Biologically Inspired Intensity and Range Image Feature Extraction

D. Kerr, S. A. Coleman, T. M. McGinnity, M. Clogenson

Abstract—The recent development of low cost cameras that capture 3-dimensional images has changed the focus of computer vision research from using solely intensity images to the use of range images, or combinations of RGB, intensity and range images. The low cost and widespread availability of the hardware to capture these images has realised many possible applications in areas such as robotics, object recognition, surveillance, manipulation, navigation and interaction. Given the large volumes of data in range images, processing and extracting the relevant information from the images in real time becomes challenging. To achieve this, much research has been conducted in the area of bio-inspired feature extraction which aims to emulate the biological processes used to extract relevant features, reduce redundancy, and process images efficiently. Inspired by the behaviour of biological vision systems, an approach is presented for extracting important features from intensity and range images, using biologically inspired spiking neural networks in order to model aspects of the functional computational capabilities of the visual system.

I. INTRODUCTION

Recent years have witnessed the development of low cost consumer depth sensing technology such as the Microsoft Kinect and ASUS Wavi-Xtion. These RGB-D (Red-Green-Blue-Depth) cameras were originally designed for game console platforms but have become a popular platform for research and applications in many areas such as robotics, object recognition, surveillance, manipulation, navigation, and interaction [1, 2, 3, 4] due to the low cost and widespread availability. Depth imagery can be used to obtain reliable descriptions of 3-D scenes. The idea of computing an image of depth measurements is not a recent innovation, and researchers have been experimenting with the technology for a number of years. What has changed is the introduction of the low cost consumer technology, provided an easier gateway to this depth imaging technology for researchers and the general public.

Depth imaging technology has traditionally been known as range imaging. This description referred to the fact that the image produced contained distances (or ranges) from the imaging device to specific points in a scene. In the rest of the paper the term range imaging will be used. A range image may be described as a 2D image containing distance measurements from a selected reference point or plane to surface points of objects within a scene [5]. Range imaging provides additional information over standard intensity imaging, allowing more information about the scene to be recovered [6]. It is important to distinguish that a range image contains information only about the visible surfaces of the objects, and not their hidden surfaces, and hence range image data may be referred to as 2½-D information [5].

Artificial vision has been in development for over half a century, but the processing capabilities of biological visual systems are still vastly superior in terms of power, speed, and performance. Existing artificial vision techniques are at an advanced level of development, and able to perform complex tasks such as face detection and classification in still images, but artificial vision techniques are still simplified enormously compared with biological vision systems. To overcome the failings of conventional artificial vision techniques, researchers have started to examine and take inspiration from biological vision systems. Whilst many approaches have been proposed, most are variations of second generation neural networks [12].

Spiking neural networks (SNNs) are the third generation class of neural networks that more accurately mimic the biological information processing in the visual system, increasing computational power and speed when compared with traditional neural networks and therefore enabling real-time processing [13] which is essential for data intensive applications such as robotics applications. SNNs use simple neuronal models and communicate using spikes in a manner similar to action potentials found in biological neurons. There has been some research investigating the application of SNNs to visual processing; a spiking neural network model that performs segmentation and edge detection is proposed in [14]; in [15] a spiking neural network is proposed to detect contours in images through the synchronisation of integrate and fire neurons using simple synthetic images; in [16, 17] a spiking neural network is proposed for real-time edge detection. Additionally, spiking neural networks have been previously used as controllers in evolutionary robotics to perform vision based obstacle avoidance [18, 19], laser-based retinal model robot vision [20], spike based sensory control [21, 22], flying robot visual control [23] and sonar based control [24]. Whilst some of these techniques may be modified to enable processing of range images, none of these algorithms have been specifically developed to make use of the combined range data and intensity data and the complementary

Manuscript received February 22, 2013. This research is funded by the Centre of Excellence in Intelligent Systems project, funded by InvestNI and the Integrated Development Fund.

D. Kerr is with the Intelligent Systems Research Centre, University of Ulster, Magee, Northern Ireland, BT48 7JL, U. K. (phone: +44 28716 75330; e-mail: d.kerr@ulster.ac.uk).

S.A. Coleman is with the Intelligent Systems Research Centre, University of Ulster, Magee, Northern Ireland, BT48 7JL, U. K. (e-mail: sa.coleman@ulster.ac.uk).

T.M. McGinnity is with the Intelligent Systems Research Centre, University of Ulster, Magee, Northern Ireland, BT48 7JL, U. K. (e-mail: tm.mcginnty@ulster.ac.uk).

M. Clogenson, was with Intelligent Systems Research Centre, University of Ulster, Magee, Northern Ireland, BT48 7JL, U. K. She is now with CPE Lyon, domaine Scientifique de la Doua, BP 82077-69616, Villeurbanne, France (e-mail: marine.clogenson@cpe.fr).
information that the different imaging techniques provide.

This paper presents an approach to feature detection using biologically inspired spiking neural networks to develop an artificial vision system that reflects a stronger correlation with biological visual systems than second generation neural networks. The primary objective of this research focuses on the simultaneous processing and feature extraction from registered intensity and range images using biologically inspired techniques. The spiking neural network provides a means to combine the processing of the different image modalities, and produce as output, a single processed image that illustrates the complementary features from both intensity and range image modalities. Section II describes how RGB-D images are captured and pre-processed. Section III presents the spiking neuron model used in the experiments, the spiking neural network structure, the implementation and discusses the biological basis of the proposed technique. Section IV presents experimental results and visual comparison with conclusions presented in Section V.

II. RGB-D IMAGE CAPTURE AND STRUCTURE

RGB-D cameras capture Red-Green-Blue images with per-pixel Depth information. Many techniques have been developed to compute range (i.e. depth) measurements of a scene, but typically range measurements are computed using either active stereo or time-of-flight sensing. Depth measurement techniques used in RGB-D cameras have some shortcomings as they only provide depth measurements at a limited distance (< 5m), depth estimates can be noisy and the field of view is constrained when compared to specialised range cameras and 3D laser scanners.

A typical RGB-D camera (see Figure 1) has a number of components to capture RGB-D images. It consists of an infrared laser emitter (to produce the structured light pattern), an infrared camera and an RGB camera. Depth is computed using a triangulation process [25]. The laser emits a single beam (split into multiple beams by a diffraction grating) to create a constant pattern of speckles projected onto the scene that is captured by the infrared camera and is compared with a reference pattern. Speckle shifts are measured using a simple image correlation procedure to produce a disparity image. The distance to the sensor can then be retrieved from the corresponding disparity. Each pixel in an RGB-D image consists of four channels: red, green, blue and depth. The red, green, and blue channels are captured using the RGB camera and the depth measurements are computed using the method described previously. Thus, the 3D location of each RGB pixel in physical space is computed using speckle shifts and known sensor parameters.

Figure 2 illustrates an example frame captured with an RGB-D camera with the range and RGB image components shown separately. To represent the range image visually the depth measurements are converted to intensity values where values close to the camera are represented in darker greys and values further away from the camera are represented with lighter greys. With depth sensing technology there may be locations in the image where no depth measurement has been measured due to reflections, occlusions, or various other reasons. In the range image displayed in Figure 2(b) these locations are visualised as black areas in the image.

Fig. 1. PrimeSense RGB-D camera used to capture RGB-D images.

Fig. 2. Example images captured with RGB-D style camera ([9]).

In the experiments presented in this paper the RGB image is transformed to an intensity (greyscale) image by summing and averaging individual R, G, and B components. This results in both an intensity image and a range image providing complementary information for further processing.

III. SPIKING NETWORK DESIGN AND BIOLOGICAL BASIS

This section describes the biological basis for the proposed
artificial vision system, the spiking neuron model used in the network, and the spiking neural network design.

A. Biological basis and receptive fields

The retina is a light sensitive tissue lining within the inner surface of the eye. It is an extension of the brain and formed embryonically from neural tissue and connected to the brain by the optic nerve. The retina is the only source of visual information to the brain; visual processing begins when photons stimulate the light-sensitive photoreceptor rod and cone cells within it. These cells convert the information into electrical signals and send them through intermediate networked layers of cells to around 15-20 types of retinal ganglion cells. The electrical signals are then sent to higher visual processing areas in the brain via the optic nerve. Within the network structure proposed in this paper there is a layer that emulates the photoreceptors in the retina.

Recent research has identified that computations of functional visual characteristics such as orientation and motion detection begin in the retina, through the formation of receptive fields, and are not restricted to higher stages in the visual system as was previously thought [26]. In a biological system a receptive field is where a spiking neuron integrates the spikes from a group of afferent neurons. The receptive fields of neurons in the visual system are comprised of a 2-D region in visual space with varying size. Receptive fields are present in many areas of the visual system including the retina and visual cortex. An example receptive field is illustrated in Figure 3 where neuron N has a receptive field with a 9-neuron array. Each neuron in the receptive field connects to neuron N through both excitatory and inhibitory synapses. It should be noted however, that construction of receptive fields is not restricted to only 9-neuron arrays; arrays of any size may be used to construct receptive fields.

Some receptive fields are elongated with an excitatory central oval, and an inhibitory surrounding region. Some receptive fields are rectangular, with one long side being excitatory and the other being inhibitory. The light pattern needed to stimulate these receptive fields needs to have a particular orientation in order to excite the cell. These orientation specific receptive fields are emulated in the proposed network structure and form the basis of the experimental work presented. By emulating the processing of these orientation specific receptive fields, it is possible to create an artificial vision system capable of detecting light patterns at particular orientations in the case of intensity images, or depth discontinuities at particular orientations in the case of range images. We use four types of orientation (up, down, left and right) specific receptive fields corresponding to horizontal inhibitory, horizontal excitatory, vertical inhibitory and vertical excitatory respectively.

B. Spiking Neuron Model

A widely used spiking neuron model is that of Hodgkin and Huxley [27] based on recordings obtained from experiments on the giant squid axon using a voltage clamp method. However, even though this model is biologically plausible, the complexity in simulating the model is very high due to the number of differential equations. Thus, most computer simulations of neuron models choose to use a simplified neuron model such as the integrate-and-fire model (I&F), leaky I&F model, conductance-based I&F or Izhikevich’s model. A full review of the biological behaviour of single neurons can be found in [28] and a comparison of different neuron models can be found in [29]. The conductance-based I&F model has been selected to model the network neurons in this approach as it offers similar neuron behaviour to the Hodgkin-Huxley whilst providing a reduction in computational complexity. In the conductance-based I&F model the membrane potential is governed by the following equation:

\[
c_m \frac{dv(t)}{dt} = g_r (E_r - v(t)) + \frac{w_{ex} g_{ex}(t)}{A_{ex}} (E_{ex} - v(t)) + \frac{w_{ih} g_{ih}(t)}{A_{ih}} (E_{ih} - v(t))
\]

where \( c_m \) is the membrane capacitance, \( E_r \) is the membrane reversal potential, \( g_r \) is the conductance of the membrane, \( E_{ex} \) and \( E_{ih} \) are the reversal potential of the excitatory and inhibitory synapses respectively, \( w_{ex} \) and \( w_{ih} \) are weights for excitatory and inhibitory synapses respectively, \( A_{ex} \) and \( A_{ih} \) are the membrane surface areas connected to the excitatory and inhibitory synapses respectively. If the membrane potential \( v(t) \) exceeds the threshold voltage \( v_{th} \), an action potential is generated; then \( v(t) \) is then reset to \( v_{reset} \) for a time \( \tau_{ref} \) which is called the refractory duration. For simplicity in this work \( \tau_{ref} \) is set to 0. The variables \( g_{ex}(t) \) and \( g_{ih}(t) \) represent the conductance’s of excitatory and inhibitory synapses respectively, which vary with time. The output spike train is then represented by a series of 1’s or 0’s representing whether or not a neuron fires at time \( t \), i.e. \([S_{out}(t_1), S_{out}(t_2), \ldots, S_{out}(t_{1878})]\).
C. Spiking Neural Network Structure

The proposed spiking neural network structure is illustrated in Figure 4. Suppose that the first network layer in Figure 4 represents photoreceptors. Each pixel in the input intensity and range image corresponds to an intensity-photoreceptor and range-photoreceptor respectively. The intermediate layer is composed of eight types of neurons, four for the intensity image and four for the range image. Each of these sets of 4 neurons represents the excitatory and inhibitory synapses forming the receptive field of the neurons in horizontal and vertical directions. The intermediate layer is connected to the photoreceptor layer via weight filters representing the connectivity and synaptic weights of the receptive field under consideration. The size of the weight filters may be increased by changing parameter values, thus increasing the connectivity and ultimately the size of the receptive field. In Figure 4, within the receptive field, ‘X’ in the synapse connections represents an excitatory synapse and ‘Δ’ represents an inhibitory synapse. The synapse weights are calculated using the function provided in equation (2) [17] based on the description of biological receptive fields [30]. The excitatory and inhibitory receptive fields to detect vertical edges in the intensity image are denoted as $w_{\text{vert-ex}}^i$ and $w_{\text{vert-ih}}^i$. The excitatory and inhibitory receptive fields to detect the horizontal edges in the intensity image are denoted as $w_{\text{hori-ex}}^i$ and $w_{\text{hori-ih}}^i$ respectively. Similarly, $w_{\text{vert-ex}}^r$, $w_{\text{vert-ih}}^r$, $w_{\text{hori-ex}}^r$ and $w_{\text{hori-ih}}^r$ denote the vertical and horizontal excitatory and inhibitory receptive fields in the range image.
The values for the excitatory weights in the receptive field \( w_{\text{hori-ex}} \) in the horizontal direction are computed using the function,

\[
w_{\text{hori-ex}}(x, y) = \begin{cases} 
0 & \text{if } (y - y_s) \leq 0 \\
\frac{(y - y_s)^2 - (y - y_i)^2}{\delta_i^2 - \delta_s^2} & \text{if } (y - y_s) > 0 \\
w_{\text{max}} - e^{-\frac{(y - y_s)^2 - (y - y_i)^2}{\delta_i^2 - \delta_s^2}} & \text{if } (y - y_s) < 0
\end{cases}
\]

(2)

where \((x_s, y_s)\) is the centre of the receptive field, \(\delta_i\) and \(\delta_s\) are constants that determine the shape of the receptive field, and \(w_{\text{max}}\) is the maximal weight of the excitatory and inhibitory synapses. In the experiments presented in this paper \(\delta_i = 1\), \(\delta_s = 6\) and \(w_{\text{max}} = 1\) ensuring consistency with the work presented in [17].

Similarly, the values for the inhibitory weights in the receptive field \(w_{\text{hori-ih}}\) are computed using the function,

\[
w_{\text{hori-ih}}(x, y) = \begin{cases} 
0 & \text{if } (y - y_s) > 0 \\
\frac{(y - y_i)^2 - (y - y_s)^2}{\delta_i^2 - \delta_s^2} & \text{if } (y - y_i) \leq 0 \\
w_{\text{max}} - e^{-\frac{(y - y_i)^2 - (y - y_s)^2}{\delta_i^2 - \delta_s^2}} & \text{if } (y - y_i) > 0
\end{cases}
\]

(4)

When constructing a \(3\times3\) \(w_{\text{hori-ex}}\) receptive field using these parameters the following weight filter is produced,

\[
w_{\text{hori-ex}} = \begin{bmatrix} 
0 & 0 & 0 \\
0 & 0 & 0 \\
0.97 & 1.0 & 0.97
\end{bmatrix}.
\]

(3)

Similarly, the values for the inhibitory weights in the receptive field \(w_{\text{hori-ih}}\) are computed using the function,

\[
w_{\text{hori-ih}} = \begin{bmatrix} 
0 & 0 & 0 \\
0 & 0 & 0 \\
0.97 & 1.0 & 0.97
\end{bmatrix}.
\]

(5)

The \(w_{\text{vert-ex}}\) and \(w_{\text{vert-ih}}\) receptive fields can be obtained by simply rotating the \(w_{\text{hori-ex}}\) and \(w_{\text{hori-ih}}\) weight filters through 90°. The intensity weight filters \(w_{\text{vert-ex}}, w_{\text{vert-ih}}, w_{\text{hori-ex}}, w_{\text{hori-ih}}\) and \(w_{\text{vert-ih}}\) are obtained in a similar manner using equations (2) and (4) resulting in the same weight filters providing the parameters and receptive field size under consideration are identical.

Receptive field sizes vary in biology to enable ‘tuning’ to a particular size of features, a technique commonly called multi-scale processing in the image processing community. Whilst not illustrated in the work presented here, the size of the receptive field in the proposed method may be varied by adjusting the parameters used in equations (2) and (4).

There are eight parallel arrays of neurons in the intermediate layer, each with the same dimension as the receptor layer. (Only one neuron, \(n_i\) for each intermediate layer array is illustrated in Figure 4, and some of the connections between the photoreceptor layer and the synapse filters have been omitted for visual clarity). Each intermediate layer neuron corresponds to an orientation specific receptive field and is capable of detecting intensity changes or depth discontinuities at specific orientations.

Each integrator neuron in the output layer integrates each of the eight corresponding neuronal outputs from the intermediate orientation specific receptive field neurons. The firing rate map of the output layer forms an edge map corresponding to the input images with intensity changes and depth discontinuities pooled into a single output. Thresholding is applied by simply selecting an appropriate threshold value, \(T\), either empirically or scientifically, and output layer neurons that have a firing rate equal or above the output neuron threshold result in the output neuron firing and that neuron to be considered as an edge feature point.

The network model was implemented using the Python programming language and the NumPy module, with the following parameters for the network: \(v_{\text{th}} = 60\text{mv}, v_{\text{reset}} = -70\text{mv}, E_{\text{ex}} = 0\text{mv}, E_{\text{ih}} = -70\text{mv}, g_t = 1.0\mu\text{S/mm}^2, c_{\text{m}} = 10\text{nF/mm}^2, \tau_{\text{ref}} = 6\text{ms}, A_{\text{ref}} = 0.028953\text{mm}^2\) and \(A_{\text{ref}} = 0.014103\text{mm}^2\). These parameters are consistent with biological neurons [17, 28]. The SNN is initialised with the default network parameters and image values are normalised such that the intensity values are in the range [0...1], range image values are in the range [0...1]. Range images positions with missing depth measurements, such as areas of the image scene located outside the sensor range, are considered to be equal to 0 in order to account for depth discontinuities. Synaptic strengths are adjusted to ensure that the neuron does not fire in response to uniform image data within its receptive field. During the simulation each receptive field is processed simultaneously in time and the output neuron potential is determined through the summation of the combined response from each receptive field.

IV. EXPERIMENTAL SIMULATION

In this section we provide a visual comparison of the extracted features using the SNN. The RGB-D images used in the experiments have been captured using a Prime Sense RGB-D camera and obtained from the public database [9].

Figure 5(a) illustrates the depth image and Figure 5(b) illustrates the intensity image (obtained by averaging the individual components from the captured RGB image). Note the upper half of the kettle bottle in Figure 5(b) is not clearly visible as it is only half full of ketchup. Figure 5(c) illustrates the feature outputs from the combined intensity and range SNN edge detection process discussed in Section III. In Figure 5(c) it can be seen that the edges from the bottle of ketchup have been identified using the network model, highlighting the complementary information available using both intensity and range images. If only intensity image data was available these edge features would not be identified as the transparent glass bottle is of a similar intensity value to the surrounding worktop.

To compare the performance of the SNN edge detector we present a visual comparison in Figure 6 with the well known Canny edge detector presented in [31] and the Jiang Scanline range image edge detector [32, 33] where the edges in the intensity and range images are detected separately using each respective detector. An advantage of the SNN edge
detector presented here is that the SNN naturally combines the multiple image inputs resulting in a single image output.

In Figure 6 we can see that the tissue paper sitting on top of the tissue paper box has some details missing in both the range edges in Figure 6(c) and the intensity edges in Figure 6(d). The depth edge detector struggles to detect a difference in depth profile along the top edge of the tissue box due to the small change in depth, but the intensity image edge detector can clearly differentiate between the different intensity values. Conversely, the intensity edge detector struggles to detect a difference in intensity values between the tissue paper and the background wall in areas, but the depth edge detector can clearly differentiate the shape of the tissue from the background wall due to the change in depth profile. An advantage provided by using the combined intensity and range feature detector is illustrated in the ability of the detector output in Figure 6(e) to detect all these features using complementary information from the range and intensity images, resulting in an edge feature output image with no requirement for individual thresholding. Similar example results are presented in Figure 7.

V. DISCUSSION AND FUTURE WORK

In this paper we presented a biologically motivated approach to feature extraction taking inspiration from the visual system. We used range and intensity images as input to the network with different types of photoreceptors in the input layer representing the different types of photoreceptors found in biology. The spiking neural network presented is constructed by a hierarchical structure that is composed of spiking neurons with scalable receptive fields as found in the visual cortex. The intermediate layer in our network consisted of various orientated receptive fields incorporating excitatory and inhibitory synapses that are capable of detecting depth discontinuities and intensity changes at various orientations. In the output layer our network uses integrator neurons that integrate the response from the intermediate layer neurons. The output layer neurons incorporate a threshold mechanism that when stimulated with an appropriate number of spikes produce a firing map corresponding to edge features from both the range and intensity images. The conductance based integrate and fire spiking neuron model used in this paper provide powerful biologically realistic functionality for integration of inputs and generation of spikes. Excitatory and inhibitory synapses are able to perform different complicated computations including detection of depth discontinuities, detection of intensity changes, integration of neuronal outputs and thresholding.

Range and intensity images are useful for many robot tasks such as creating models of physical objects or providing depth information that complements standard intensity images for various robot tasks. This paper demonstrates how the spiking neural network can detect edge features using range and intensity image inputs and illustrates visual performance and computational improvements over the standard approaches.

ACKNOWLEDGEMENT

This research is funded by the Centre of Excellence in Intelligent Systems project, funded by InvestNI and the Integrated Development Fund.
Fig. 6. - Visual comparison of SNN detector and standard edge and range detectors with tissue image

Fig. 7. - Visual comparison of SNN detector and standard edge and range detectors with table image
REFERENCES