker:5,4,2

```
@spinn-3:1,1,0 > as 404000 1fff
@spinn-3:1,1,0 >
@spinn-3:1,1,0 > sp 0 1 0
@spinn-3:0,1,0 > as 404000 1fff
@spinn-3:0,1,0 >
@spinn-3:0,1,0 > sp 4 0 0
@spinn-3:4,0,0 > as 404000 1fff
@spinn-3:4,0,0 >
@spinn-3:4,0,0 > sp 3 0 0
@spinn-3:3,0,0 > as 404000 1fff
@spinn-3:3,0,0 >
@spinn-3:3,0,0 > sp 2 0 0
@spinn-3:2,0,0 > as 404000 1fff
@spinn-3:2,0,0 >
@spinn-3:2,0,0 > sp 1 0 0
@spinn-3:1,0,0 > as 404000 1fff
@spinn-3:1,0,0 >
@spinn-3:1,0,0 > sp 0 0 0
@spinn-3:0,0,0 > as 404000 3fff
@spinn-3:0,0,0 >
liuq@frodo:~/apt/old_spinnaker_
```

```
    for i in range(0, len(inhi_list)):
        p.Projection(pool_integrate, recognition_pops[iTempla
...
```

```
ello, World!
pp_data.synaptic_row_length: 0, row size: 12 bytes
pplication booted!
tarting simulation...
imulation complete.
pp_data.synaptic_row_length: 0, row size: 12 bytes
pplication booted!
tarting simulation...
```

(c) Snapshot of the neuron responses of the template matching layer [85]

Figure 4.1: Snapshots of the real-time dynamic posture recognition system on SpiNNaker.

15,216,512 synapses, use up to 19 chips (290 cores) on a 48-node board, see Table 3.1a. Regarding the lower resolution of 32×32 retinal input, the network (Table 3.1b) consists of 5,925 neurons and 318,420 synapses taking up only two chips (31 cores) of the board.

Figure 4.1 shows snapshots of neural responses of some populations during real-time recognition. Figure 4.1a is a snapshot of the Gabor population which prefers the horizontal direction, given the input posture of a 'Fist'; and Figure 4.1b shows the activity of the neurons in the integration layer, given a 'Victory Sign'. And the active neurons in the visualiser in Figure 4.1c are pointing out the position of the recognised pattern the 'Index finger'. All the supporting demonstrative videos can be found on YouTube [83, 84, 85].

4.3 Recognition of Recorded Data

To compare with the results of the experiments carried out with Matlab (in Section 3.3), the same recorded retinal data is conducted into SpiNNaker. Only Model 1 is tested on the neuromorphic hardware platform, since tracking is still need to investigate using SNN (for Model 2) in the future. The recorded data is presented as spike source array in the system with 128×128 input (see Figure 4.3a) while the data is forwarded to a sub-sampling layer of 32×32 resolution in the system of the smaller network (see Figure 4.4a). The output spikes generated from the recognition populations with time are shown in Figures 4.3 and 4.4 for full resolution and lower systems respectively. More spikes are generated during the period when the preferred input posture is shown.

Correspondingly, the spiking rates of each recognition population is sampled into frames (Figure 4.2) to make a comparison with the Matlab simulation. Each colour represents one recognition population, and the spike activity goes higher when the input posture matches the template. Firstly, the spike rates are sampled into 30 ms frames which is in accordance with the Matlab experiments. In the Matlab simulation, the templates are trained with cut frames and so the test images are also fixed to the same length frames. Otherwise, the recogniser will not work properly because of the replications of the moving posture. Contrasting this, the spiking rates can be sampled to various frame lengths. Thus, the other two plots in the figure illustrate the classification in a wider window of 300 ms. From Table 4.1, the recognition and rejection rates are quantified as percentages.

Comparing with the results of Matlab simulation (Table 3.2), the recognition rate

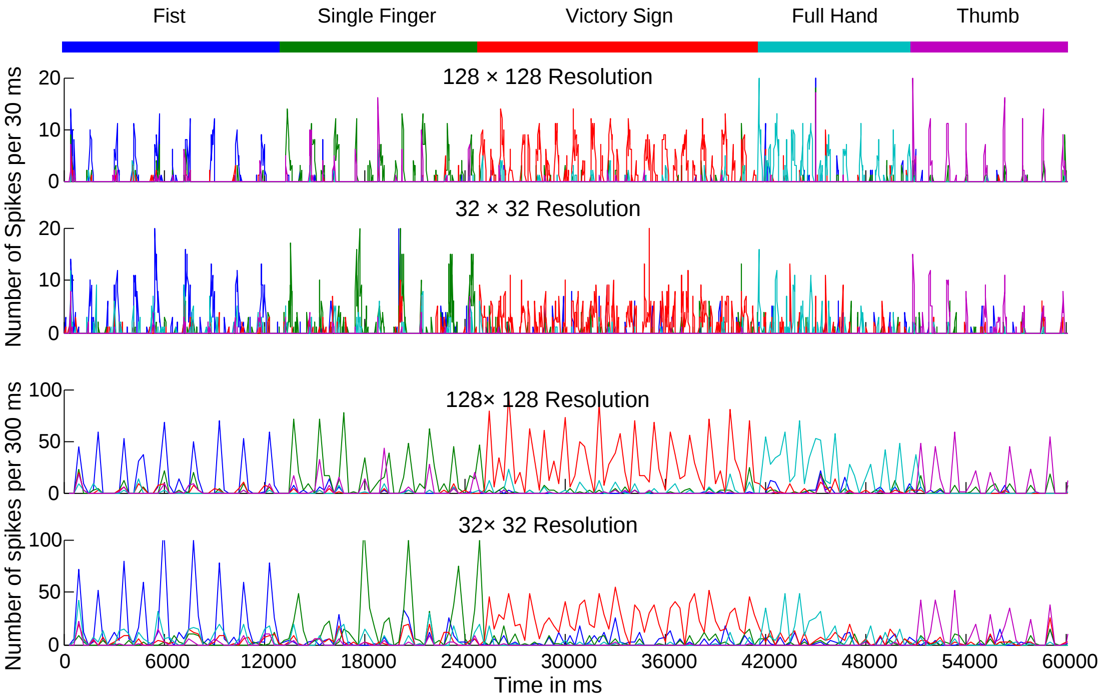
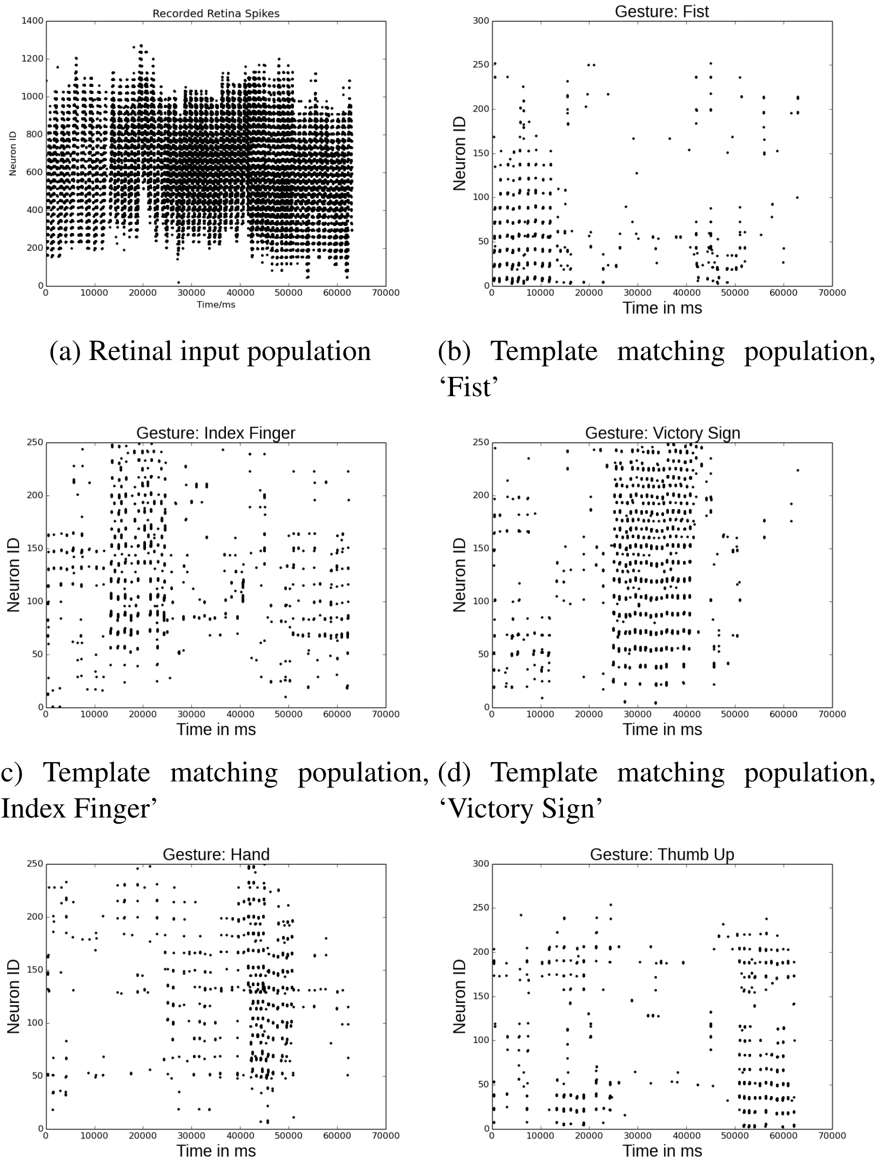


Figure 4.2: Real-time neural responses of two experiments on SpiNNaker with time to the same recorded postures. These two experiments only differ in input resolution. The result of the high input resolution test is plotted the first with a sample frame of 30 ms; while the 3rd plot shows the same result with a sample frame of 300 ms. The other two plots refer to the smaller input resolution. Every point represents the over all number of spikes of a specific population (different colour) in a 'frame'.

is about 7.6% lower at both high and low resolutions, and the rejection rate remains the same slightly above 75%. However, by changing the frame length to 300 ms recognition rates reach (93.0% for the larger network) or exceed (86.4% for smaller network) the Matlab simulation, meanwhile the rejection rates also drop dramatically by 26.0% and 22.4%. This is in accordance with natural visual responses, which means, the longer an object shows, the more accurate the recognition will be. Between the two network scales there is also a smaller gap in recognition rates as the window length grows, i.e. 8.9% and 6.6% respectively.



(a) Retinal input population

(b) Template matching population, 'Fist'

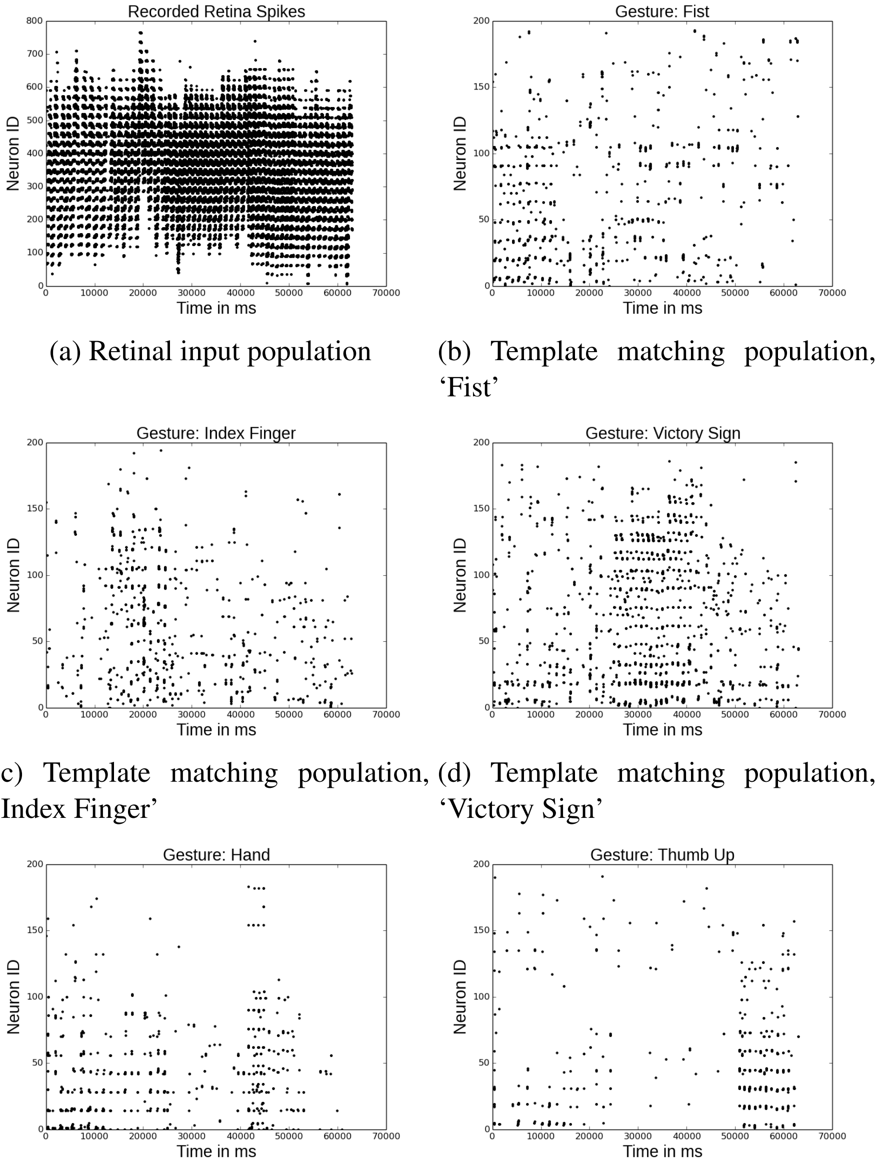
(c) Template matching population, 'Index Finger'

(d) Template matching population, 'Victory Sign'

(e) Template matching population, 'Full Hand'

(f) Template matching population, 'Thumb Up'

Figure 4.3: Spikes captured during the live recognition of the recorded retinal input with the resolution of 128×128.



(a) Retinal input population

(b) Template matching population, 'Fist'

(c) Template matching population, 'Index Finger'

(d) Template matching population, 'Victory Sign'

(e) Template matching population, 'Full Hand'

(f) Template matching population, 'Thumb Up'

Figure 4.4: Spikes captured during the live recognition of the recorded retinal input with the resolution of 32×32.

Table 4.1: Real-time recognition results on SpiNNaker in %

		30 ms per frame		**300 ms per frame**	
		High Resolution	Low Resolution	High Resolution	Low Resolution
Fist	Correct	91.78	78.02	100	92.31
	Reject	82.78	78.54	70.73	68.29
Index Finger	Correct	78.25	78.25	88.24	72.22
	Reject	80.46	73.56	57.50	55.00
Victory Sign	Correct	96.48	86.27	95.00	92.50
	Reject	64.46	72.68	28.57	28.57
Full Hand	Correct	85.29	60.78	90.00	75.00
	Reject	67.31	83.65	35.48	61.29
Thumb up	Correct	84.09	88.10	91.67	100
	Reject	87.54	73.81	66.67	66.67
Average	Correct	87.18	78.28	92.98	86.41
	Reject	76.51	76.45	51.79	55.96

Chapter 5

Contributions and Research Plan

5.1 Contributions

To explore how brain may recognise objects in its general, accurate, invariant and energy-efficient manner, this work proposes the use of a neuromorphic hardware system which includes a DVS retina connected to SpiNNaker, a real-time SNN simulator. Building a hand gesture recognition system based on this bespoke hardware for dynamic hand postures is a first step in the study of the ventral visual pathway in the brain. Inspired by the structures of the primary visual cortex, convolutional neural networks are modelled using both linear perceptrons and LIF neurons as V1-liked neurons. This model is position invariant to recognise moving postures.

The detailed contributions are listed below:

- modelled a convolutional neural network to recognise moving postures (position invariant) with V1-liked neurons.

- translated the conventional artificial neural networks of perceptrons to spiking neural networks of LIF neurons by Siegert function.

- configured the neuromorphic platform to communicate with the retina to perform real-time posture recognition using spiking neurons.

- maintain the software to make SpiNNaker receive correct recorded spikes to compare the performance with perceptrons and visualise the live spikes to probe the neural activities in real time.

5.2 Publications

- Q. Liu and S. Furber, ”Real-time recognition of dynamic hand posture on a neuromorphic system.” Artificial Neural Networks ICANN 2015. Springer Berlin Heidelberg. (Under review)

- Q. Liu, X. Lagorce, E. Stromatias, D. Emmanouilidou, R. Benosman, S. Furber and S. Liu, ”A large-scale, real-time sound localization on a neuromorphic platform.” Neuromorphic Engineering. (Under proof reading of co-authors)

- Q. Liu, C. Patterson, S. Furber, Z. Huang, Y. Hou, and H. Zhang, ”Modeling populations of spiking neurons for fine timing sound localization.” in Neural Networks (IJCNN), The 2013 International Joint Conference on, pp. 18, Aug 2013.

5.3 Future Work

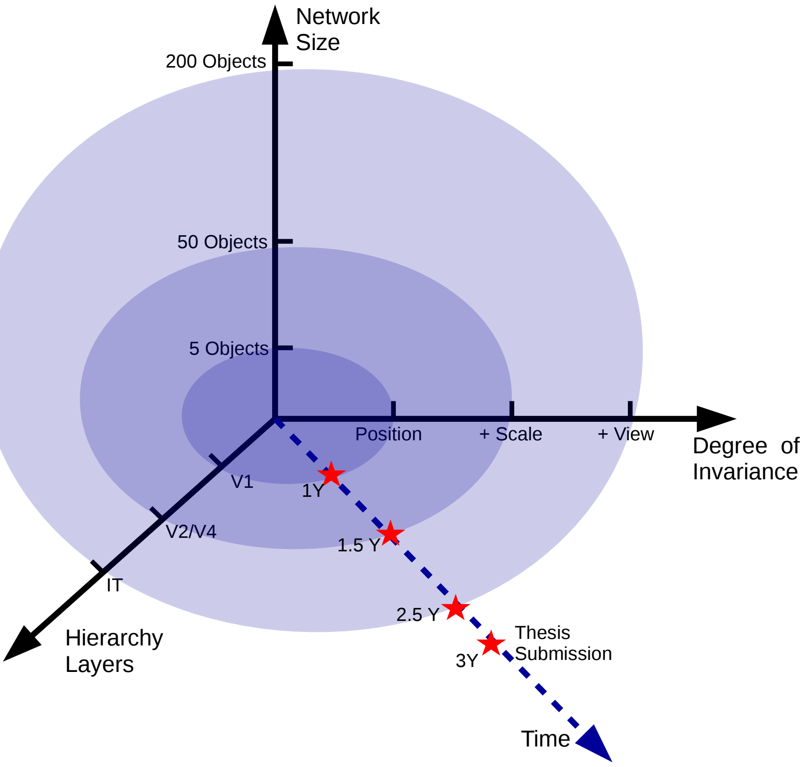


Figure 5.1: 3D representation of the research plan on the transformation-invariant object recognition system. Three milestones are highlighted with red stars indicating the expected targets of the object recognition networks.

The proposed research plan is illustrated in Figure 5.1. To build a biologically-plausible object recognition system with spiking neurons, this work will be completed with different scopes in three stages. The recognition ability of the system is measured in three dimensions: the hierarchy layers, degree of invariance and the network size.

This work will contribute to the understanding of biological visual processing by means of mimicking the neural activities in the ventral stream. More importantly, the research will apply the accurate, rapid, robust and effortless approaches to artificial systems by exploring the brain's invariant object recognition.

The performance of the real-time recognition system will be tested on each milestone to validate the success of the models. The neural activities and recognition rate will be compared with biological data.

The key research steps are listed in Figure 5.2. Since the increment work flow is hard to present in Gantt charts, only the work for the first milestone is drawn. The subsequent sections will outline the key research stages.

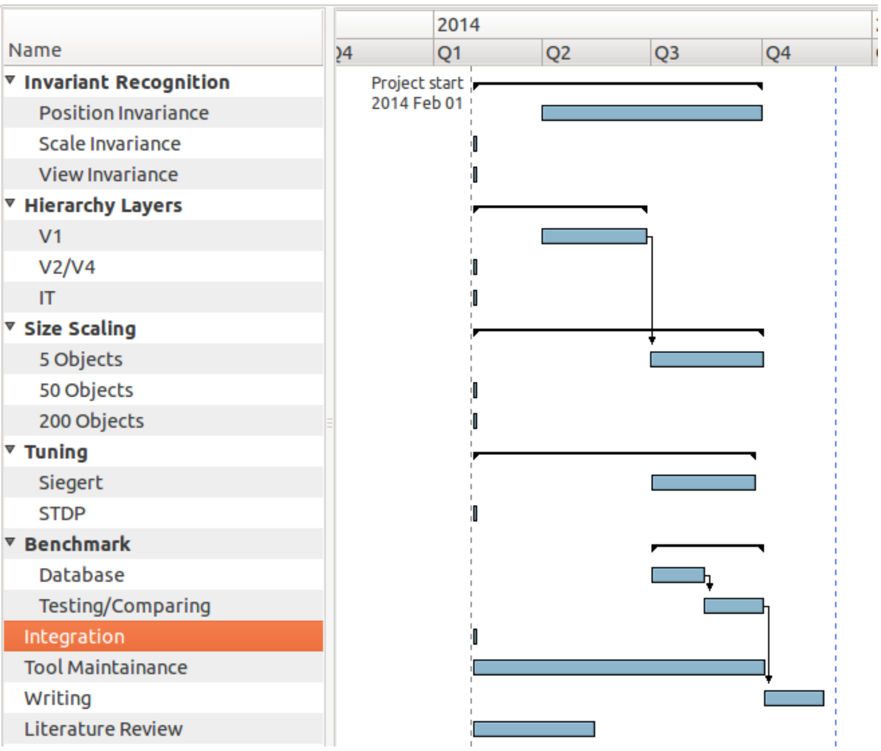


Figure 5.2: Gantt chart of the work flow for the first milestone. The main research works are listed on the left. Different from the example, in the following work, not only achieving a milestone but also any increase in any dimension will result in tuning and benchmark testing.

5.3.1 Invariant Object Recognition

As stated above the brain recognises huge amount of objects rapidly and effortlessly even in cluttered and natural scenes. While the major stumbling crux of the computer object recognition systems lies in the invariance problem. To explore the invariant object recognition of the brain in a biologically plausible way is the right place to solve the computational difficulty.

Position Invariance

Position invariance in the lower level of V1-liked neurons has been achieved in the preliminary work by convolving receptive fields with Gabor kernels. The following work in accordance with Figure 5.1 will focus on how to expand the position invariance to higher hierarchical level of the ventral stream.

Scale Invariance

Similar to orientation detection, V1 provides overcomplete population re-representations of visual image on the features of scale, frequency and orientation. It forms the basis of scale invariant object recognitions. The complexity of this work will increase as the number of hierarchy layers grows.

View Invariance

A difficult specificity-invariance trade-off occurs in the view invariant recognition tasks, since the recogniser should be able to discriminate different objects while at the same time it is also tolerant to viewing angel transformations. Learning will play a very important role in this work, where objects observed with multiple view points can be recognised even only single view point is trained.

5.3.2 Modelling the Hierarchy Layers

As the visual information conducts along the ventral stream, neurons become selective for increasingly complex features. Along with this growing complexity of the preferred stimulus, neurons become more and more tolerant to the exact position and scale of the stimulus within their receptive fields. To satisfy the functional discoveries, this work will employ learning and compare with biological data.

5.3.3 Size Scaling

The milestones set for the dimension of size scaling is in accordance with experiments data. In paper [86], the classical receptive field of the V2 cell consists of 48 grating stimuli and 80 contour stimuli; while Zoccolan et al. [43] tested the IT neurons with 213 images.

Thanks to the massive-parallel neural simulation of SpiNNaker system, to make a great real-time invariant object recognition becomes possible. However, it also requires the software development to support huge neural networks.

5.3.4 Integration

To reach the milestone of building an object recognition system with position, scale and view invariance, integration of these separate models will be a challenge. It does not only require placing the models physically together, but also merge the functions. As illustrated in 2.1.2, single neurons are tuned to different features and object identities. This work asks for investigation on population coding and learning.

5.3.5 Tuning

Tuning is the key to make the object recognition system success. In the preliminary work, Siegert transformation function is used to adjust perceptral weights to spiking LIF neurons. It is a strong backer to guarantee the feasibility of the work. However, learning algorithms such of STDP of spiking neural networks are supposed to be employed to make the system more biologically plausible. On the other hand, this work will provoke the learning algorithm study in SpiNNaker group.

5.3.6 Benchmark Performance

The performance of the real-time recognition system will be tested on each milestone to validate the success of the models. The neural activities and recognition rate will be compared with biological data.

Building Dataset

Building a well-labelled retinal output dataset is essential in spike-based object recognition study. Unified benchmarks with AER format will be ideal for SNN study, because of its non-frame, event-based fashion. These benchmarks make it possible for research groups to test their SNN model without a silicon retina present. It will boost the communication, comparison and collaboration in the community. In addition, it requires a lively discussion and cooperation with neuroscientists, where data can be derived and tested.

Testing/Comparing

The testing and comparing on the dataset will verify the reliability of the models. By comparing with the biological data, the model can be rectified and improved. The more data it compares with, the closer it could untangle the object representation.

Bibliography

[1] S. R. Lehky and A. B. Sereno, "Comparison of shape encoding in primate dorsal and ventral visual pathways," *Journal of neurophysiology*, vol. 97, no. 1, pp. 307–319, 2007.

[2] J. A. Bednar, "Topographica: building and analyzing map-level simulations from python, c/c++, matlab, nest, or neuron components," *Frontiers in neuroinformatics*, vol. 3, 2009.

[3] W. H. Bosking, J. C. Crowley, and D. Fitzpatrick, "Spatial coding of position and orientation in primary visual cortex," *Nature neuroscience*, vol. 5, no. 9, pp. 874–882, 2002.

[4] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?," *Neuron*, vol. 73, no. 3, pp. 415–434, 2012.

[5] T. Serre and T. Poggio, "A neuromorphic approach to computer vision," *Communications of the ACM*, vol. 53, no. 10, pp. 54–61, 2010.

[6] T. R. Vidyasagar, "Reading into neuronal oscillations in the visual system: implications for developmental dyslexia," *Frontiers in human neuroscience*, vol. 7, 2013.

[7] S. G. Wysoski, L. Benuskova, and N. Kasabov, "Fast and adaptive network of spiking neurons for multi-view visual pattern recognition," *Neurocomputing*, vol. 71, no. 13, pp. 2563–2575, 2008.

[8] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.

[9] Ö. Toygar and A. Acan, "Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition," *Pattern Recognition Letters*, vol. 25, no. 12, pp. 1421–1430, 2004.

[10] S.-D. Wei and S.-H. Lai, “Robust and efficient image alignment based on relative gradient matching,” *Image Processing, IEEE Transactions on*, vol. 15, no. 10, pp. 2936–2943, 2006.

[11] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[13] M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” *Nature neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[14] M. Fabre-Thorpe, G. Richard, and S. J. Thorpe, “Rapid categorization of natural images by rhesus monkeys,” *Neuroreport*, vol. 9, no. 2, pp. 303–308, 1998.

[15] C. Keysers, D.-K. Xiao, P. Földiák, and D. Perrett, “The speed of sight,” *Journal of cognitive neuroscience*, vol. 13, no. 1, pp. 90–101, 2001.

[16] J. J. DiCarlo and D. D. Cox, “Untangling invariant object recognition,” *Trends in cognitive sciences*, vol. 11, no. 8, pp. 333–341, 2007.

[17] J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, “A 3.6 s latency asynchronous frame-free event-driven dynamic-vision-sensor,” *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 6, pp. 1443–1455, 2011.

[18] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, “The SpiNNaker Project,” 2014.

[19] D. J. Felleman and D. C. Van Essen, “Distributed hierarchical processing in the primate cerebral cortex,” *Cerebral cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[20] J. Prado, S. Clavagnier, H. Otzenberger, C. Scheiber, H. Kennedy, and M.-T. Perenin, “Two cortical systems for reaching in central and peripheral vision,” *Neuron*, vol. 48, no. 5, pp. 849–858, 2005.

[21] L. G. Ungerleider and J. V. Haxby, “what and where in the human brain,” *Current Opinion in Neurobiology*, vol. 4, no. 2, pp. 157 – 165, 1994.

[22] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action," *Trends in neurosciences*, vol. 15, no. 1, pp. 20–25, 1992.

[23] P. Janssen, R. Vogels, and G. A. Orban, "Selectivity for 3d shape that reveals distinct areas within macaque inferior temporal cortex," *Science*, vol. 288, no. 5473, pp. 2054–2056, 2000.

[24] G. Von Bonin and P. Bailey, "The neocortex of macaca mulatta.(illinois monogr. med. sci., 5, no. 4.).," 1947.

[25] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, no. 3, p. 574, 1959.

[26] R. B. Tootell, M. S. Silverman, E. Switkes, and R. L. De Valois, "Deoxyglucose analysis of retinotopic organization in primate striate cortex," *Science*, vol. 218, no. 4575, pp. 902–904, 1982.

[27] X. An, H. Gong, L. Qian, X. Wang, Y. Pan, X. Zhang, Y. Yang, and W. Wang, "Distinct functional organizations for processing different motion signals in v1, v2, and v4 of macaque," *The Journal of Neuroscience*, vol. 32, no. 39, pp. 13363–13379, 2012.

[28] J. Wu, *Early Detection and Rehabilitation Technologies for Dementia.* IGI Global, 2011.

[29] J. Hegdé and D. C. Van Essen, "Selectivity for complex shapes in primate visual area v2," *J Neurosci*, vol. 20, no. 5, pp. 61–66, 2000.

[30] A. Anzai, X. Peng, and D. C. Van Essen, "Neurons in monkey visual area v2 encode combinations of orientations," *Nature neuroscience*, vol. 10, no. 10, pp. 1313–1321, 2007.

[31] Y. Daniel, M. Zarella, and G. Burkitt, "Whither the hypercolumn?," *The Journal of physiology*, vol. 587, no. 12, pp. 2791–2805, 2009.

[32] F. T. Qiu and R. Von Der Heydt, "Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules," *Neuron*, vol. 47, no. 1, pp. 155–166, 2005.

[33] S. Filipe and L. A. Alexandre, “From the human visual system to the computational models of visual attention: a survey,” *Artificial Intelligence Review*, pp. 1–47, 2013.

[34] J. Moran and R. Desimone, “Selective attention gates visual processing in the extrastriate cortex,” *Science*, vol. 229, no. 4715, pp. 782–784, 1985.

[35] T. O. Williams, *Biological Cybernetics Research Trends*. Nova Publishers, 2007.

[36] K. Tanaka, H.-a. Saito, Y. Fukada, and M. Moriya, “Coding visual images of objects in the inferotemporal cortex of the macaque monkey,” *J Neurophysiol*, vol. 66, no. 1, pp. 170–189, 1991.

[37] P. Dean, “Effects of inferotemporal lesions on the behavior of monkeys.,” *Psychological bulletin*, vol. 83, no. 1, p. 41, 1976.

[38] C. G. Gross, “Single neuron studies of inferior temporal cortex,” *Neuropsychologia*, vol. 46, no. 3, pp. 841–852, 2008.

[39] E. L. Schwartz, R. Desimone, T. D. Albright, and C. G. Gross, “Shape recognition and inferior temporal neurons,” *Proceedings of the National Academy of Sciences*, vol. 80, no. 18, pp. 5776–5778, 1983.

[40] N. Logothetis and J. Pauls, “Psychophysical and physiological evidence for viewer-centered object representations in the primate,” *Cerebral Cortex*, vol. 5, no. 3, pp. 270–288, 1995.

[41] G. Sary, R. Vogels, and G. A. Orban, “Cue-invariant shape selectivity of macaque inferior temporal neurons,” *Science*, vol. 260, no. 5110, pp. 995–997, 1993.

[42] C. J. Perry and M. Fallah, “Feature integration and object representations along the dorsal stream visual hierarchy,” *Frontiers in computational neuroscience*, vol. 8, 2014.

[43] D. Zoccolan, M. Kouh, T. Poggio, and J. J. DiCarlo, “Trade-off between object selectivity and tolerance in monkey inferotemporal cortex,” *The Journal of Neuroscience*, vol. 27, no. 45, pp. 12292–12307, 2007.

[44] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, “Stimulus-selective properties of inferior temporal neurons in the macaque,” *The Journal of Neuroscience*, vol. 4, no. 8, pp. 2051–2062, 1984.

[45] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, *et al.*, "Sequence analysis of the genome of the unicellular cyanobacterium synechocystis sp. strain pcc6803. ii. sequence determination of the entire genome and assignment of potential protein-coding regions," *DNA research*, vol. 3, no. 3, pp. 109–136, 1996.

[46] N. K. Logothetis and D. L. Sheinberg, "Visual object recognition," *Annual review of neuroscience*, vol. 19, no. 1, pp. 577–621, 1996.

[47] R. Vogels and I. Biederman, "Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex," *Cerebral Cortex*, vol. 12, no. 7, pp. 756–766, 2002.

[48] D. Zoccolan, D. D. Cox, and J. J. DiCarlo, "Multiple object response normalization in monkey inferotemporal cortex," *The Journal of Neuroscience*, vol. 25, no. 36, pp. 8150–8164, 2005.

[49] W. De Baene, E. Premereur, and R. Vogels, "Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation," *Journal of Neurophysiology*, vol. 97, no. 4, pp. 2900–2916, 2007.

[50] G. B. Ermentrout, R. F. Galán, and N. N. Urban, "Reliability, synchrony and noise," *Trends in neurosciences*, vol. 31, no. 8, pp. 428–434, 2008.

[51] C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex," *Science*, vol. 310, no. 5749, pp. 863–866, 2005.

[52] N. Majaj, H. Najib, E. Solomon, and J. DiCarlo, "A unified neuronal population code fully explains human object recognition," *Computational and Systems Neuroscience (COSYNE)*, 2012.

[53] S. L. Brincat and C. E. Connor, "Dynamic shape synthesis in posterior inferotemporal cortex," *Neuron*, vol. 49, no. 1, pp. 17–24, 2006.

[54] L. G. Nowak and J. Bullier, "The timing of information transfer in the visual system," in *Extrastriate cortex in primates*, pp. 205–241, Springer, 1997.

[55] C. F. Stevens, "An evolutionary scaling law for the primate visual system and its basis in cortical function," *Nature*, vol. 411, no. 6834, pp. 193–195, 2001.

[56] M. Lewicki and T. Sejnowski, "Learning overcomplete representations," *Neural computation*, vol. 12, no. 2, pp. 337–365, 2000.

[57] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[58] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of physiology*, vol. 117, no. 4, p. 500, 1952.

[59] J.-P. Pfister and W. Gerstner, "Triplets of spikes in a model of spike timing-dependent plasticity," *The Journal of neuroscience*, vol. 26, no. 38, pp. 9673–9682, 2006.

[60] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *The Journal of Neuroscience*, vol. 2, no. 1, pp. 32–48, 1982.

[61] S. A. Bamford, A. F. Murray, and D. J. Willshaw, "Synaptic rewiring for topographic mapping and receptive field development," *Neural Networks*, vol. 23, no. 4, pp. 517–527, 2010.

[62] A. Gupta and L. N. Long, "Character recognition using spiking neural networks," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, pp. 53–58, IEEE, 2007.

[63] J. H. Lee, P. Park, C.-W. Shin, H. Ryu, B. C. Kang, and T. Delbruck, "Touchless hand gesture UI with instantaneous responses," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pp. 1957–1960, Sept 2012.

[64] L. Camunas-Mesa, C. Zamarreno-Ramos, A. Linares-Barranco, A. J. Acosta-Jimenez, T. Serrano-Gotarredona, and B. Linares-Barranco, "An event-driven multi-kernel convolution processor module for event-driven vision sensors," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 2, pp. 504–517, 2012.

[65] M. Rehn and F. T. Sommer, "A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields," *Journal of computational neuroscience*, vol. 22, no. 2, pp. 135–146, 2007.

[66] A. Delorme, L. Perrinet, and S. J. Thorpe, “Networks of integrate-and-fire neurons using rank order coding b: spike timing dependent plasticity and emergence of orientation selectivity,” *Neurocomputing*, vol. 38, pp. 539–545, 2001.

[67] C. Eliasmith and T. C. Stewart, “Nengo and the neural engineering framework: connecting cognitive theory to neuroscience,” in *Proceedings of the 33rd annual meeting of the cognitive science society*, pp. 1–2, 2011.

[68] C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, Y. Tang, and D. Rasmussen, “A large-scale model of the functioning brain,” *science*, vol. 338, no. 6111, pp. 1202–1205, 2012.

[69] M. Naylor, P. J. Fox, A. T. Markettos, and S. W. Moore, “Managing the fpga memory wall: Custom computing or vector processing?,” in *Field Programmable Logic and Applications (FPL), 2013 23rd International Conference on*, pp. 1–6, IEEE, 2013.

[70] P. O’Connor, D. Neil, S.-C. Liu, T. Delbruck, and M. Pfeiffer, “Real-time classification and sensor fusion with a spiking deep belief network,” *Frontiers in neuroscience*, vol. 7, 2013.

[71] T. Delbruck, “Frame-free dynamic digital vision,” in *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, pp. 21–26, 2008.

[72] C. Patterson, F. Galluppi, A. Rast, and S. Furber, “Visualising large-scale neural network models in real-time,” in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pp. 1–8, 2012.

[73] F. Galluppi, K. Brohan, S. Davidson, T. Serrano-Gotarredona, J.-A. P. Carrasco, B. Linares-Barranco, and S. Furber, “A real-time, event-driven neuromorphic system for goal-directed attentional selection,” in *Neural Information Processing*, pp. 226–233, Springer, 2012.

[74] J. Lazzaro and J. Wawrzynek, “A multi-sender asynchronous extension to the aer protocol,” in *Advanced Research in VLSI, Conference on*, pp. 158–158, IEEE Computer Society, 1995.

[75] L. A. Plana, “AppNote 8 - Interfacing AER devices to SpiNNaker using an FPGA.” https://spinnaker.cs.man.ac.uk/tiki-download_wiki_attachment.php?attId=20, 4 2013.

[76] A. P. Davison, D. Brüderle, J. Eppler, J. Kremkow, E. Muller, D. Pecevski, L. Perrinet, and P. Yger, “Pynn: a common interface for neuronal network simulators,” *Frontiers in neuroinformatics*, vol. 2, 2008.

[77] S.-C. Liu, A. van Schaik, B. Minch, and T. Delbruck, “Event-based 64-channel binaural silicon cochlea with q enhancement mechanisms,” in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pp. 2027–2030, May 2010.

[78] Q. Liu, C. Patterson, S. Furber, Z. Huang, Y. Hou, and H. Zhang, “Modeling populations of spiking neurons for fine timing sound localization,” in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pp. 1–8, Aug 2013.

[79] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[80] G. La Camera, M. Giugliano, W. Senn, and S. Fusi, “The response of cortical neurons to in vivo-like input current: theory and experiment,” *Biological cybernetics*, vol. 99, no. 4-5, pp. 279–301, 2008.

[81] A. N. Burkitt, “A review of the integrate-and-fire neuron model: I. homogeneous synaptic input,” *Biological cybernetics*, vol. 95, no. 1, pp. 1–19, 2006.

[82] A. J. Siegert, “On the first passage time probability problem,” *Physical Review*, vol. 81, no. 4, p. 617, 1951.

[83] Q. Liu, “A gabor filter prefers the horizontal lines running on SpiNNaker in real-time .” https://www.youtube.com/watch?v=PvJy6RKAJhw&feature=youtu.be&list=PLxZ1W-Upr3eoQuLxq87qpUL-CwSphtEBJ, Sept. 2014.

[84] Q. Liu, “Feature extraction of live retinal input.” http://youtu.be/FZJshPCJ1pg?list=PLxZ1W-Upr3eoQuLxq87qpUL-CwSphtEBJ, Sept. 2014.

[85] Q. Liu, "Live dynamic posture recognition on SpiNNaker." http://youtu.be/yxN90aGGKvg?list=PLxZ1W-Upr3eoQuLxq87qpUL-CwSphtEBJ, Sept. 2014.

[86] J. Hegdé and D. C. Van Essen, "Temporal dynamics of shape analysis in macaque visual area v2," *Journal of neurophysiology*, vol. 92, no. 5, pp. 3030–3042, 2004.