# Real-Time Recognition of Dynamic Hand Postures on a Neuromorphic System

Qian Liu

*Abstract*—The major stumbling problem of the computer object recognition lies in the poor robustness to the transformations. Exploring and mimicking invariant object recognition within the brain is a promising approach to tackling the computational difficulty; in turn it also contributes to the understanding biological visual processing by means of mimicking neural activity in the brain.

Thanks to its high-performance massively parallel processing, SpiNNaker makes it possible to simulate large-scale neural networks in real-time. As a first milestone of this study, a recognition system for dynamic hand postures is developed on a neuromorphic hardware platform. Future work is proposed to build an object recognition system with position, scale and view invariance by modelling the hierarchical visual pathway up to the inferotemporal (IT) cortex. The proposed plan of the system will be able to recognise 200 objects in real time exploiting Integrate-and-Fire(LIF) neurons.

## I. Introduction

Patterns or objects in two-dimensional images can be described with four properties [1]: 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 [2], divide-and-conquer [3], gradient matching [4] and feature based methods [5], [6] are used. Finding an appropriate feature for a specific object still remains an open question and there is no process as general, accurate, or energy-efficient as that demonstrated by the brain. It is not a new idea to turn to nature for inspiration. Riesenhuber et al. [7], 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.

### A. What Is Object Recognition?

Object recognition is the process of assigning labels to particular objects, ranging from precise labels (‘identification’) to coarse labels (‘categorisation’) [8]. This includes the ability to accomplish these tasks under various identity preserving transformations such as object position, scale, viewing angle, background clutter and etc.

The brain can accurately recognise and categorise objects remarkably quickly, for example object recognition time in monkeys is under 200 ms [9] and the images are presented sequentially in spikes less than 100 ms [10]. This research focuses on this rapid and highly accurate object recognition, ‘core recognition’, which is defined in [11].

### B. Why Is It Important?

The human brain recognises huge amount of objects rapidly with ease even in cluttered and natural scenes. This robust object recognition of the biological system is invariant to the change of position, scale, viewing angle and etc. (known as transformation invariance). While the major stumbling problem of the computer object recognition lies in the poor robustness to the transformations. Each encounter of an object on the retina is unique because of differing illumination (lighting conditions), position (projection locations on the retina), scale (distances and sizes), pose (viewing angles), and clutter (visual contexts). 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 remaining tolerant to image transformations.

Exploring and mimicking invariant object recognition within the brain is a promising approach to tackling the computational difficulty; in turn it also contributes to the understanding biological visual processing by means of mimicking neural activity in the visual system of the brain. Moreover, energy-efficiency improvements following from the great energy efficiency of biological systems will help in building object recognition systems, e.g. posture recognition for human-machine interfaces in mobile devices.

### C. How to Mimic The Brain?

To explore how brain may recognise objects, we have employed a biologically-inspired DVS silicon retina [12]. This is a good example of low-cost visual processing due to its event-driven and redundancy-reducing style of computation; and a SpiNNaker system [13], which is a massive parallel computing platform aimed at real-time simulation of SNNs. Thanks to its high-performance processing of large-scale neural networks, we explore biological approaches of visual processing by mimicking the functions of different layers along the ventral 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 performance 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. This preliminary work achieved the first milestone of the research which aims at building a position-invariant object recognition system exploiting V1-like neurons (primary visual cortex: area V1) to classify five hand postures.

## II. Biological Aspects

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 [14]. There are two basic streams locating in the visual area: a dorsal and a ventral pathway .This research mainly focuses on the ventral visual pathway, since it dominates the object recognition among the cortical areas.

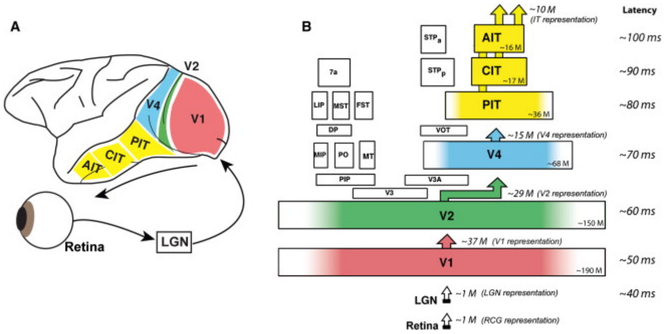


Fig. 1: The ventral visual pathway and abstraction layers [8].

### A. The ventral visual pathway

The ventral visual pathway (Figure 1) starts from the primary visual cortex V1 in the occipital cortex through areas such as V2 and V4 to the Inferotemporal (IT) cortex.

**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 [15] in 1958. In the spatial domain, V1 neurons are tuned to Gabor-like transforms applied to their small local receptive field. In theory, these Gabor-like filters together can carry out neuronal processing of spatial frequency, orientation, motion, direction, speed, and many other spatio-temporal features.

**Visual Areas V2/V4.** The responses of many V2 neurons are also modulated for complex properties: orientation of illusory contours [16], binocular disparity [17], and whether the stimulus is part of the figure or the ground [18].

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.

**Inferotemporal Cortex: IT.** The complexity increases along the ventral stream towards anterior IT (AIT) where objects are represented and recognised [19]. The high-order complex features includes the combinations of colour or texture with complicated shapes [20], and body parts such as faces and hands [21]. The distinguishing features of the IT cortex is that the neuronal responses are position and size invariant [22], and also invariant to changes in luminance, texture, and relative motion  [23].

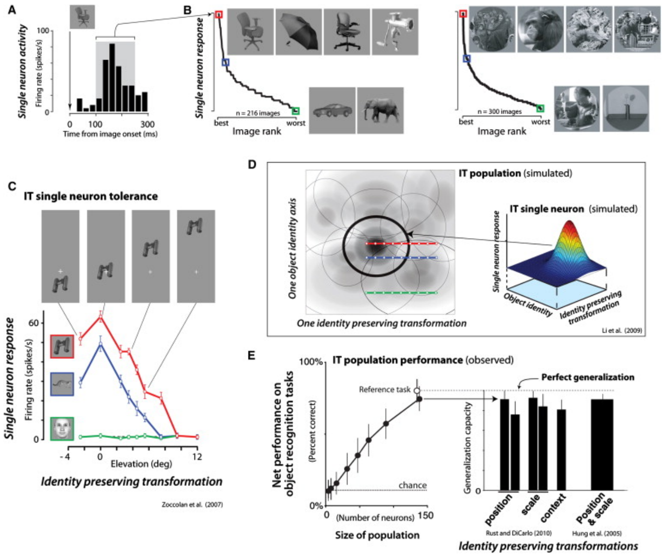


Fig. 2: IT single-neuron properties and their relationship to population performance [8].

### B. 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.

**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(A) [24], 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 [25], see example in Figure 2(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 indicated the highest respond while the green the lowest and the blue the medium. Another example in Figure 2(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.

Although IT neurons are commonly described as narrowly selective object identifier, neurophysiological studies have shown a diverse selectivity of single neurons [25]. As illustrated in Figure 2(D), a single neuron (right) is modulated to both object identities and variables of identity-preserving transformations.

**Population of neurons.** 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 [26]; and the simple weighted summation explains a wide range of invariant object recognition behaviour sufficiently [27].

Figure 2(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 [26]. 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 [28] (see Figure 2(A), an appropriate decoding window can be 100-150 ms after image onset).

### *C. Hierarchical Feed-forward Organisation*

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 [8].

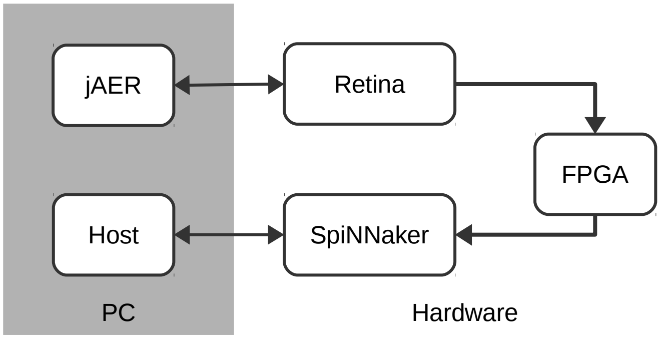
The corresponding hierarchical organisation is showed in Figure 1(B). Each area is plotted with the size proportional to its cortical surface size. Approximate total number of neurons of both hemispheres is shown in the corner of the cortical areas. The approximate number of projections is written above 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.

## III. Preliminary Work

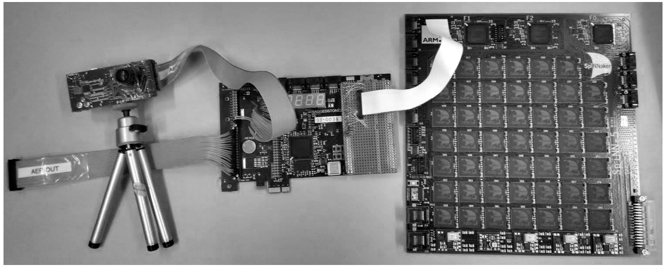
To explore how the brain may recognise objects in its general,accurate and energy-efficient manner, this work employs a neuromorphic hardware system formed from a Dynamic Video Sensor (DVS) silicon retina in concert with the SpiNNaker real-time Spiking Neural Network (SNN) simulator. Inspired by the behaviours of the primary visual cortex, Convolutional Neural Networks (CNNs) are modelled using both linear perceptrons and LIF neurons.

### *A. Platform*

The outline of the platform is illustrated in Figure 3a, where the hardware system is configured, controlled and monitored by the PC. The jAER [29] 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 offline. It visualises [30] the spiking activity of the network in real-time. The photograph of the hardware platform, Figure 3b, shows that the silicon retina connects to the SpiNNaker 48-node system via a Spartan-6 FPGA board [31].



(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.

Fig. 3: System overview of the object recognition platform.

**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 [32]) 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 [12].

**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.

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.

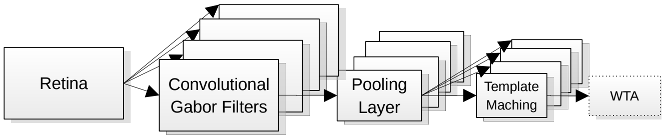


Fig. 4: Model 1. The retina input is convolved with Gabor filters in the second layer, and then shrinks the sizes in the pooling layer. The templates are considered as convolution kernels in the last layer. The WTA circuit can be used as an option to show the template matching result more clearly.

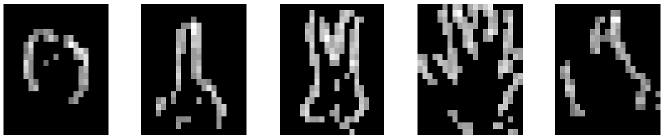


Fig. 5: Templates of the five postures: 'Fist','Index Finger', 'Victory Sign', 'Full Hand' and 'Thumb up'.

### B. CNN Nodels

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.

**Model 1:** Template Matching. Shown in Figure 4 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 5.

**Model 2:** Trained MLP. Inspired by the research of Lecun [33], we designed a combined network model with MLP and CNN (Figure 6). The first three layers are 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.

### C. Moving from Perceptrons to Spiking Neurons

It remains a challenge to transform traditional artificial neural networks into spiking ones. There are attempts [34] [35] to estimate the output firing rate of the LIF neurons (Equation 1) under certain conditions.

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\frac{V(t) - V_{rest}}{\tau_m} + \frac{I(t)}{C_m} \tag{1}$$

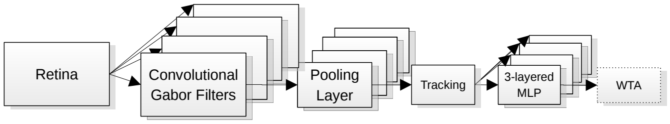


Fig. 6: Model 2. The retina input convolves with Gabor filters in the second layer, and then shrinks the sizes in the pooling layer. The following tracking layer finds the most active area of some fixed size, moves the posture to the centre and pushes the image to the trained MLP. The winner-take-all (WTA) layer can be used as an option to show the template matching result more clearly.

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

$$\lambda_{out} = \left[ t_{ref} - \tau_m \ln \left( 1 - \frac{V_{th} - V_{rest}}{I R_m} \right) \right]^{-1} \tag{2}$$

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 1 can be modelled as a stochastic differential equation referring to Ornstein-Uhlenbeck process,

$$\tau_m \frac{\mathrm{d}V(t)}{\mathrm{d}t} = -[V(t) - V_{rest}] + \mu + \sigma \sqrt{2\tau_m} \xi(t) \tag{3}$$

where

$$\begin{aligned} \mu &= \tau_m (\mathbf{w_E} \cdot \lambda_\mathbf{E} - \mathbf{w_I} \cdot \lambda_\mathbf{I}) \\ \sigma^2 &= \tfrac{\tau_m}{2} \left( \mathbf{w_E^2} \cdot \lambda_\mathbf{E} + \mathbf{w_I^2} \cdot \lambda_\mathbf{I} \right) \end{aligned} \tag{4}$$

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 $\lambda$ 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 [36],

$$\begin{aligned} \lambda_{out} = \Bigg( & \tau_{ref} + \frac{\tau_Q}{\sigma_Q} \sqrt{\frac{\pi}{2}} \int_{V_{rest}}^{V_{th}} \mathrm{d}u \exp \left( \frac{u - \mu_Q}{\sqrt{2}\sigma_Q} \right)^2 \\ & \cdot \left[ 1 + \mathrm{erf} \left( \frac{u - \mu_Q}{\sqrt{2}\sigma_Q} \right) \right] \Bigg)^{-1} \end{aligned} \tag{5}$$

where $\tau_Q, \mu_Q, \sigma_Q$ are identical to $\tau_m, \mu, \sigma$ in Equation 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.

### *D. Experiments*

**Experiment 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.

Model 1 is tested on both percetrons in Matlab and LIF neurons on SpiNNaker; whereas, Model 2 is only validated on Matlab since it merely estimates the highest performance of the system but in a non-biological way. All the details can be found in the latest submitted paper.

**Recognition using LIF neurons.** The output spikes generated from the recognition populations with time are shown in Figures 8 for full resolution system. More spikes are generated during the period when the preferred input posture is shown.

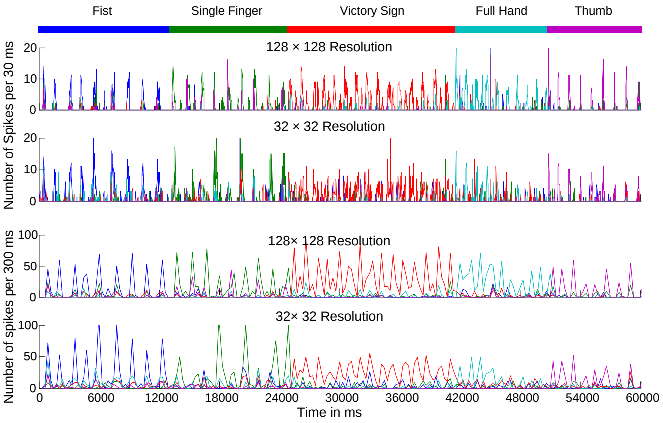
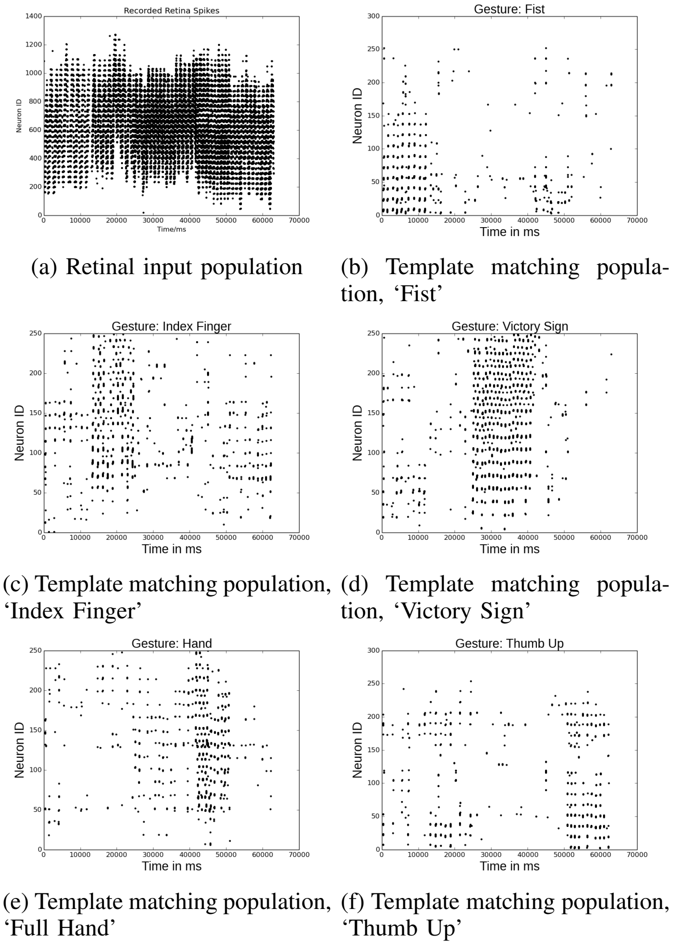


Fig. 7: Real-time neural responses of two experiments on SpiNNaker with time to recorded postures. These two experiments only differ in input resolution. Every point represents the over all number of spikes of a specific population (different colour) in a ‘frame’. First two plots are for a sample frame of 30 ms; the latter are for a frame of 300 ms.

In this study’s largest configuration using these approaches, a network of 74,210 neurons and 15,216,512 synapses is created and operated in real-time using 290 SpiNNaker processor cores in parallel and with 93.0% accuracy. A smaller network using only 1/10th of the resources is also created, again operating in real-time, and it is able to recognise the postures with an accuracy of around 86.4% - only 6.6% lower than the much larger system. The recognition rate of the smaller network developed on this neuromorphic system is sufficient for a successful hand posture recognition system, and demonstrates a much improved cost to performance trade-off in its approach.



(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’

Fig. 8: Spikes captured during the live recognition of the recorded retinal input with the resolution of 128×128.

## IV. Future Work

The proposed research plan is illustrated in Figure 9, where the scope of the research is estimated in three dimensions. To build a biologically-plausible object recognition system using spiking neurons, this work will be completed in three stages:

1) Year 1, building a position-invariant object recognition system exploiting V1-like neurons to classify five hand postures.
2) Year 1.5, combining scale- with position-invariance on the object recognition system, and building the hierarchy ventral pathway to the V2/V4 layer to recognise 50 simple combined features such as gratings and contours.
3) Year 2.5. integrating position-, scale- and view-invariance by modelling the hierarchical visual pathway up to the IT cortex and equipping the system with the ability to recognise 200 objects in real time.

This work will contribute to the understanding of biological visual processing by means of mimicking the neural activities

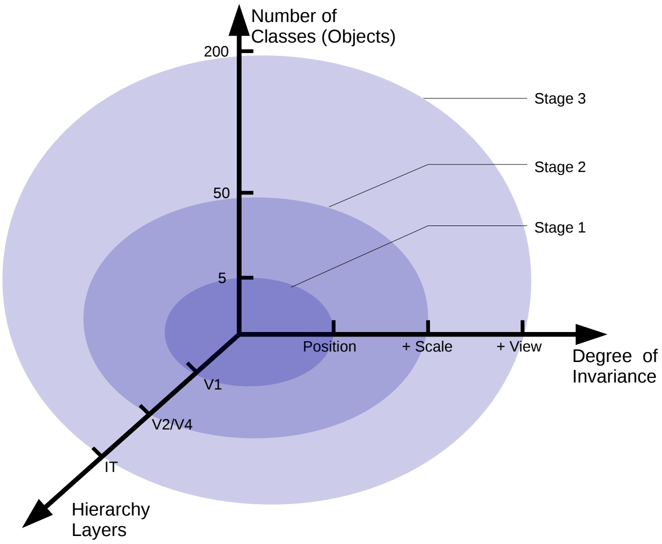


Fig. 9: 3D representation of the research plan on the transformation-invariant object recognition system. Three milestones are pointed out indicating the expected targets of the object recognition networks.

in the ventral stream. More importantly, the research will apply the accurate, rapid and robust 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 also be compared with biological data.

The key research steps are listed in Figure 10. Since the incremental work flow is hard to present in Gantt charts, only the work for the first milestone is shown. The subsequent sections will outline the key research stages.

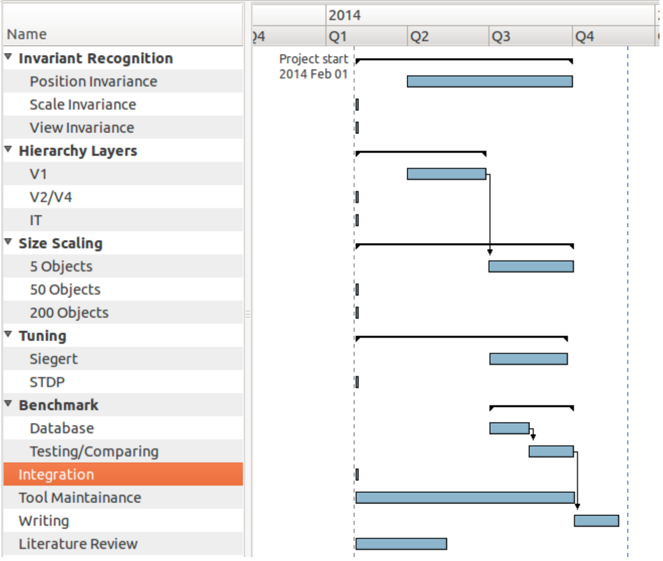


Fig. 10: Gantt chart of the work flow for the first milestone. The main research works are listed on the left.

### A. *Invariant Object Recognition*

As stated previously the brain recognises huge number of objects rapidly with ease even in noisy natural scenes. We will explore the invariant object recognition in three features: position, scale and viewing angle.

*1) 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 9 will focus on expanding the position invariance to higher hierarchical levels of the ventral stream.

*2) 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. Likewise, integrating the features into the higher abstraction of layered network to recognise more complex figures will require a tense work on tuning.

*3) View Invariance:* A difficult specificity-invariance tradeoff occurs in view invariant recognition tasks, since the recogniser should be able to discriminate different objects while at the same time also tolerating to viewing angle transformations. Learning will play a very important role in this work, where objects observed with multiple view points can be recognised even if only single view point is presented during training.

### B. *Modelling the Ventral Visual Pathway*

As the visual information propagates through the ventral stream (via visual area V1, V2/V4 and IT), neurons become selective for increasingly complex features. Along with this growing complexity of the preferred stimulus, neurons become more and more tolerant to the position and scale of the stimulus within their receptive fields. Inspired from the functional behaviour of the biological data (many have been mentioned in Section II), this work will mimic the neural activity of each hierarchy layer by LIF neurons.

### C. *Size Scaling*

The milestones set for the dimension of number of classes/objects is in accordance with experimental data from the study of neuroscience. In work by [37], the classical receptive field of the V2 cell consists of 48 grating stimuli and 80 contour stimuli; while Zoccolan et al. [24] tested and recorded the activity of the IT neurons of monkeys with 213 grayscale pictures of isolated real-world objects.

Thanks to the massively-parallel neural simulations possible in the SpiNNaker system, implementing real-time invariant object recognition becomes possible. However, it also requires the supporting software development to support larger neural networks than currently possible.

### D. *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 not only requires placing the models physically together but also merging their functions. As illustrated in Section II, single neurons are tuned to different features and object identities. This work requires further investigation into population coding and learning.

### *E. Tuning*

Tuning is the key to make the object recognition system a success. In preliminary work, Siegert transformation functions are used to adjust perceptral weights for spiking LIF neurons. This is a strong indicator of the feasibility of the work. However, learning algorithms such as STDP in spiking neural networks are must be employed to make the system more biologically plausible. It is hoped that, this work will provoke further study of learning algorithms on SpiNNaker.

### *F. Benchmarking Performance*

The performance of the real-time recognition system will be evaluated of each milestone to validate the success of the models. The neural activities and recognition rate will be compared with biological data which will act as a benchmark.

*1) Building a 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 benchmark datasets will also make it possible for other researchers to test their SNN model without a silicon retina present.

*2) Testing/Comparing:* The testing and comparing on the dataset will verify the reliability of the models. The neural responses of single or populated neurons to the same dataset will be analysed in firing rate and response time. 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.

## References

[1] 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.

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

[3] 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.

[4] 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.

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

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, "The SpiNNaker Project," 2014.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] 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.