The Hodgkin-Huxley neuron model for motion detection in image sequences

Hayat Yedjour, Boudjelal Meftah, Dounia Yedjour, Olivier Lézoray

The Hodgkin–Huxley neuron model for motion detection in image sequences

**Hayat Yedjour · Boudjelal Meftah ·
Dounia Yedjour · Olivier Lézoray**

Abstract In this paper, we consider a biologically-inspired spiking neural network model for motion detection. The proposed model simulates the neurons' behavior in the cortical area MT to detect different kinds of motion in image sequences. We choose the conductance-based neuron model of the Hodgkin-Huxley to define MT cell responses. Based on the center-surround antagonism of MT receptive fields, we model the area MT by its great proportion of cells with directional selective responses. The network's spiking output corresponds to an MT neuron population's firing rates and enables to extract motion boundaries. We conduct several experiments on real image sequences. The experimental results show the proposed network's ability to segregate multiple moving objects from an image sequence and reproduce the MT cells' responses. We perform a quantitative evaluation on the YouTube Motion Boundaries (YMB) dataset, and we compare the result to state-of-the-art methods for boundary detection in videos: boundary flow estimation (BF) and temporal boundary difference (BD). The proposed network model provides the best results on YMB compared to BF and BD methods.

H. Yedjour
Université Mohamed Boudiaf (USTO), Oran, 31000, Algérie

B. Meftah
LRSBG, Université Mustapha Stambouli, Mascara, 29000, Algérie
Corresponding author E-mail: boudjelal.meftah@univ-mascara.dz

O. Lézoray
GREYC, Université de Caen Normandie, 6 Bd. Maréchal Juin, Caen, 14050, France

D. Yedjour
Université Mohamed Boudiaf (USTO), Oran, 31000, Algérie

1 Introduction

Motion detection is a basic task that the visual system performs very efficiently. Several research studies have tried to relate our visual motion perception with the activation of the primary visual cortex (V1, or striate cortex), with extrastriate visual areas such as the medial temporal area (MT or V5) and with medial superior temporal area (MST) [1],[2], [3]. It seems that the area the most involved in motion processing is the MT area [4],[5], which receives input motion afferent mainly from V1 [6], [7]. In [8], the author has described the columnar structure of the middle temporal area MT, i.e., neurons in each column respond optimally to motion in a particular direction. This propriety is similar to that found in the striate cortex for contour orientation [9]. Adjacent columns of the columnar organization in MT sometimes have opposite preferred motion directions [8], [10]. The majority of the neurons in the middle temporal (MT) area have receptive fields with antagonistic surrounds [11],[12]. An essential utility of surround mechanisms is to detect motion discontinuities or motion boundaries [13],[14] based on the differential motion. There are different kinds of surround geometry [15],[16]. Neurons whose receptive fields have circularly symmetric surrounds are postulated to underlie figure-ground segregation [11],[17]. In a visual scene, spatial variations of brightness create contrasts that define edges with particular orientations in space. Detecting these edges' motion requires temporal delays to compare the current input with an earlier visual input. Recent anatomical and physiological studies suggest that motion-sensitive neurons in the primate receive two sets of inputs, one delayed for the other [18]. However, the origin of these delays remains unknown [20]. In the visual motion information domain, it has been found that the human visual system presents delays. These delays reside in separate classes of slow and fast synaptic transmission [22], [21].

Spiking Neuron Networks (SNN) are often referred to as the third generation of neural networks that can solve biological stimuli problems. Opportunities of using spiking neural networks as the most advanced technic for image processing have been proved in the literature by several authors that have developed a bioinspired spiking neural network models [23],[24],[25].

Various Spiking Neuron Networks models have been proposed for motion processing, such as [26], the authors have established a large-scale spiking model of the visual area MT capable of exhibiting both component and pattern motion selectivity. In [27], the authors implemented a biologically inspired visual system for motion detection and pursuit with a Discrete Leaky Integrate-and-Fire (DLIF) neuron model. In [28], the authors proposed bio–inspired motion features for action recognition and modeled different MT cells. Each MT cell was modeled as a conductance-driven integrate-and-fire neuron.

Some approaches have focused on axonal delay for motion detection, and those models were also validated, capturing specific properties of biological visual systems. For example, in [29], the authors proposed a spiking neural network to detect moving objects in a visual image sequence using the integrate and fire neuron model and axonal delay. In [30], a recurrent neural

network model has been used to capture the responses of a representative subset of MT neurons. Their model can successfully create temporal delays to reproduce the response of MT neurons. In [31], the authors have used the axonal delays to construct a spiking neural network for moving target detection and classification. In [32], the authors have developed a spiking neural network architecture for visual motion estimation that relies on synaptic delays to create motion-sensitive receptive fields.

More recently, new approaches based on deep learning have been proposed for motion detection from video sequences and have achieved promising performance on challenging benchmarks [33], [34], [35].

Most of the spiking neural network models proposed in the literature are mainly based on integrate-and-fire neuron models. However, while relatively accurate in describing the biophysical processes that occur during potential action generation, the Hodgkin–Huxley neuron model [36] is not widely used in theoretical studies, especially in motion detection. To test its effectiveness for visual motion detection, we introduce the conductance-based Hodgkin–Huxley neuron model [36] to reproduce the complex dynamics of action potential generation. We conceive a spiking neural network model to simulate visual motion detection so that more biological neuronal behaviors can be reflected in the proposed network model.

Our model exhibits axonal delay [29] at network input to consider intrinsic differences between slow and fast synaptic transmission. We focus on the cortical area MT, considering the results found by Born and Bradley [11] on center-surround interactions in the MT area and the preferred direction of motion for describing the direction selectivity of MT cells. Our goal is to propose a simplified bio–inspired model for motion detection, capturing motion information's critical biological properties. This paper allows demonstrating how a neural network based on both the Hodgkin-Huxley model and the center–surround properties of neurons in the middle temporal MT visual area can contribute to extract moving objects from a visual image sequence.

The paper is organized as follows. Section 2 introduces the Hodgkin-Huxley model. Section 3 describes the proposed spiking network structure for motion detection in which circularly symmetric surrounds receptive fields are described. Simulation results are presented in Section 4. The last section concludes.

2 Hodgkin–Huxley Model

The Hodgkin-Huxley (H-H) describes how action potentials in neurons are initiated and propagated as a mathematical model. It is a ensemble of nonlinear differential equations that approximates the electrical characteristics of excitable cells [37]. Based on experimentations using the giant axon of the squid, Hodgkin and Huxley found three different types of ion current: sodium, potassium and a leak current that consists mainly of Cl ions [36]. Their model consists of four equations:

$$C\frac{dV}{dt} = I - G_K n^4(V - E_K) - G_{Na}m^3 h(V - E_{Na}) - G_L(V - E_L) \qquad (1)$$

Where V is the membrane potential, C represents the specific membrane capacitance per unit area, I is the total membrane current per unit area, G_K and G_{Na} are the potassium and sodium conductances per unit area, E_K and E_{Na} are the potassium and sodium reversal potentials, respectively, and G_L and E_L are the leak conductance per unit area and leak reversal potential, respectively. This equation's time-dependent elements are V, G_K and G_{Na}, where the last two conductances depend explicitly on voltage as well.

$$\frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n \qquad (2)$$

$$\frac{dm}{dt} = \alpha_m(V)(1 - m) - \beta_m(V)m \qquad (3)$$

$$\frac{dh}{dt} = \alpha_h(V)(1 - h) - \beta_h(V)h \qquad (4)$$

α_i and β_i are rate constants for the ith ion channel, that depend on the voltage but not on time. n, m and h are dimensionless quantities between 0 and 1 that are associated with potassium channel activation, sodium channel activation and sodium channel inactivation, respectively. A conductance model was generated to approximate synaptic activity. The total synaptic current I was decomposed into a sum of two independent conductances as proposed in [38]:

$$I = g^{ex}(V - E_{ex}) + g^{in}(V - E_{in}) \qquad (5)$$

where g^{ex} and g^{in} are, respectively, time-dependent excitatory and inhibitory conductances. E_{ex} and E_{in} are their respective reversal potentials.

As this is generally the case for all spiking neuron models, when the membrane potential V of neuron reaches the spike threshold V_{th} at time t, a spike S is generated and the membrane potential V is reset to V_{reset}. Any further inputs to the neuron are ignored during a time τ_{ref} called absolute refractory period. S represents a spike train array encoded by binary ones and zeros, generated by the neuron.

$$\begin{aligned} &\text{if } (V(t) > V_{th}) \text{ then } S(t) = 1 \text{ and } V(t) = V_{reset} \\ &\text{else } S(t) = 0 \end{aligned} \qquad (6)$$

3 Spiking Network structure for motion detection

The architecture of the proposed motion model is composed of three layers as illustrated in Figure 1. The first layer represents the input image sequence in grayscale denoted by I named receptor surface. It is supposed that the receptor layer can transform an image $I(x, y)$ at time t into a spike train which is fed to spiking neurons in an intermediate layer. Each pixel $I(x, y)$ in the receptor layer is then converted into a resulting spike train.
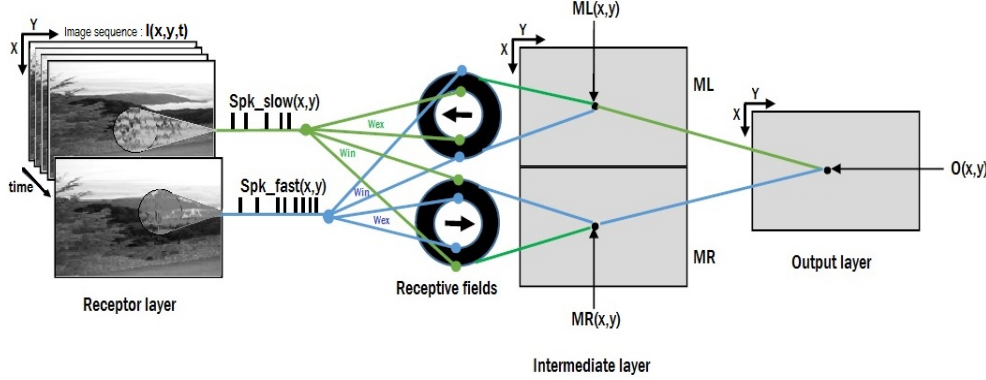

Fig. 1 SNN structure proposed for motion detection.

To models temporal delays of synaptic transmission, we consider two delayed spike trains generated at position (x, y) in the receptor layer referred to as fast $Spk_{fast}(x, y)$ and slow $Spk_{slow}(x, y)$ synaptic transmissions [21],[39]. These spike trains cause a change of peak conductances q_{fast} and q_{slow} where $q_{fast}(x, y)$ represents the peak conductance caused by the current $Spk_{fast}(x, y)$ from a receptor at (x, y) and $q_{slow}(x, y)$ represents the peak conductance caused by the current $Spk_{slow}(x, y)$ from a receptor at (x, y). They are computed using the following formulas:

$$q(x, y, t) = g_{norm} I(x, y, t) \qquad (7)$$

Where g_{norm} is a normalization constant, i.e., $g_{norm} = 1/I_{max}$. I_{max} is the maximum pixel value in a grayscale image at time t.

To model the direction selectivity of the area MT [8] ; the intermediate layer includes two neuron subpopulations responsive to motion directions according to the center and surrounds of their receptive fields [11]. They are modeled with a Difference-of-Gaussians (DoG) filter. The first subpopulation in the intermediate layer is denoted MR. Each neuron MR(x,y) has excitatory connections to the receptor layer with fast synaptic transmission. Furthermore, inhibitory connections with slow synaptic transmission and can respond

to a rightward motion. The second subpopulation in the intermediate layer is denoted ML. Each of its neuron ML(x,y) has excitatory connections to the receptor layer with slow synaptic transmission and inhibitory connections with fast synaptic transmission and can respond to a leftward motion within the receptive field.

The distribution of the weights for excitatory and inhibitory synapses connections W_{ex} and W_{in} are computed as follows [40]:

$$W^{ex} = \begin{cases} \frac{1}{2\pi\sigma_c^2} exp\left[-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma_c^2}\right)\right] & \text{if } \sqrt{x^2+y^2} \leq R \\ 0 & \text{if } \sqrt{x^2+y^2} > R \end{cases} \tag{8}$$

$$W^{in} = \begin{cases} 0 & \text{if } \sqrt{x^2+y^2} \leq R \\ \frac{1}{2\pi\sigma_s^2} exp\left[-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma_s^2}\right)\right] & \text{if } \sqrt{x^2+y^2} > R \end{cases} \tag{9}$$

R is the DoG center radius, σ_c and σ_s are the standard deviations of the center and the surrounding elements of the DoG filter.

For MR neurons at the intermediate layer, the conductances $g^{ex}_{MR(x,y)}(t)$ and $g^{in}_{MR(x,y)}(t)$ are governed by the following equations:

$$\frac{g^{ex}_{MR}(t)}{dt} = -\frac{1}{\tau_{ex}} g^{ex}_{MR}(t) + \sum_{r=-R}^{R}\sum_{f=-R}^{R} W^{ex}_{MR(r,f)} q_{fast}^{ex}{}_{(x-r,y-f)} \tag{10}$$

$$\frac{g^{in}_{MR}(t)}{dt} = -\frac{1}{\tau_{in}} g^{in}_{MR}(t) + \sum_{r=-R}^{R}\sum_{f=-R}^{R} W^{in}_{MR(r,f)} q_{slow}^{in}{}_{(x-r,y-f)} \tag{11}$$

By analogy to the MR array, the conductances $g^{ex}_{ML(x,y)}(t)$ and $g^{in}_{ML(x,y)}(t)$ for each neuron in the ML array are computed as follows:

$$\frac{g^{ex}_{ML}(t)}{dt} = -\frac{1}{\tau_{ex}} g^{ex}_{ML}(t) + \sum_{r=-R}^{R}\sum_{f=-R}^{R} W^{ex}_{ML(r,f)} q_{slow}^{ex}{}_{(x-r,y-f)} \tag{12}$$

$$\frac{g^{in}_{ML}(t)}{dt} = -\frac{1}{\tau_{in}} g^{in}_{ML}(t) + \sum_{r=-R}^{R}\sum_{f=-R}^{R} W^{in}_{ML(r,f)} q_{fast}^{in}{}_{(x-r,y-f)} \tag{13}$$

Where τ_{ex} and τ_{in} are respectively synaptic decay times for excitatory and inhibitory synapses respectively. W^{ex}_{MR} and W^{in}_{MR} represent the strength for excitatory and inhibitory synapses respectively for MR neurons. W^{ex}_{ML} and W^{in}_{ML} represent the strength for excitatory and inhibitory synapses respectively for ML neurons. q^{ex}_{fast} and q^{in}_{fast} represent the peak conductances for excitatory and inhibitory of fast synapses respectively. q^{ex}_{slow} and q^{in}_{slow} represent the peak conductances for excitatory and inhibitory slow synapses respectively.

According to Eq. 1 of the conductance-based (H-H) model, the dynamics of the membrane potential V for a neuron MR and ML at the intermediate layer are given by the following equations:

$$C\frac{dV_{MR}}{dt} = g_{MR}^{ex}(V-E^{ex})+g_{MR}^{in}(V-E^{in})-G_Kn^4(V-E_K)-G_{Na}m^3h(V-E_{Na})-G_L(V-E_L)$$
(14)

$$C\frac{dV_{ML}}{dt} = g_{ML}^{ex}(V-E^{ex})+g_{ML}^{in}(V-E^{in})-G_Kn^4(V-E_K)-G_{Na}m^3h(V-E_{Na})-G_L(V-E_L)$$
(15)

The spike trains generated in the intermediate layer feed the output layer, where the activation of each output neuron depends on the activation of the intermediate layer. If the current $Spk_{fast}(x,y)$ is equal to the current $Spk_{slow}(x,y)$, MR and ML neurons are silents, and consequently, the output layer will not generate spikes. If the currents $Spk_{fast}(x,y)$ and $Spk_{slow}(x,y)$ are different, then a movement from two consecutive input stimulus can be assumed. In that case, MR neurons will generate a spike while respecting the spike condition of (Eq. 6) if $Spk_{fast}(x,y) < Spk_{slow}(x,y)$ and therefore, it can respond to a rightward motion detected with MR DoG receptive fields. For ML neurons, it will generate a spike if S$Spk_{slow}(x,y) > Spk_{fast}(x,y)$, and therefore, it can respond to a leftward motion within the receptive field. The output layer denoted by O receives feedforward connections outputted from the intermediate layer. Each neuron $O(x,y)$ integrates the corresponding neurons' outputs from the intermediate layer and can fire if a neuron MR or a neuron ML fires.

If we denote by $S_{MR}(t)$ and $S_{ML}(t)$ the spike trains for neurons MR and ML respectively, each output neuron $O(x,y)$ is computed using the following equations:

$$\frac{g_{O(x,y)}^{ex}(t)}{dt} = -\frac{1}{\tau_{ex}}g_{O(x,y)}^{ex}(t) + (W_1S_{MR}(t) + W_2S_{ML}(t))$$
(16)

$$C\frac{dV_{O(x,y)}(t)}{dt} = g_{O(x,y)}^{ex}(V-E^{ex})-G_Kn^4(V-E_K)-G_{Na}m^3h(V-E_{Na})-G_L(V-E_L)$$
(17)

Spike condition in Eq. 6 is applied to the output neuron $O(x,y)$ to generate a spike train array denoted by $S_{O(x,y)}(t)$. The firing rate $F_{O(x,y)}$ for neuron $O(x,y)$ during the simulation time T in the output layer is computed by the following expression:

$$F_{O(x,y)} = \frac{\sum_{t=1}^{T} S_{O(x,y)}(t)}{T}$$
(18)

4 Simulation results and discussion

We have choosen to implement our model in MATLAB environment using a set of parameters values as recommended by biophysical model [36]. They

are summarized as follows: $C = 1\mu F/cm^2$, $E_L = 10.613mV$, $E_{Na} = 115mV$, $E_K = -12mV$, $G_L = 0.3mS/cm^2$, $G_{Na} = 120mS/cm^2$, $G_K = 36mS/cm^2$, $V_{th} = -65mV$, $V_{reset} = -70mV$, $E_{ex} = 0mV$, $E_{in} = -75mV$, $\tau_{ex} = 3ms$, $\tau_{in} = 10ms$, $T = 100ms$, $\tau_{ref} = 10ms$. Synapse weights $W_1 = W_2 = 1$. (n ; m, and h) for activation and inactivation and are between 0 and 1. (n_0 ; m_0, and h_0) are their initial values, and are calculated using the resting states defined by : $dn/dt = 0$; $dm/dt = 0$ and $dh/dt = 0$, of Eqs. (2)-(4) respectively. With the value $V = 0$ of membrane potential, the following initial point is found: ($n_0 = 0.1765$; $m_0 = 0.0529$ and $h_0 = 0.5961$).

Table 1 summarizes the analogy between the real parameters of the (H-H) [36] and those choosen in our simulation for action potentials generation and propagation in layers of the proposed SNN network.

Table 1 Biological correspondence of the (H-H) model parameters used for the proposed SNN network

(H-H) model parameters	Real Values	Experimental values
Membrane capacitance per unit area ($\mu F/cm^2$)	$C = 0.91$	$C = 1$
Leakage reversal potential (mV)	$E_L = 10.6$	$E_L = 10.613$
Sodium reversal potential (mV)	$E_{Na} = 115$	$E_{Na} = 115$
Potassium reversal potential (mV)	$E_K = -12$	$E_K = -12$
Maximal channel conductance per unit area (mS/cm^2)	$G_L = 0.3$	$G_L = 0.3$
Maximal sodium conductance per unit area (mS/cm^2)	$G_{Na} = 120$	$G_{Na} = 120$
Maximal potassium conductance per unit area (mS/cm^2)	$G_K = 36$	$G_K = 36$
Initial activation value of potassium channel	$n_0 = 0.31773241094$	$n_0 = 0.1765$
Initial activation value of sodium channel	$m_0 = 0.0529551709$	$m_0 = 0.0529$
Initial inactivation value of sodium channel	$h_0 = 0.5959943932$	$h_0 = 0.5961$

To ensure the DoG filter approximates the receptive fields of biological MT Cells, the center-surround size ratio is set to 1.5 ($\sigma_s/\sigma_c = 1.5$) [41].

By plotting the firing rates of the output layer as an image, we can obtain visual motion boundaries representing the moving objects corresponding to the input image sequence as shown in Figure 2. The input is an image sequence. The directionally-selective of intermediate layer filters are applied over each frame of the input sequence in a DoG distribution. The edges of firing neuron groups [25] are used to produce spikes in output layer according to equations 14-17.

Figure 2 presents an example of a video sequence processing from the Weizmann dataset [42]. The top row shows a video sequence where a person walks in the rightward direction, the neurons that respond strongly to a right direction of movement fail to respond to the left direction, as shown in the
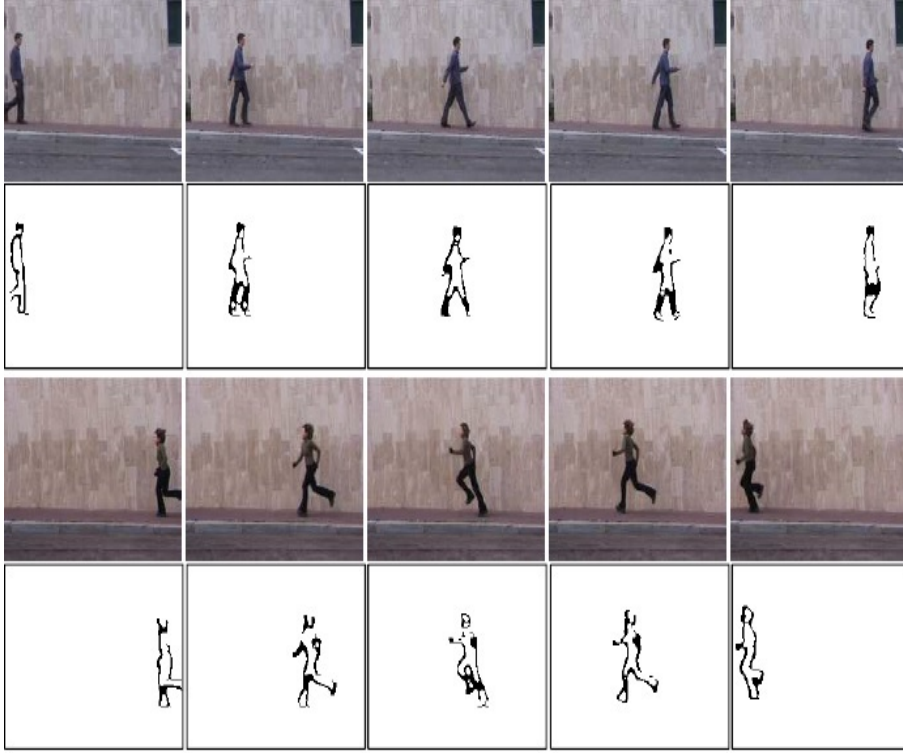
Fig. 2 Motion detecting Snn. First and third rows show frames from an input video sequence from the Weizmann dataset that shows a moving object. Second and fourth rows show the contours of the moving object obtained by the proposed approach.

second row. The second video sequence presented in the third row shows a person runs in the leftward direction. The neurons that respond strongly to a left direction of movement fail to respond to the right direction, as shown in the fourth row. These simulation results show that the proposed SNN can segregate moving objects from an image sequence consistent with the known properties of area MT [8].

The proposed SNN has experimented with many image sequences from the YouTube Motion Boundaries dataset (YMB) [35]. This dataset contains 60 sequences captured from real-world scenes with complex background including varieties of persons, objects and poses. The number of frames per sequence varies and can be reach until 100 frames. In each sequence, motion boundaries in one frame are provided by three independent annotators. The dataset includes two types of videos: 30 sequences from the YouTube Objects dataset [44] representing 10 object categories, such as train, car or dog, and 30 others from the Sports1M [45] dataset comprises 487 classes. Videos from YouTube Objects

have a lower resolution (225x400) than the ones from Sports1M (1280x720). Both datasets contain large motions.

The SNN takes input an image from an image sequence where motion detection is performed at each time step. The presented image sequences concern position invariant patterns positions invariance and also contain right and left translations. In Figure 3, we present results from the YMB dataset using the proposed approach.

Fig. 3 Motion detection results on YMB dataset. First row shows a frame from a video sequence. Second row shows the associated ground truth frames. Third row shows the estimated motion boundaries with our proposed approach.

4.1 Motion detection performance

In order to test the robustness of our approach. We compare our approach against two state-of-the-art motion detection boundaries methods. The first one is based on the temporal difference (will be denoted by BD) computed on Canny edges, and the second one is based on optical flow (will be denoted by BF) using the Lucas-Kanade optical flow method [43].

The BD method [46] used in this work is inspired by Haynes technique [47] which consists of performing the simple product between the absolute frame difference FD and the gradient image G (Equation 19).

$$FD(x, y, t) = |CF(x, y, t) - PF(x, y, t)| \tag{19}$$

Where CF represents of current frame and PF the previous frame in an image sequence.

The gradient image G in Equation 20 is computed on Canny edges technique. The result of BD method is a motion boundaries detection [46].

$$BD(x, y, t) = FD(x, y, t) * G[CF(x, y, t)] \tag{20}$$

For motion boundaries based on optical flow estimation (BF) and in order to determine a vector motion field within an image sequence. First, we calculate the motion information based on the Lucas-Kanade algorithm between two consecutive frames. Then, based on maximum flow present, we proceed to flow field color coding to highlight the motion estimation result and finally, a segmentation is performed on the obtained motion estimation that will be assigned to the motion boundaries computed as the gradient of the optical flow estimation (BF) [48].

The advantage of combining motion information with segmentation allows it possible to more clearly highlight objects of interest among more complex scenes.

To quantify the performance of the proposed SNN for motion detection, we performed an analysis in terms of peak signal-to-noise ratio (PSNR), mean squared error (MSE), the figure of merit (FOM), and the Hausdorff distance measure (HD). They are computed as follows:

Let $A = \{a_i\}$, $i = 1, ..., n$ (as ground truth motion points) and $B = \{b_j\}$, $j = 1, ..., m$ (as image motion points).

$$PSNR = 10 log_{10} \left(\frac{M^2}{MSE} \right) \tag{21}$$

Where M is the maximum pixel value.

$$MSE = \frac{1}{M * N} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [A(i, j) - B(i, j)]^2 \tag{22}$$

$$HD = max[h(A, B), h(B, A)] \tag{23}$$

$$h(A, B) = sup_{a \in A} inf_{b \in B} \|a - b\| \tag{24}$$

Where $\|.\|$ is the distance between two points a_i and b_j measured by some norm (generally the Euclidean norm).

$$FOM = \frac{1}{Max(N_A, N_B)} \sum_{k=1}^{N_B} \frac{1}{1 + \alpha * d^2(i)} \tag{25}$$

Where N_A is the number of edge points on the ground truth image. N_B is the number of detected edge points. d is the distance between the kth detected

edge pixel and the nearest edge pixel of the ground truth and α is a scaling factor, most commonly chosen to be 1/9.

Given two consecutive video frames and their corresponding binary ground-truth motion boundary for each sequence from the YMB dataset, we compute the binary version of the moving boundary detectors for the proposed approach, for the flow estimation (BF), and the temporal difference (BD) methods. The similarity between the motion boundary detectors to the ground-truth motion boundary is evaluated using MSE, PSNR, HD, and FOM for each sequence. The quantitative results are finally averaged for all image sequences for the whole YMB dataset and are summarized in Table 2.

Table 2 Evaluation of motion detection approaches

Performance measures	BD	BF	Proposed approach
MSE	0.1028	0.0863	**0.0224**
PSNR	17.1526	17.5775	**18.4856**
HD	8.4068	6.7665	**5.1630**
FOM	0.9672	0.9774	**0.9802**

For the PSNR and FOM measures, the higher is their evaluation score, the stronger is a motion boundary. Unlike for MSE and HD measures, the lower the evaluation score, the stronger the motion boundary for a given approach.

MSE is used as a measure to evaluate our approach, this measure can indicate how close a motion boundary model is to the ground-truth motion boundary. From the table 2, all positives values indicating there is some difference in detectors and ground-truth. Lower the MSE indicates model is closer to ground truth. In our proposed approach, MSE is minimum comparing to (BD) [46] and (BF) methods [48], demonstrating that the proposed motion boundary model and the ground-truth motion boundary are getting closer to each other.

The PSNR represents the ratio for measuring the quality of the motion boundaries results produced by a motion boundary detectors and the ground-truth motion boundary. The higher the PSNR, the better the quality of the moving boundary detector to match the ground-truth's one. This would occur because we wish to minimize the MSE between the two compared motion maps. The proposed approach resulted in the highest PSNR compared (BD) [46] and (BF) approaches [48].

FOM is a measure for accuracy assessment of motion boundary detection, it represents the deviation from ground-truth to the detected motion boundary, and this measure varies in the range of [0, 1], where 1 represents the optimal value. In Table 2, it can easily be noticed that the proposed approach offer a high FOM value, which explains the great accurately matching with the ground truth motion boundary.

The Hausdorff distance is an algorithm used to find a given ground truth motion boundary in an arbitrary motion boundary detector. These must be in

binary images. The algorithm then tries to minimize the Hausdorff distance between these two resultants binary images. The minimal Hausdorff distance can be considered the best candidate for locating the ground truth motion boundary and arbitrary motion boundary detector. Since this measure is used to measure the difference between two different representations of the same frame in image sequence, it is observed in Table 2, that the proposed approach gives a lesser Hausdorff distance. Therefore, the proposed model yields bet-ter performance compared to the (BD) [46] and (BF) [48] motion boundary detectors.

From Table 2, we can conclude that our approach outperforms (BF) and (BD) methods significantly by allowing optimized extraction of moving contours.

Figure 4 provides qualitative results of our approach compared to BF and BD methods for four images from the YMB dataset.

Fig. 4 Results on YMB dataset. In each row from top to bottom, frame from a video sequence, the associated ground truth frame annotation, motion boundaries based on temporal difference (BD), motion boundaries based on optical flow estimation (BF), motion boundaries computed as the gradient of the optical flow estimation (BF), and our boundary detection approach.

Motion boundaries based on the temporal difference (BD) are not robust to noise; this is proved especially with the second image. In contrast, the proposed approach successfully deals with noise by means of temporal integration and inhibiting texture edges while preserving object contours and region boundaries. The fourth image indicates that some object motions are missed in the flow estimation (BF). As can be seen, the resulting motion representation is considerably accurate; since it produces sharp motion boundaries even on small objects such as the balloon in the first frame.

In order to test the performance of our model against new approaches for motion boundary detection based on deep learning, we compare our model to two dedicated deep approach for automatic motion boundary detection. The first is a learning-based approach using structured random forests for motion boundary detection (LDMB) [35]. During training, they use images and ground-truth optical flows from the MPI-Sintel benchmark [49]. The output of the different trees are then averaged to produce the final boundary detection. The second termed Motion Boundary Network (MoBoNet) [33]. MoBoNet takes as input source images, flow estimations and warping errors, and directly produces the corresponding motion boundaries.

The motion boundary detection evaluation is carried on motion boundary datasets generated from the optical flow datasets: Sintel [49], KITTI [50], Middleburry [51] and YMB dataset [35].

The MPI-Sintel dataset [49] is composed of animated sequences generated using computer graphics. It provides dense ground truth for optical flow for each frame. This makes the dataset a very reliable test set for comparison of methods. This dataset is available in two versions: clean and final.

The KITTI dataset produced in 2012 [50] was used to test our model. This dataset contains 194 real-world image pairs of cameras mounted on a car. Ground truth for optical flow is obtained from a 3D laser scanner combined with the egomotion data of the car.

The Middlebury dataset [51] is composed of 8 training sequences of 2 to 8 frames with ground-truth optical flow for a central frame.

The performance indicators called recall R and precision P are used to measure the similarity between binary motion map obtained by a given model and those from ground truth. Recall R is defined as the fraction of true contour pixels (according to the ground truth) that are successfully detected by a given model. Precision P is defined as the fraction of true contour pixels from all the detected ones. The evaluation code is provided by edge detection benchmark BSDS [52]. In order to come to a single performance measure that would allow for results comparison, as suggested in [52], we compute the mean Average-Precision of recall and precision, commonly referred to as the (mAP), for this, Precision-recall are finally averaged for all images and all binary versions of the ground-truth. The higher is the evaluation score, the stronger is a motion boundary detection.

The (mAP) values of both (LDMB) [35] and (MoBoNet) [33] approaches depend on the algorithm used to estimate the optical flow. In Table 3, there are three different algorithms : DeepFlow [53], FlowNet [54] and DC-Flow [55].

For evaluation, binary motion boundaries must be computed from ground-truth optical flow. For YMB dataset [35], the binary motion boundary ground-truth is given. For Sintel, KITTI and Meddleburry datasets, the Binary motion boundary ground truth is computed from corresponding optical flow datasets. For the proposed SNN, we compare the binary motion boundaries directly computed from the proposed SNN approach. The comparison of the performance (mAP) of our approach to LDMB [35] and MoBoNet [33] approaches for different input flows for all datasets is given in Table 3.

Table 3 Comparison of the performance (mAP) of SNN approach to LDMB and MoBoNet approaches for different input flows

	MPI-sintel clean	MPI-sintel final	KITTI	Middleburry	YMB
LDMB based DeepFlow	75.8	67.7	65.2	89.0	68.6
LDMB based FlowNet	68.4	59.7	62.6	82.5	64.3
LDMB based DC-Flow	83.2	75.6	74.3	94.2	77.5
MoBoNet based DeepFlow	75.8	67.7	65.2	89.0	68.6
MoBoNet based FlowNet	75.7	67.9	68.7	87.2	70.6
MoBoNet based DC-Flow	85.4	78.1	75.7	95.6	79.1
Proposed approach	**65.3**	**68.5**	**62.0**	**96.2**	**79.7**

Compared to LDMB [35] and MoBoNet [33] methods for different input flows, the proposed SNN achieves superior performance on the Middleburry and YMB motion boundary datasets. The values obtained on YMB dataset are slightly lower compared to Middleburry dataset. This can be explained as Middlebury dataset mainly contains small motions that can be easily estimated, while YMB dataset contains scenes with a moving and complex background with large motions. LDMB and MoBoNet methods achieves slightly supe-rior performance compared to the proposed SNN on MPI-Sintel and KITTI datasets. The reason is that these approaches instead a learning-based approaches for the motion boundary prediction problem. This requires a high volume of training data. However, the learning mechanism was not considered in the proposed SNN model.

Based on results of Table 3, the proposed SNN model achieves motion boundary detection accuracies higher than LDMB and MoBoNet methods. Generally, indicating more accurately matching with the ground-truth. Our SNN model can generate a quality motion boundaries despite the learning mechanism was not considered in the SNN model. Therefore, the proposed SNN can be improved to perform a set of learning mechanisms to process complex information and therefore having motion boundaries predicted with higher accuracy.

Finally, the directionally-selective of the proposed model allows optimized extraction of moving contours. Also, it takes advantage of the luminance correction provided by the edges of firing neuron groups [25] used in the input layer, which allows extraction of mobile contours even in case of noise.

5 Conclusion

We have developed a hierarchical feedforward model based on spiking neurons for the MT area. The spike generation process uses the conductance-based of the Hodgkin-Huxley neuron model that can well reproduce the MT neurons' response properties. Our model implements two different MT receptive fields with center-surround interactions. The neuron's center-surround properties in the MT area has been proved in [11] to contribute to the behavioral segregation of objects from the background. Synapses play the role of filters as well as motion differentiator. Our experiments on the YMB dataset for motion boundary detection demonstrate that the proposed model performs better than flow-based motion boundaries (BF) and temporal difference based motion boundaries (BD). As it can distinguish between the object's motion from that of the background and it succeeds in detecting boundaries motion corresponding to various motion patterns in real image sequences. The obtained results showed a high similarity with the known properties of neurons in motion processing. In future work, we will include object tracking or action recognition based on this model's generated motion maps.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest statement.

References

1. T.D. Albright, Form–cue invariant motion processing in primate visual cortex,Science, Vol. 255, pp.1141-1143 (1992)
2. J.H.R. Maunsell, W.T. Newsome, Visual processing in monkey extrastriate cortex,Annual Review of Neuroscience, volume 10(1), pp. 363-401 (1987)
3. Z. Lu, X. Li, M. Meng, Encodings of implied motion for animate and inanimate object categories in the two visual pathways,NeuroImage, volume 125,pp. 668-680 (2016)
4. R. Dubner, S. M. Zeki, Response properites and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey, Brain Research, volume 35, pp. 528-532 (1971)
5. R.M. Ruhl, T. Bauermann, M. Dieterich, P.Z. Eulenburg,Functional correlate and delineated connectivity pattern of human motion after effect responses substantiate a subjacent visual–vestibular interaction, Neuroimage, volume 10(174), pp. 22-34 (2018)
6. D.J. Felleman, D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, Cereb. Cortex, volume 1(1), pp. 1-47 (1991)
7. S.A. Kozhukhov, K.A. Saltykov, N.A. Lazareva, The Structure of Primary Visual Cortex Neuron Responses and its Interaction with the Dynamics of the Preferred Orientation, Neurosci Behav Physi, volume 47, pp. 402 (2017)
8. T.D. Albright, Direction and orientation selectivity of neurons in visual area of the macaque, J Neurophysiol, volume 52(6), pp. 1106-1130 (1984)
9. D.H. Hubel, T.N Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J Physiol, volume 160, pp. 106-154 (1962)
10. A. Roy, I.K. Christie, G.M. Escobar, Does experience provide a permissive or instructive influence on the development of direction selectivity in visual cortex?, Neural Dev, volume 13, pp. 16 (2018)

11. R.T. Born, D.C. Bradley, Structure and function of visual area MT,Annual Review Neuroscience, volume 28, pp. 157-189 (2005)
12. D.K. Khachvankian, A.L. Ghazaryan, B.A. Harutiunian-Kozak, Expansion of Visual Receptive Fields in the Extrastriate Visual Cortex: Dependence on the Trajectory of a Moving Stimulus, Neurophysiology, volume 49(2), pp. 122-129 (2017)
13. K. Nakayama, J.M. Loomis, Optical velocity patterns, velocity–sensitive neurons, and space perception: A hypothesis, Perception, volume 3(1), pp. 63-80 (1974)
14. P. Zarei, T. Kameneva, M.R. Ibbotson,The interaction between integration and segmentation neurons for motion perception, BMC Neuroscience, volume 16(86) (2015)
15. D.K Xiao, S.E. Raiguel, V. Markar, G.A Orban, The spatial distribtion of the antagonistic surround of MT/V5 neurons, Cerebral Cortex, volume 7(7), pp. 662-667 (1997)
16. H Wang, Z Wang, Y Zhou, T Tzvetanov, Near and Far Surround Suppression in Human Motion Discrimination, Front Neuroscience, volume 12, pp. 206 (2018)
17. H.B Turkozer, Z Pamir, H Boyaci, Contrast affects fMRI activity in middle temporal cortex related to center–surround interaction in motion perception, Front Psychol, volume 7, pp. 454 (2016)
18. Y Ohnishi, K Kawano, K Miura, Temporal impulse response function of the visual system estimated from ocular following responses in humans, Neurosci Res, volume 113, pp. 56-62 (2016)
19. R.L DeValois, N.P Cottaris, L.E Mahon, S.D Elfar, J.A Wilson, Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity, Vision Res, volume 40, pp. 36853702 (2000)
20. N.J Priebe, D Ferster, Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex, Neuron, volume 45, pp. 133-145 (2005)
21. R Maex, G.A Orban, Model circuit of spiking neurons generating directional selectivity in simple cells, J Neurophysiol, volume 75, pp. 1515-1545 (1996)
22. C Oluk, A Pavan, H Kafaligonul, Rapid motion adaptation reveals the temporal dynamics of spatiotemporal correlation between ON and OFF pathways, Sci Rep, volume 6, pp. 1-10 (2016)
23. B Meftah, O Lézoray, A Benyettou, Segmentation and edge detection based on spiking neuron networks, Neural Processing Letters, volume 32(2), pp. 131-146 (2010)
24. B Meftah, Lézoray, S Chaturvedi, A Khurshid, A Benyettou, Image processing with spiking neuron networks, Artificial Intelligence, Evolutionary Computing and Metaheuristics, Studies in Computational Intelligence, Yang, Xin-She (Ed.), volume 427, pp. 525-544 (2013)
25. H Yedjour, B Meftah, Lézoray, A Benyettou, Edge detection based on Hodgkin-Huxley neuron model simulation, Cognitive Processing, volume 18 (3), pp. 315-323 (2017)
26. M Beyeler , M Richert , N.D Dutt , J.L Krichmar, Efficient spiking neural network model of pattern motion selectivity in visual cortex, Neuroinformatics, volume 12(3), pp. 435-454 (2014)
27. L Risinger, K Kaikhah, Motion detection and object tracking with discrete leaky integrate–and–fire neurons, Appl Intell, volume 29, pp. 248-262 (2008)
28. M.J Escobar , P Kornprobst, Action recognition via bio-inspired features: The richness of center-surround interaction, Computer Vision and Image Understanding, volume 116 (5), pp. 593-605 (2012)
29. Q Wu , T.M McGinnity , L.P Maguire , J Cai , G.D Valderrama–Gonzalez, Motion Detection Using Spiking Neural Network Model, Huang, D.-S., Wunsch II, D.C., Levine, D.S., Jo, K.-H. (eds.) ICIC 2008. LNCS (LNAI), 7683, (2008,
30. J Joukes , T.S Hartmann , B Krekelberg, Motion detection based on recurrent network dynamics, Front Syst Neurosci, volume 8, pp. 239 (2014)
31. R Cai , Q Wu , P Wang , H Sun , Z Wang, Moving Target Detection and Classification Using Spiking Neural Networks, Intelligent Science and Intelligent Data Engineering IScIDE 2011, October 23-25, 2011, Xi'an, China, volume 7202 (2012)
32. G Orchard , R Benosman , R Etienne-Cummings , N.V Thakor, A spiking neural network architecture for visual motion estimation, IEEE Biomedical Circuits and Systems Conf. (BioCAS), 31 Oct.-2 Nov. 2013, Rotterdam, Netherlands, pp. 298-301 (2013)
33. X Yin, X Dai, X Wang, M Zhang, D Tao, L Davis, Deep motion boundary detection, CoRR abs/1804.04785, (2018)

34. J Cheng, Y Tsai, S Wang, M Yang, SegFlow: Joint Learning for Video Object Segmentation and Optical Flow, IEEE International Conference on Computer Vision (ICCV), (2017)
35. P Weinzaepfel , J Revaud , Z Harchaoui , C Schmid, Learning to detect motion boundaries, CVPR - IEEE Conference on Computer Vision and Pattern Recognition, Jun 2015, Boston, United States IEEE, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 2578-2586 (2015)
36. A Hodgkin , A Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J Physiol, volume 117, pp. 500-544 (1952)
37. M.E Nelson, Electrophysiological models of neural processing, Wiley Interdisciplinary Reviews: Systems Biology and Medicine, volume 3(1), pp. 74-93 (2011)
38. A Destexhe , M Rudolph , J.M Fellous , T Sejnowski, Fluctuating synaptic conductances recreate in vivo–like activity in neocortical neurons, Neuroscience, volume 107, pp. 13-24 (2001)
39. B Conway , M Livingstone, Space–time maps and two bar interactions of different classes of direction-selective cells in macaque V1, of Neurophysiology, volume 89, pp. 2726-2742 (2003)
40. Q.X Wu , T.M McGinnity , L.P Maguire , A Belatreche , B Glackin, Processing visual stimuli using hierarchical spiking neural networks, Neurocomputing, volume 71, pp. 2055-2068 (2008)
41. N Petkov , W.T Visser, Modifications of center–surround, spot detection and dot–pattern selective operators, Technical Report 2005-9-01, Institute of Mathematics and Computing Science, University of Groningen, Netherlands (2005)
42. S Alpert , M Galun , R Basri , A Brandt, Image segmentation by probabilistic bottom–up aggregation and cue integration, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 34 (2), pp. 315-327 (2012)
43. D Lucas , T Kanade, An iterative image registration technique with an application to stereo vision, Proceedings of Imaging Understanding Workshop, pp. 121-130 (1981)
44. A Prest, C Leistner, J Civera, C Schmid, V Ferrari, Learning object class detectors from weakly annotated video, In CVPR, (2012)
45. A Karpathy, G Toderici, S Shetty, T Leung, R Sukthankar, L Fei-Fei, Large-scale video classification with convolutional neural networks, In CVPR, (2014)
46. H Yedjour, D Yedjour, B Meftah, Motion Boundary Detection Improved by Bio-Inspired Approach, International Journal on Communications Antenna and Propagation (IRE-CAP), volume 9 (5) pp. 343-350 (2019)
47. S M Haynes, R Tain, Detection of moving edges, Computer Vision, Graphics, and Image Processing, volume 21(3), pp. 345367 (1983)
48. H Yedjour, Optical Flow Based on Lucas-Kanade Method for Motion Estimation, Hatti M. (eds) Artificial Intelligence and Renewables Towards an Energy Transition. ICAIRES 2020. Lecture Notes in Networks and Systems, volume 174 (2021)
49. D J Butler, J Wulff, G B Stanley, M J Black, A naturalistic open source movie for optical flow evaluation, In European Conference on Computer Vision (ECCV), pp. 611-625 (2012)
50. A Geiger, P Lenz, R Urtasun, Are we ready for autonomous driving? the kitti vision benchmark suite, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354-3361 (2012)
51. S Baker, D Scharstein, J P Lewis, S Roth, M J Black, R Szeliski, A database and evaluation methodology for optical flow, 11th IEEE International Conference on Computer Vision (ICCV), (2011)
52. D Martin, C Fowlkes, D Tal, J Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, In IEEE International Conference on Computer Vision (ICCV), (2001)
53. P Weinzaepfel, J Revaud, Z Harchaoui, C Schmid, Deepow: Large displacement optical flow with deep matching, In IEEE International Conference on Computer Vision (ICCV), pp. pages 1385-1392 (2013)
54. A Dosovitskiy, P Fischer, E Ilg, P Hausser, C Hazirbas, V Golkov, P Smagt, D Cremers, T Brox, Flownet: Learning optical flow with convolutional networks, In IEEE International Conference on Computer Vision, pp. 2758-2766 (2015)
55. J Xu, R Ranftl, V Koltun, Accurate optical flow via direct cost volume processing, arXiv preprint arXiv:1704.07325, (2017)