Visual Motion Pattern Extraction and Fusion for Collision Detection in Complex Dynamic Scenes

Shigang Yue[↓] and F. Claire Rind
Ridley Building, School of Biology and Psychology
Faculty of Science, Agriculture and Engineering
University of Newcastle upon Tyne
Newcastle upon Tyne, NE1 7RU United Kingdom

Abstract: *Detecting colliding objects in complex dynamic scenes is a difficult task for conventional computer vision techniques. However, visual processing mechanisms in animals such as insects may provide very simple and effective solutions for detecting colliding objects in complex dynamic scenes. In this paper, we propose a robust collision detecting system, which consists of a lobula giant movement detector (LGMD) based neural network and a translating sensitive neural network (TSNN), to recognise objects on a direct collision course in complex dynamic scenes. The LGMD based neural network is specialized for recognizing looming objects that are on a direct collision course. The TSNN, which fuses the extracted visual motion cues from several whole field direction selective neural networks, is only sensitive to translating movements in the dynamic scenes. The looming cue and translating cue revealed by the two specialized visual motion detectors are fused in the present system via a decision making mechanism. In the system, the LGMD plays a key role in detecting imminent collision; the decision from TSNN becomes useful only when a collision alarm has been issued by the LGMD network. Using driving scenarios as an example, we showed that the bio-inspired system can reliably detect imminent colliding objects in complex driving scenes.*

Keywords: *visual motion; pattern recognition; collision detection; neural network; locust; LGMD; direction selectivity; complex dynamic scene; nature-inspired information processing.*

[↓] Corresponding author:
Dr. Shigang YUE
Ridley Building, School of Biology and Psychology
Faculty of Science, Agriculture and Engineering
University of Newcastle
Newcastle upon Tyne, NE1 7RU
United Kingdom
Tel: (0044) 191 222 5720
Fax: (0044) 191 222 5229
E-Mail: shigang.yue@ncl.ac.uk or shigang.yue@ieee.org

1. Introduction

The ability to detect and avoid collision is very important for animals and mobile intelligent machines. However, many artificial vision systems are not yet able to quickly and cheaply extract the wealth information in the visual scenes [1]. Detecting colliding objects in complex dynamic scenes has been a difficult task for current conventional robotics vision technologies, especially with a limited computing source [2].

The visual collision avoidance systems in insects have evolved over millions of years, and are efficient and reliable in the insect's visual environments. The neural circuits processing visual information in insects are relatively simple compared to those in the human brain and can be a good model for optical sensors for collision detection [3]. The visual processing mechanisms in insects revealed by neurobiologists over the past decades have already begun to provide solutions for collision avoidance or visual based robotic navigation (for example, [4-5], a review is available [6]).

The lobula giant movement detector (LGMD) is a large visual interneuron in the optic lobe of the locust that responds most strongly to approaching objects [7-9]. The functional model of the LGMD's input circuitry showed the same selectivity as the LGMD neuron [10] and has also been used for collision avoidance and detection in a mobile robot and a car [5, 11-12]. The high efficiency in detecting collisions using vision has made it possible to use a LGMD based neural network to detect collision in real time applications. However, in complex driving scenes, the tuned LGMD based neural network also responds briefly to nearby fast translating objects in its visual field [11]. Unfortunately, fast translating visual events can occur in road scenes, for examples, at a roundabout, T-junction or a cross road.

To deal with these fast translating objects, the outputs of four directionally sensitive neurons, based on crossed strips of correlated EMDs (elementary movement detectors), were used to either directly suppress the LGMD's spiking response or to combine with the LGMD output to arrive at a decision [13]. One disadvantage of the direct integration is that the mistakes made by these EMDs can directly affect a correct collision detection by the LGMD. An independent decision on translating events made by an independent specialized neural network could provide a better solution. To increase the chance of detecting translating events, whole field direction selective neural networks (DSNNs) were used to form that specialized translating sensitive neural network (TSNN).

Direction selective neurons have been found in animals for decades, for example, in insects, i.e. locust [14-15], beetles [16] and -flies [17] and in vertebrates, i.e. rabbits [18-21] and cats [22-23]. A recent survey is available [24]. There are many ways to form a computational DSNN (for example, [25-26]). Recent results suggest that asymmetric lateral inhibition ensures robust directional selectivity in the rabbit's retina [21]. In this paper, we use an asymmetric lateral inhibitory mechanism to form the whole-field DSNNs. These DSNNs have a similar network structure to the LGMD neural network but with asymmetric lateral inhibition. A TSNN, as a further level of organisation of these DSNNs, fuses the extracted visual motion cues from the DSNNs so that it only responds to fast translating objects.

In our system, the LGMD and the TSNN make their own decisions based on the visual cues extracted simultaneously and independently. The LGMD plays a key role in detecting imminent collision. The TSNN's decision is based on translating cues and becomes useful only when the LGMD has issued a collision alarm. The system checks the TSNN's decision in order to eliminate a possible false collision alarm resulting from fast translating objects. We demonstrate the system's reliability in detecting dangerous imminent collision by challenging it with driving scenes.

2. Formulation of the system

The system for detecting colliding objects has three main parts: an LGMD based neural network for extracting looming cues in depth, a TSNN for translating cues and a decision making mechanism to fuse the looming and translating cues (Figure 1). Details of the three parts will be given in the following sub-sections.

2.1. LGMD based neural network

The LGMD based neural network (Figure 1 (a)) used in this study is based on the previous neural network described in [5, 10, 11, 27] with minor changes. The LGMD neural network responds to looming cues in depth, however, with the current structure, it also responds to nearby rapidly translating objects [11]. The network is composed of four groups of cells - photoreceptor P, excitatory E, inhibitory I and summing S, and two single cells - feed-forward inhibition (FFI) and LGMD.

P layer The first layer of the neural network are the photoreceptor P cells which are arranged in matrix form; the luminance L_f of each pixel in the input image at frame f is captured by each photoreceptor cell, the change of luminance P_f between frames of the image

sequence is then calculated and forms the output of this layer. The output of a cell in this layer is defined by equation [11, 27]:

$$P_f(x,y) = \sum_i p_i P_{f-i}(x,y) + (L_f(x,y) - L_{f-1}(x,y)) \tag{1}$$

where $P_f(x,y)$ is the change of luminance corresponds pixel (x,y) at frame f, x and y are the pixel coordinates, L_f and L_{f-1} are the luminance, subscript f denotes the current frame and $f-1$ denotes the previous frame, the persistence coefficient p_i is defined by $p_i = (1+e^{\mu i})^{-1}$ and $\mu \in (-\infty, +\infty)$.

I E layer The output of the *P* cells forms the inputs to two separate cell types in the next layer. One type is called the excitatory cells, through which excitation is passed directly to the retinotopical counterpart of the cell in the third layer, the *S* layer. The second cell types are lateral inhibition cells, which pass inhibition, after 1 image frame delay, to their retinotopical counterpart's neighbouring cells in the *S* layer. The strength of inhibition spread to a cell in this layer is given by:

$$I_f(x,y) = \sum_{i=-n}^{n} \sum_{j=-n}^{n} P_{f-1}(x+i, y+j) w_I(i,j), (i \neq j, if\ i = 0) \tag{2}$$

where $I_f(x,y)$ is the inhibition in pixel (x,y) at current frame f; $w_I(i, j)$ are the local inhibition weights; n defines the size of the inhibited area.

S layer The excitatory flow from the *E* cells and inhibition from the *I* cells is summed by the S cells using the following equation:

$$S_f(x,y) = abs(P_f(x,y)) - abs(I_f(x,y)) W_I \tag{3}$$

where W_I is the global inhibition weight. Excitations that exceed a threshold value are able to reach the summation cell LGMD:

$$\widetilde{S}_f(x,y) = \begin{cases} S_f(x,y) & if\ S_f(x,y) \geq T_r \\ 0 & if\ S_f(x,y) < T_r \end{cases} \tag{4}$$

where T_r is the threshold.

LGMD cell The membrane potential of the LGMD cell U_f is the summation of all the excitations in S cells as described by the following equation,

$$U_f = \sum_{x=1}^{k}\sum_{y=1}^{l} \sum\sum abs(\widetilde{S}_f(x,y)) \tag{5}$$

The membrane potential U_f is then transformed to a spiking output using a sigmoid transformation,

$$u_f = (1+e^{-U_f n_{cell}^{-1}})^{-1} \tag{6}$$

where n_{cell} is the total number of the cells in S layer. Since U_f is greater than or equal to zero (as equation (5) is a sum of absolute value), the sigmoid membrane potential u_f varies from 0.5 to 1. The collision alarm is decided by the spiking of cell LGMD. If the membrane potential u_f exceeds the threshold T_s, a spike is produced. A certain number of successive spikes, which is denoted by S^{LGMD}, will trigger the collision alarm in the LGMD cell. However, spikes may be suppressed by the FFI cell when whole field movement occurs [28].

FFI cell In the absence of feed forward inhibition (FFI), the network may produce spikes and a false collision signal when challenged by a sudden change of visual scene, for example during a rapid turn. The feed forward inhibition cell works to cope with such whole field movement when a large number of P cells are activated [10, 28]. The FFI at a given frame is taken from the summed output of the photoreceptor cells with one frame delay,

$$F_f = \sum_{i=1}^{m} \alpha_{f-j}^{F} F_{f-j} + \sum_{x=1}^{k}\sum_{y=1}^{l} abs(P_{f-1}(x,y)) n_{cell}^{-1} \tag{7}$$

where α^{F}_{f-j} is the persistence coefficient for FFI and varies from 0 to 1, and m indicates the maximum number of successive frames the persistence can last. Once F_f exceeds its threshold T_{FFI}, spikes in the LGMD are inhibited immediately.

2.2. The TSNN

The proposed TSNN (Figure 1 (b)) fuses the visual motion cues extracted by the DSNNs. It shares the same photoreceptor P cells with the LGMD network; it has its own excitatory E cells and inhibitory I cells which are similar to those in the LGMD network; it has four groups

of summing cells- *SL*, *SR*, *SU* and *SD* cells, four direction selective cells- *L*, *R*, *U* and *D*, two intermediate cells *a* and *b*, and a spiking cell *TS*. We will take the left inhibitory summing cells *SL* and left inhibitory cell *L* as examples to illustrate the process.

SL layer The inhibition from an *I* cell is passed on to its retinotopical counterpart's neighbouring cells in the next layer. The inhibition is passed, with one image frame delay, asymmetrically from between one to eight cells away. The summed strength of inhibition to a cell in this layer is

$$I_f^L(x,y) = \sum_{i=1}^{m_I} \sum_{j=-n_I}^{n_I} P_{f-1}(x+i, y+j) w_f^L(i,j), (m_I > n_I) \tag{8}$$

where $I^L_f(x,y)$ is the summed inhibition to the *SL* cell and $w^L_I(i,j)$ are the local inhibition weights. In the above equation, inhibition can spread in four directions: up down, left and right, though in an asymmetrical way. The spread to the left is stronger than that to the right since m_I is greater than n_I. At this stage we found that it was not necessary to use all three inhibition directions because the outputs of several direction selective neurons are combined at the next level to extract and then fuse the visual motion cues. To save computing time, we set n_I to 0 (and m_I to 8), so that inhibition has a maximum spread of 8 pixels to the left resulting in directional selectivity with a single nonpreferred direction (leftward in this instance). The local inhibition weights were set to ensure a strong inhibitory effect. With the strong inhibition from the right side, the excitation caused by left translating movements will be reduced or even cancelled. Therefore, the summing cell *L* keeps silent to objects moving to the left but is excited by motion in the other three directions (R, U, and D).

The excitatory flow gathered in an *SL* cell will be

$$S_f^L(x,y) = abs(P_f(x,y)) - abs(I_f^L(x,y))W_I^L \tag{9}$$

where W_I^L is the global inhibition weight.

L cell The excitations in the *SL* cells are summed by the left inhibitory cell *L*. However, to reach the summation cell, excitations should be able to exceed the threshold T_{rL}.

$$\widetilde{S}_f^L(x,y) = \begin{cases} S_f^L(x,y) & if\ S_f^L(x,y) \geq T_{rL} \\ 0 & if\ S_f^L(x,y) < T_{rL} \end{cases} \tag{10a}$$

The membrane potential of the left inhibitory cell L is,

$$U_f^L = \sum_{x=1}^{k}\sum_{y=1}^{l} abs(\widetilde{S}_f^L(x,y)) \tag{10b}$$

The membrane potential of the L cell is then transformed using a sigmoid function,

$$u_f^L = (1+e^{-U_f^L n_{cellL}^{-1}})^{-1} \tag{11}$$

where n_{cellL} is the total number of the cells in SL layer. Since U^L_f is not less than zero according to equation (10b), the membrane potential u^L_f varies sigmoidally from 0.5 to 1.

The membrane potential u^R_f for right inhibitory cell R, u^U_f for up inhibitory cell U and u^D_f for down inhibitory cell D can be obtained in a similar way.

Based on the above structures, the outputs of DSNNs L, R, U and D can be combined to extract visual motion cues. For example, since L, U and D respond to right visual motion with high membrane potential while R does not, rightwards visual motion patterns can be recognised via further processing the outputs of these neurons. In the following sections, further combinations of the four DSNNs will be proposed for the detection of translating objects.

The organization of *L, R, U* and *D.* The four DSNNs: L, R, U and D are further organised to form the TSNN, which fuses the extracted visual cues, as shown in Figure 1 (b). The cell a and b gather information from the four neurons,

$$\begin{pmatrix} U_f^a \\ U_f^b \end{pmatrix} = abs\left(\begin{pmatrix} w_{L-a} & w_{R-a} & w_{U-a} & w_{D-a} \\ w_{L-b} & w_{R-b} & w_{U-b} & w_{D-b} \end{pmatrix} (u_f^L\ u_f^R\ u_f^U\ u_f^D)^T)\right) \tag{12}$$

where U^a_f is the excitation in the a cell; $w_{L\text{-}a}$ is the weight of the connection between the cell L and the cell a. Other symbols are named in a similar way. The absolute operation is applied individually to the components of the vector. The weights are allowed to vary within specified domains for example from -1 to 1 and will be decided in a tuning process. The cells a and b are expected to process excitation from the four DSNNs in different ways. For example, the movement of the left and right edges are often associated strongly and the cues extracted by the left and right DSNNs may be compared through the cells a or b. The different visual

motion cues between left-right and up-down computed by the cells *a* and *b* are then computed again to eliminate other non-translating cues, for example, looming cues.

The spiking cell *TS* gathers information from cells *a* and *b*,

$$U_f^{TS} = abs(\sum(w_{a-d}U_f^a + w_{b-d}U_f^b)) \tag{13}$$

where the weights are adjusted during the network's tuning and can vary from -1 to 1. Once the membrane potential in the cell *TS* exceeds its threshold T_{TS}, a spike is produced. A certain number of successive spikes, denoted by S^{TS}, indicate that a fast translating object has been detected.

2.3. The system's decision making mechanism

The fusion of the collision cue and the translating cue is achieved at the system level. The LGMD plays a key role in extracting the imminent collision cue in the complex dynamic scenes as illustrated in Figure 2. The TSNN and the LGMD make their own decisions simultaneously. However, the TSNN's decision becomes useful only when the LGMD has issued a collision alarm. The system will then check the TSNN's decision to eliminate any possible false collision alarms that may be caused by fast translating objects. The collision alarm will be issued finally, if the TSNN does not produce a threshold activation. The output of the TSNN is only checked if the LGMD has detected a collision.

Since there is no interaction between the LGMD and TSNN until the LGMD has detected a collision, frequent interactions between the two networks have been avoided. Each of the networks can concentrate on its own expertise. These efforts may increase the system's reliability because the LGMD neural network can concentrate on detecting collisions and leave false alarms to be eliminated by the specialized TSNN. Adjusting the LGMD network to be less sensitive to avoid false alarms is difficult as mentioned in [11] and can result in mistakes in classifying collision events.

3. Parameter setting for driving scenes

We use driving scenes to test the proposed colliding objects detection system. The input video images (720 x 576 pixels) provided by Volvo Car Corporation were taken at 25 frames per second and resized to 100 (in horizontal) times 80 (in vertical) pixels by using image

resize function in Matlab[I] to feed the neural networks; images are grey scale ranging from 0 to 255. Samples of the video clips used in the study are shown in Figure 3. These video clips each represented one event, for example, collision with a car, turning, pedestrians, road symbols and high speed translating cars/vans. The car used for the collision was actually an inflatable car for economic reasons. These visual events can occur frequently in driving scenarios and can cause strong excitations in the photoreceptor layer of the two networks (LGMD and TSNN).

The neural networks are simplified by setting all the persistence coefficients to zero. Other predefined parameters are listed in Table 1 and Table 2.1. Since the LGMD based neural network and the TSNN are two different specialized neural networks, it is reasonable to tune each neural network respectively to maximize their specialized capability.

The LGMD neural network The local weights w_I of inhibition spreading from the centre pixel to neighbouring pixels are set to 0.25 for the four nearest neighbours and 0.125 for the four diagonal neighbours [12]. The global inhibition weights were set to 1.7 as shown in Table 1. The inhibition spread one pixel away to its neighbouring cells in next layer, i.e., n=1. The excitations can be sharply reduced or even eliminated by the inhibitions from its surrounding excited neighbours with these inhibition weights.

Since the size of input images is 100 x 80, the number of pixels horizontally is k=100 and vertically l=80. There are 8,000 (n_{cell}=8,000) P, E, I and S cells respectively, 1 LGMD and 1 FFI cells in the LGMD neural network. The P, E and I cells are shared by the LGMD neural network and the TNSS. The threshold T_r is set to 12 based on previous trials. Four successive spikes for S_{LGMD} has been identified as a suitable threshold activation in our previous tuning experiments.

Two crucial parameters of the LGMD neural network, the threshold of the LGMD cell T_s and the threshold of FFI T_{FFI} , were tuned for a better performance of the LGMD neural network. The tuning programme was a genetic algorithm [11] and the tuning process was run three times to pick up the best performing parameter set. The visual events used to tune the parameters were a small group of the video events shown in Figure 3 (including 1, 3, 4, 5, 6, 7, 13, 15, 16, and 18). The results of the tuning for each parameter is listed in Table 1 (right-hand column), together with other predefined parameters (left-hand column). The LGMD

[I] Nearest neighbour interpolation method has been used in resizing input images; further details please refer to 'imresize' function in Matlab® image processing toolbox. Matlab is the trade mark of the Mathworks, Inc.

network did not fail to detect a colliding object, even when the object was a polystyrene block of the approximate size of a pedestrian (not shown). Without the moderating influence of the TSNN the tuned LGMD occasionally signalled a collision when there was not one (Table 3), which is consistent with previous results [11].

TSNN In the TSNN there are 4 groups of summing cells with each group consisting of 8,000 cells (n_{cellL} = 8,000), 4 directionally selective cells, 2 intermediate cells and 1 spiking cell. The prescribed parameters of the TSNN are listed in Table 2.1 (left column). The global inhibition weight was set to 5.5 to ensure a maximum directional inhibitory effect. Different S_{TS} values can affect the overall performance of the TSNN, as shown in Table 2.2. We found that 4 was the optimal number of consecutive spikes from the TSNN (S_{TS} =4) to correctly detect a translation event in our clips (88% success rate[II] without vital failure). The S_{TS} value was thus set to four. Please note that the optimal number of consecutive spikes for the TSNN may not be 4 if the visual environment has been changed dramatically.

The connection weights in the TSNN were also tuned to optimize the TSNN's performance in detecting translating events. The same small group of video events used above in tuning LGMD network was used in tuning the TSNN. The weights vary from -1 to 1; and the threshold of the TSNN varied from 0 to 2.0 during tuning. The tuning process was run three times and the best connection weights are listed in Table 2.1 (right-hand column). The TSNN with such weights achieved an 88% success rate without a vital failure.

4. Test results and discussions

Before testing the whole collision detection system, we checked the responses of the asymmetric lateral inhibition based DSNNs when challenged with a right moving black bar and a left walking pedestrian (Figure 4 (a)). We then compared these responses with those of an Elentary Mototion Detector (EMD) based DSNN to the same stimuli (Figure 4 (b) and (c)). The EMD based DSNN had been used in a previous study [13]. The asymmetric lateral inhibition based DSNNs can distinguish the translating cue clearly in pedestrian sequences and in moving bar sequences when the translating speed (i.e., angular velocity to the camera) of the bar was 54 degrees per second (°/s) and 113°/s (Figure 4 (a)); however, the difference was not sensed clearly when translating speed was 225°/s (which is quite rare in road scenes) because this speed was beyond the range of lateral inhibition (8 pixels per frame). The EMD based DSNNs (with 10 EMD units in each DSNN) can sense the left-right cue when

[II] Success rate = ((total test events – number of failures in detection)/ total test events) 100% = ((24-3)/24) 100% ≈ 88%

translation occurs at 113°/s, however, the sudden bursts from the EMDU (up) and EMDD (down) networks at each speed may cause confusion (Figure 4 (b)). When the resolution was reduced from 10 EMD units to 3 EMD units in each DSNN, the performance was improved to motion at both 54°/s and 113°/s. To the pedestrian scenario however, the direction cue was not clear without reducing the membrane potential scale to 1/10 of its original (Figure 4 (c)). This comparison suggests that the asymmetric lateral inhibition based DSNNs are very robust without re-tuning and can be the basis for the TSNN.

We then used the driving scenes to test the proposed system for the detection of colliding objects. Samples of the 24 video clips used in the test are shown in Figure 3. These events occur frequently in driving scenarios and can cause strong excitations in the photoreceptor layer of the two networks (LGMD and TSNN). However, these translating events were not collision events and should not be detected as collisions by the system.

Details of the processed images in the LGMD and the four direction selective cells, the DSNNs' membrane potentials, the LGMD and TSNN's membrane potential and their spiking responses to four example sequences: a collision, a fast translating car/van and a pedestrian, are shown in Figure 5, Figure 6 and Figure 7 respectively. As shown in the images in Figure 5, the asymmetric inhibition resulted in different excitatory maps in the summing layers. Excitations are indicated with white. These extracted differences form the functional base of the TSNN. It should be noted that the membrane potentials (representing different visual cues) of the four DSNNs were fused in a complex way and each network did not contribute equally to the TSNN (Table 2.1 and Figure 6). As shown in detail in Figure 7, the TSNN responded to fast translating movement with trains of spikes, often before the LGMD network produced a false alarm (Figure 7 (b~d)); the LGMD and the TSNN responded quite differently to a colliding car (Figure 7 (a)).

Results of colliding and translating objects detected by the LGMD, the TSNN and the combined detection system (LGMD cooperating with TSNN) respectively have been listed in Table 3. The LGMD based neural network detected all of the collision events, however, issued false alarms to some of the translating events, for example, to sequences no.4 and no.7. The TSNN detected most of the fast translating events and did not respond to collision events. The translating objects that the TSNN ignored were relatively small and slow (sequence no.11 and no.14), and did not trigger LGMD spikes either. Working together, the false decisions issued by the LGMD were all cancelled because the decisions from the TSNN showed they were translating events.

The TSNN responds to a wide range of translating speeds. For example, the speed of the pedestrian shown in Figure 3 sequence no.10, was rather slow and the speed of the translating car shown in Figure 3 sequence 7 was fast and close to the camera. The TSNN detected them both (Table 3). This was because the DSNNs, which have formed the TSNN, worked over a wide range of translating speeds. This robustness was due to the wide spread of laterally directed inhibition in this system (Figure 4 and Figure 5).

The LGMD neural network and the DSNNs have biological counterparts in locusts [7~10, 14,15]. However, no biological evidence has yet been found showing direct inhibitory connections between neurons of these two types in the brain. This may be because direct inhibition may occur in the thoracic nervous system at the level of the premotor interneurons or at the motor neurons themselves (Rowell, 1989). Several identified premotor interneurons, such as the "C" and "M" interneurons, involved in co-ordinating an escape jump, receive strong excitation from the LGMD pathway; and these premotor interneurons then either excite ("C") or inhibit ("M") motor neurons in the leg (Pearson et al., 1980; Gynther and Pearson, 1989). In biological visual systems, it is widely accepted that visual cues are processed by different specialized visual neurons in very complex spatial-temporal ways. Therefore, it is extremely interesting to investigate how the different visual cues can be integrated and contribute to the animal's overall behaviour. On the other hand, using modelling methods to explore the possible fusion mechanisms of multiple visual neural networks for a real world application is also important.

In the above tests, we showed that different visual motion cues can be extracted with specialised neural networks and fused again for special purpose- road collision detection. The above tests have demonstrated the reliability of the system in detecting colliding objects amongst complex situations in the dynamic scenes. However, this does not mean it can detect all types of colliding objects encountered in driving. For example, it has not been trained and tested with real pedestrians on a direct collision course (for it is difficult to record such scenes in the real world), and it probably does not work when it faces a small or a very thin colliding object. We are extremely interested in using this system for pedestrian collision detection. Early efforts towards this goal included: a pedestrian-shaped polystyrene block was filmed in collision with an approaching car (as in Figure 3, scenarios 1-3) and used to challenge the combined LGMD TSNN network. The LGMD responded to the approaching pedestrian, giving a collision warning but the TSNN network was also activated and suppressed the LGMD response. When the LGMD threshold T_s was dropped by 0.01 to 0.9795 this suppression was prevented.

The LGMD and direction selective neurons are only small parts of the visual system in insects. Many new types of specialized interneurons have been identified in insects' visual pathway during the last several decades. For example, a group of direction selective neurons responsible for small target movement detection (STMD) has been reported recently in [32]. In the future, other specialized neural networks will be integrated into the system to extract other visual cues. Methods for integrating and training the complex visual systems consist of many specialized neural networks which will also be investigated systematically.

5. Conclusions

In the above sections, we proposed a collision detection system which consists of two specialised neural networks to extract and fuse different visual cues- the LGMD based neural network responding to impending objects in depth and the TSNN responding to fast translating visual movement. With the decision making mechanism to integrate the two neural networks together, the collision detection system works reliably without false alarms as demonstrated by challenging it with driving scenarios. This study suggests that a more reliable and robust response to complex visual events may be achieved by integrating specialized neural networks together.

In the future, other specialized neural networks will be integrated into the system and methods for training the complex visual systems will also be investigated systematically.

Acknowledgement

This work is supported by EU IST-2001-38097. We thank M. Soininen of Volvo Car Corporation for providing the video clips used in this paper, Dr. R. Stafford for his comments in internal review and Mr. M. Bendall for proof reading this paper. We thank the anonymous reviewers for their invaluable comments.

References

[1] G. Indiveri, R. Douglas, Neuromorphic vision sensors, Science, 288, (2000)1189-1190.

[2] G.N. DeSouza, A.C. Kak, Vision for mobile robot navigation: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(2), (2002) 237-67.

[3] F.C. Rind, Roger D. Santer, J. Mark Blanchard and Paul F.M.J. Verschure, Locust's looming detectors for robot sensors, Sensors and Sensing in Biology and Engineering, FG Barth, JAC Humphrey, and TW Secomb (Eds.), Spinger-Verlag, Wien, New York, 2003.

[4] R.R. Harrison, C. Koch, A silicon implementation of the fly's optomotor control system, Neural Computation, 12, (2000) 2291-2304.

[5] M. Blanchard, F.C. Rind and P.F.M.J. Verschure, Collision avoidance using a model of the locust LGMD neuron, Robotics and Automonous Systems, 30, (2000)17-38.

[6] F.C. Rind, Bioinspired sensors: from insect eyes to robot vision, Frontiers in Neuroscience: Methods in Insect Sensory Neuroscience, T.A. Christensen (ed) CRC Press, Boca Raton. Chpt 8 pp.213-238, 2005.

[7] M. O'Shea, C.H.F. Rowell, J.L.D. Williams, The anatomy of a locust visual interneurone: The descending contralateral movement detector, Journal of Experimental Biology, 60, (1974)1–12.

[8] C.H.F. Rowell, M. O'Shea, J.L. Williams, The neuronal basis of a sensory analyser, the acridid movement detector system .IV. The preference for small field stimuli, J. Experimental Biology, 68, (1977) 157-185.

[9] F.C. Rind, P.J. Simmons, Orthopteran DCMD neuron: A reevaluation of responses to moving objects. I. Selective responses to approaching objects, Journal of Neurophysiology, 68, (1992) 1654–1666.

[10] F.C. Rind, D.I. Bramwell, Neural network based on the input organization of an identified neuron signaling impending collision, Journal of Neurophysiology, 75, (1996) 967– 985.

[11] S. Yue, F.C. Rind, M.S. Keil, J. Cuadri and R. Stafford, A bio-inspired visual collision detection mechanism for cars: optimisation of a model of a locust neuron to a novel environment, Neurocomputing, 2004 (online 26 October 2005, doi:10.1016/j.neucom.2005.06.017).

[12] S. Yue, F.C. Rind, A Collision detection system for a mobile robot inspired by locust visual system, IEEE Int. Conf. on Robotics and Autom., Spain, Barcelona, 2005, pp.3843-3848.

[13] R. Stafford, R.D. Santer and F.C. Rind, A bio-inspired visual collision detection mechanism for cars: combining insect inspired neurons to create a robust system, Bio Systems: Sixth International Workshop on Information Processing in Cells and Tissues (IPCT2005) (in press).

[14] F.C. Rind, A directionally selective motion-detecting neurone in the brain of the locust: physiological and morphological characterization, J. Exp. Biol., 149, (1990) 1-19.

[15] F.C. Rind, Identification of directionally selective motion-detecting neurones in the locust lobula and their synaptic connections with an identified descending neurone, J. Exp. Biol., 149, (1990) 21-43.

[16] B. Hassenstein, W. Reichardt, Systemtheorische analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus, Zeitschrift für Naturforschung, 11b, (1956) 513–524.

[17] A. Borst, J. Haag, Neural networks in the cockpit of the fly, J. Comp. Physiology, 188, (2002) 419-437.

[18] H. B. Barlow, R. M. Hill, Selective sensitivity to direction of movement in ganglion cells of rabbit retina, Science, 139, (1963) 412–414.

[19] H.B. Barlow, W.R. Levick, Mechanism of directionally selective units in rabbits retina, J. Physiol. (Lond.), 178, (1965) 477-504.

[20] S.F. Stasheff, R.H. Masland, Functional inhibition in direction-selective retinal ganglion cells: spatiotemporal extent and intralaminar interactions, J. Neurophysiology, 88, (2002) 1026-1039.

[21] S.I. Fried, T.A. Muench, and F.S. Werblin, Mechanisms and circuitry underlying directional selectivity in the retina, Nature, 420, (2002) 411-414.

[22] N.J. Priebe. D. Ferster, Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex, Neuron, 45, (2005) 133-145.

[23] M.S. Livingstone, Direction inhibition: a new slant on an old question, Neuron, 45, (20005) 5-7.

[24] D.I. Vaney, W.R. Taylor, Direction selectivity in the retina, Current Opinion in Neurobiology, 12, (2002) 405-410.

[25] J. A. Marshall, Self-organizing neural networks for perception of visual motion, Neural Networks, 3, (1990) 45-74.

[26] T.Tversky, R. Miikkulainen, Modeling direction selectivity using self-organizing delay-adaptation maps, Neurocomputing, 44-46, (2002) 679-684.

[27] F.C. Rind, R. Stafford and S. Yue, LOCUST Technical Report D11: Biological Model Report, Project EU IST-2001-38097, LOCUST: Life-Like Object Detection for Collision-Avoidance Using Spatio-Temporal Image Processing, February 2004.

[28] R. D. Santer, R. Stafford and F. C. Rind, Retinally-generated saccadic suppression of a locust looming detector neuron: investigations using a robot locust, Journal of the Royal Society , London, Interface, 1, (2004) 61-77.

[29] C.H.F. Rowell, Descending interneurones of the locust reporting deviation from flight course:what is their role in steering. Journal of Experimental Biology, 146, (1989) 177-194.

[30] I.C. Gynther, and K.G. Pearson, An evalution of the role of identified interneurons in triggering kicks and jumps in the locust. Journal of Neurophysiology, 61, (1989) 45 - 57.

[31] K.G. Pearson, W.J. Heitler, and J.D. Steeves. 1980. Triggering of locust jump by multimodal inhibitory interneurons. Journal of Neurophysiology, 43, (1980) 257 - 278.

[32] K. Nordström, P.D. Barnett and D. C. O'Carroll, Insect detection of small targets moving in visual clutter, PLoS Biology, 4(3): e54, 2006.

Table 1. The parameters of the LGMD based neural network

name	value	name	value
p_i	0	T_{FFI}	35.8798
μ	0	T_s	0.9895
W_I	1.7		
T_r	12		
n_{cell}	8,000		
k	80		
$\boldsymbol{l}$	100		
w_I	0.125~0.25		
$\alpha_{f\text{-}1}^{F}$	0		
n	1		
m	1		
S^{LGMD}	4		

Table 2.1 The parameters of the TSNN

name	value	name	value
p_i	0	$w_{L\text{-}a}$	0.8519
μ	0	$w_{R\text{-}a}$	-0.5127
W^L_I	1.7	$w_{U\text{-}a}$	-0.3905
T_{rL}	12	$w_{D\text{-}a}$	0.3905
n_{cellL}	8,000	$w_{L\text{-}b}$	0.1334
k	80	$w_{R\text{-}b}$	-0.2273
$\boldsymbol{l}$	100	$w_{U\text{-}b}$	1.3993
w^L_I	5.5	$w_{D\text{-}b}$	-0.5743
$\alpha_{f\text{-}1}^{F}$	0	$w_{a\text{-}TS}$	0.7336
n	1	$w_{b\text{-}TS}$	-1.4697
m	1	T_{TS}	0.4996
S^{TS}	4		

Table 2.2 The effect of different S_{TS} value on overall performance of TSNN

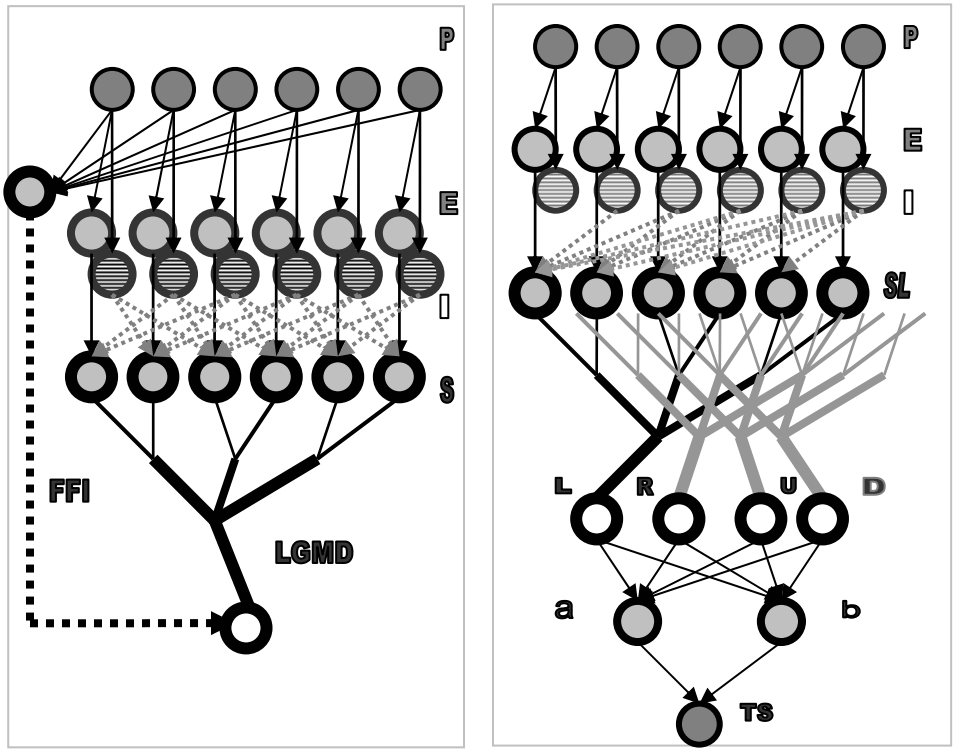
S_{TS} value (successive spikes)	*Number of failures*	*Number of vital failures (in collision situations)*	Success rate
1	5	*3*	79% X
2	4	*2*	83% X
3	4	*1*	83% X
4	3	*0*	88%
5	4	*0*	83%
6	4	*0*	83%
7	4	*0*	83%
15	9	*0*	63%

Note: X represents at least one vital failure has happened; vital failure means failure in collision events. The total number of visual events used in this test is 24.

Table 3. The performance of the LGMD neural network, the TSNN and the system when challenged with video sequences in a driving scene database sampled in Figure 3.

Sequences	LGMD	TSNN	system	Sequences	LGMD	TSNN	system
1*	√	√	√	13	√	√	√
2*	√	√	√	14	√	√	√
3*	√	√	√	15	√	√	√
4	×	√	√	16	√	×	√
5	×	√	√	17	×	√	√
6	√	√	√	18	×	√	√
7	×	√	√	19	√	√	√
8	√	√	√	20	√	√	√
9	√	√	√	21	√	√	√
10	√	√	√	22	√	√	√
11	√	×	√	23	√	√	√
12	√	√	√	24	√	×	√
Success rate					79%	88%	100%

Note: * 1~3 are collision sequences. 4~24 are all non-collision sequences. For the LGMD and the system, √ means responds correctly to collision and non-collision events; for the TSNN, √ means respond to translating and non-translating events correctly. × means respond to the events incorrectly.



(a). the LGMD based neural network (b). the TSNN

Figure 1. The schematic illustration of the LGMD based neural network and the asymmetric lateral inhibition based TSNN for colliding objects detection. (a), The LGMD based neural network. There are four groups of cells and two single cells: photoreceptor cells (P); excitatory and inhibitory cells (E and I); summing cells (S); a LGMD cell and a feed forward inhibition cell (FFI). The output of P cells is the luminance change. The lateral inhibition is indicated with thin grey lines and has one frame delay. The lateral inhibition in I layer spreads to its neighbouring cells in S layer without preferred direction. The excitation is indicated with black lines and has no delay. The FFI also has one frame delay. The input to FFI is luminance change from photoreceptor cells. (b), The asymmetric lateral inhibition based TSNN. There are seven groups of cells and seven single cells: photoreceptor cells (P) shared with LGMD network; excitatory cells (E) and inhibitory cells (I); left inhibitory summing cells (SL), which only sums the excitation and the left spreading inhibition; similarly SR, SU and SD cells sums excitation and right, up and downwards spreading lateral inhibitions respectively; the L, R, U and D cell; the intermediate a and b cells; the spiking translating sensitive cell TS. The spread inhibitions are in dotted lines. Only the left inhibited cell L (highlighted) and its network are illustrated as the example of the four direction selective cells.

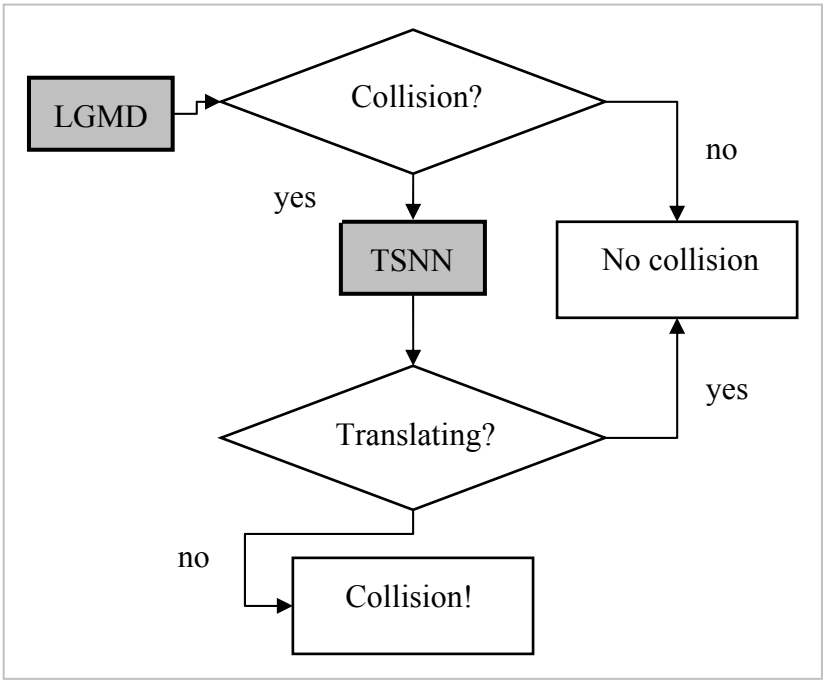
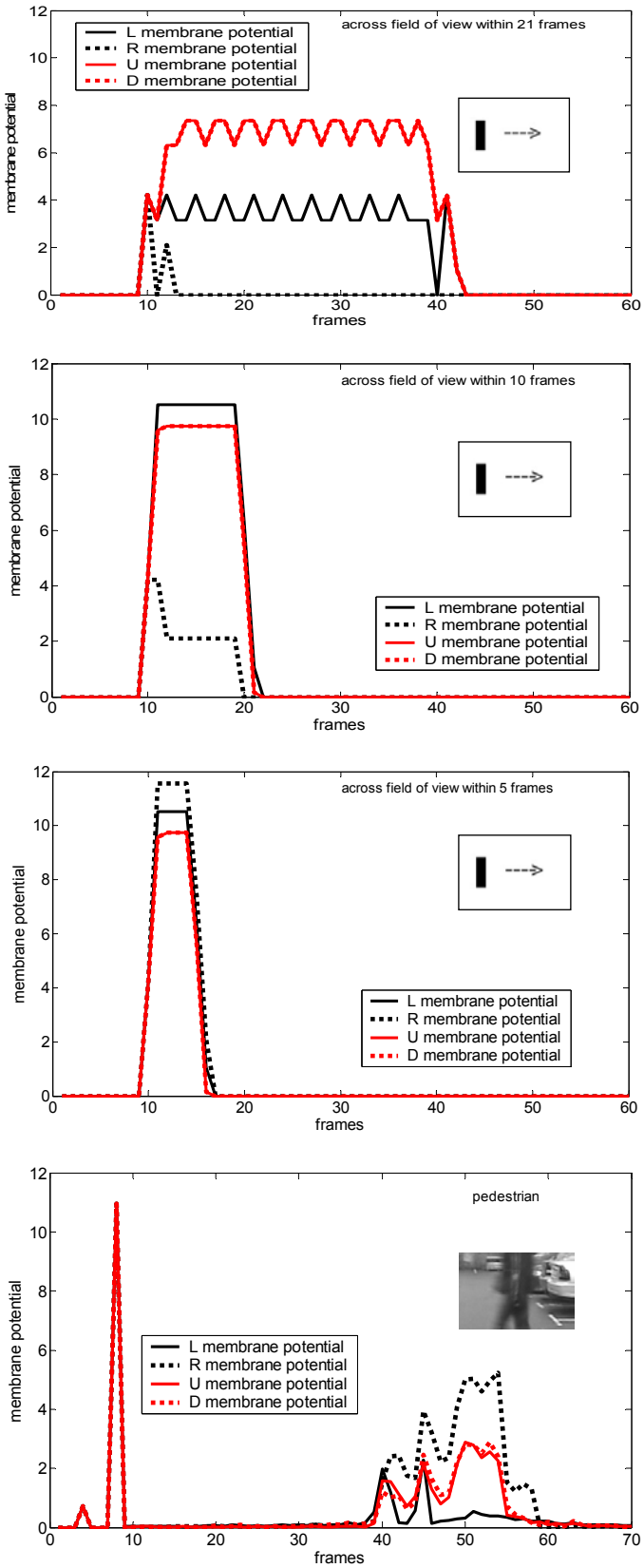
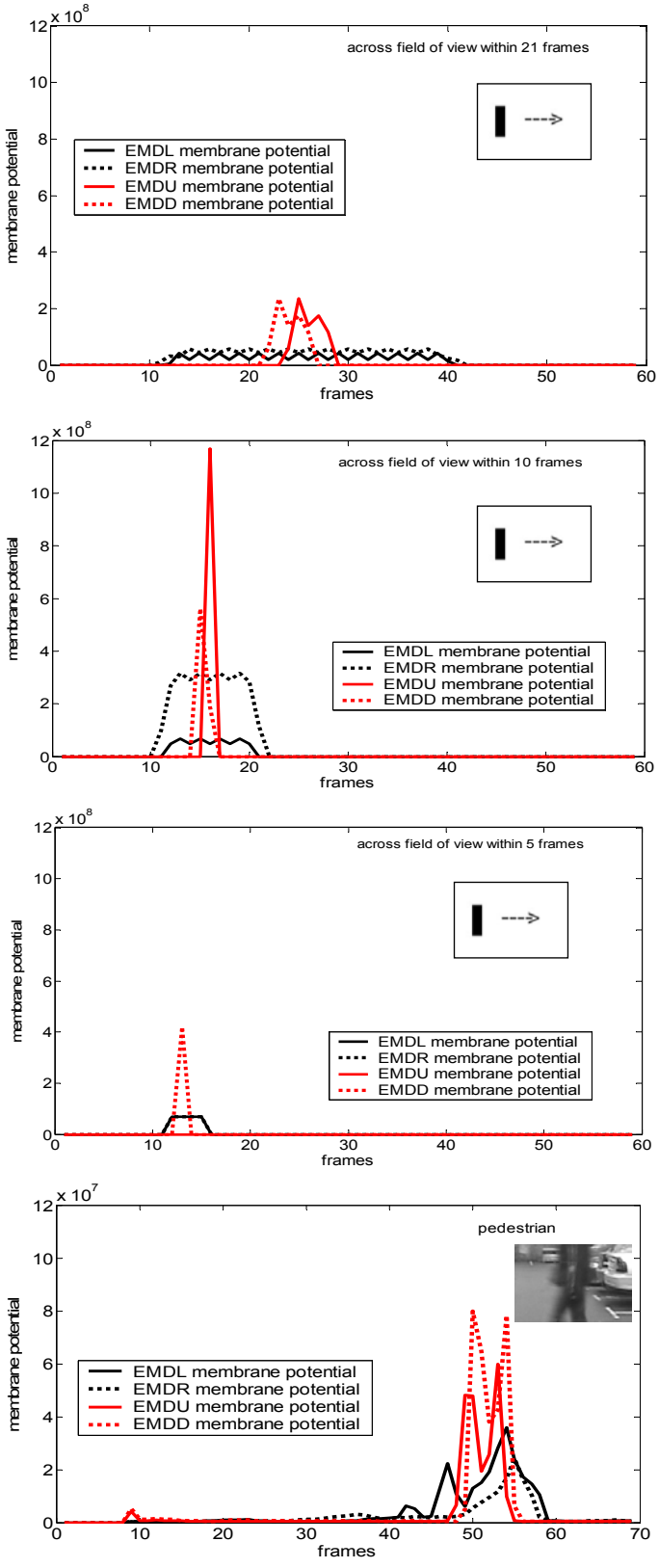


Figure 2. The integration of the LGMD and the TSNN in the system. The extracted visual motion cues are further fused in system level via a decision making mechanism.

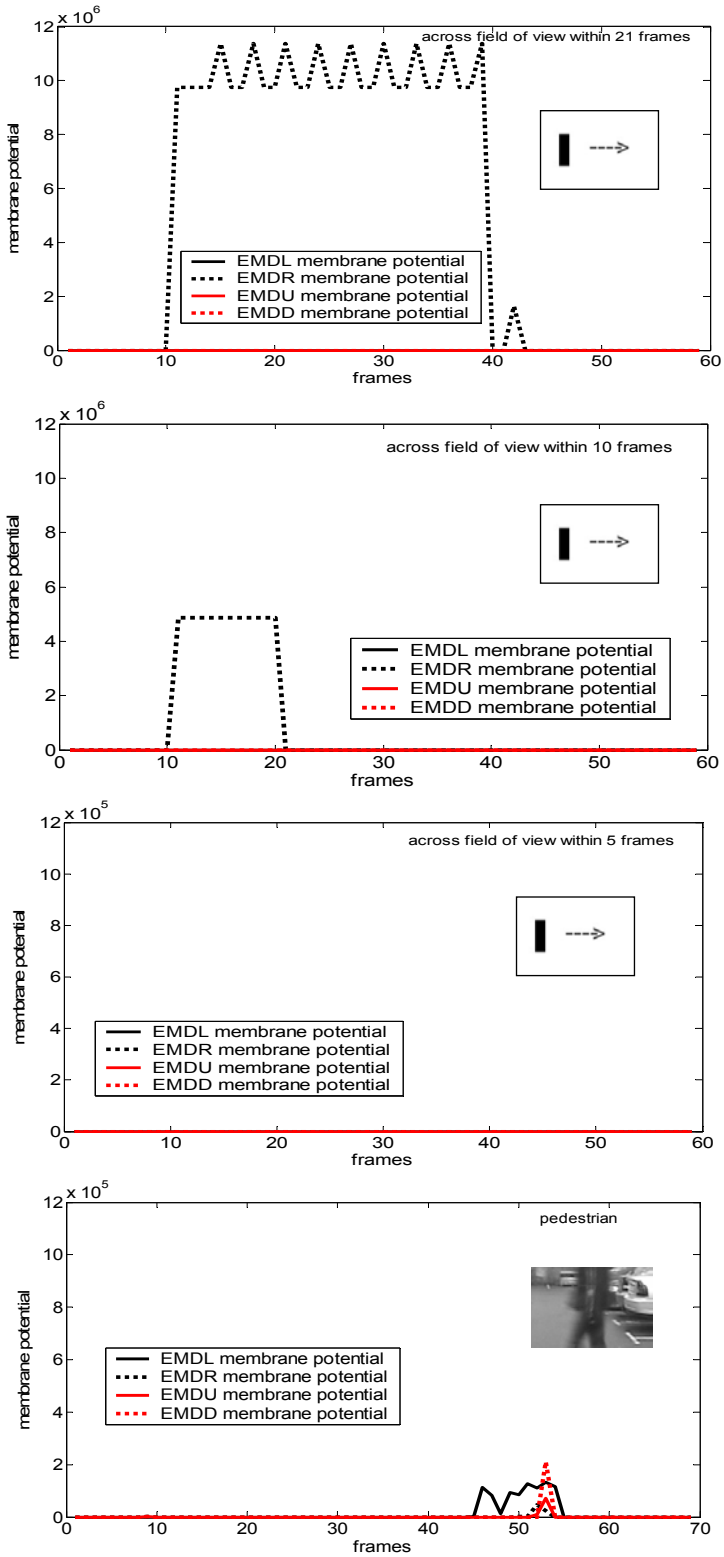
Figure 3. Samples of driving scenes for testing the collision detection system. 1~3 collision events, collision with car at three speed levels: relatively slow, mediate and fast respectively; 4~5 translating events, fast translating figures, while driving at low speed; 6~7 translating events, fast translating cars; 8~9 translating events, turning and translating cars, normal driving speed; 10~11 translating events, translating human figures; 12~13 translating events, car cutting in; 14 approaching but not collision event, driving after a truck; 15~16 translating events, road symbols, driving at high speed; 17~18 translating events, fast translating car/van; 19 translating man with a trolley; 20 translating event, turning; 21 driving into, in and out of tunnel; 22~23 translating events at night, pedestrians; 24 approaching but not collision event at night.



(a). The asymmetric lateral inhibition based DSNNs.



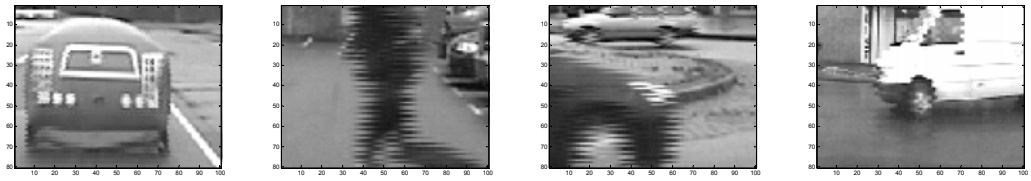
(b). The EMD based DSNNs with 10 EMDs in each DSNN. Note the scale of the last plot is 1/10 of the other three.



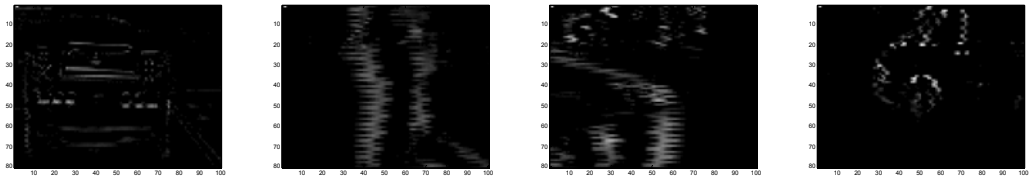
(c). The EMD based DSNNs with 3 EMDs in each DSNN. Note the scale of the last plot is 1/10 of the first three.

Figure 4. The pre-sigmoid membrane potential of the DSNNs challenged with different visual sequences. The computer generated black bar moved from left to right at different speeds, i.e., crossing the whole field of view in 21 frames, 10 frames and 5 frames respectively which corresponded to 54°/s, 113°/s and 225°/s *with a 45° field of view and an image acquisition rate of 25 frames per second. The pedestrian took 17 frames to move over the field of view. The image showed in the plot was taken at frame 50.*

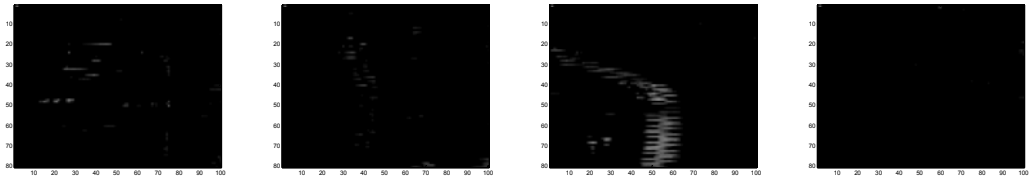
Original images:



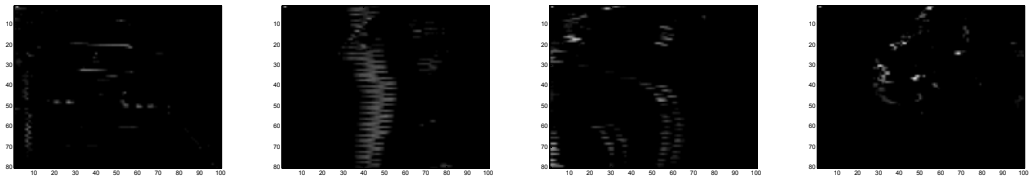
In S layer of the LGMD:



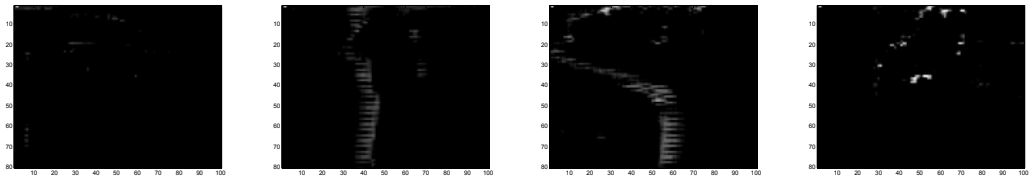
In SL layer of TSNN:



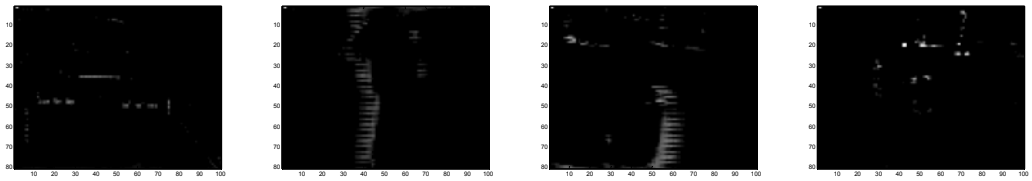
In SR layer of TSNN:



In SU layer of TSNN:

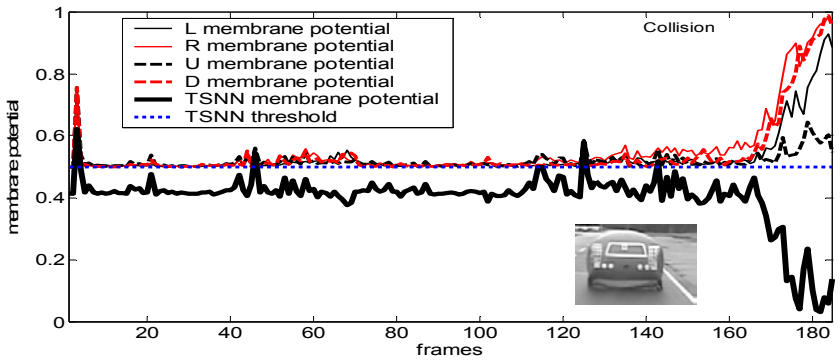


In SD layer of TSNN:

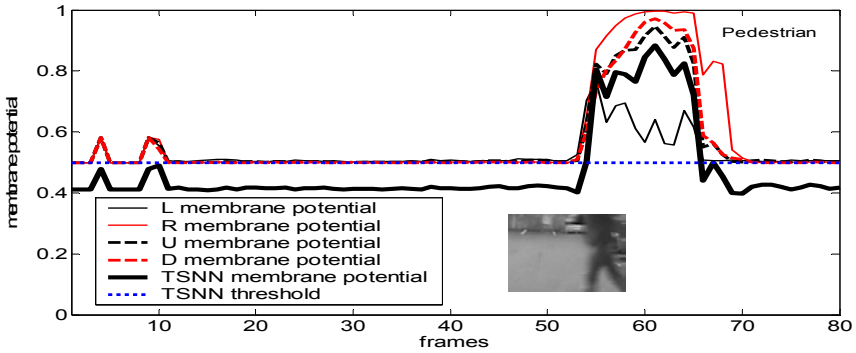


(a).sequence no.1 (b). sequence no.4 (c).sequence no.7 (d). sequence no.17

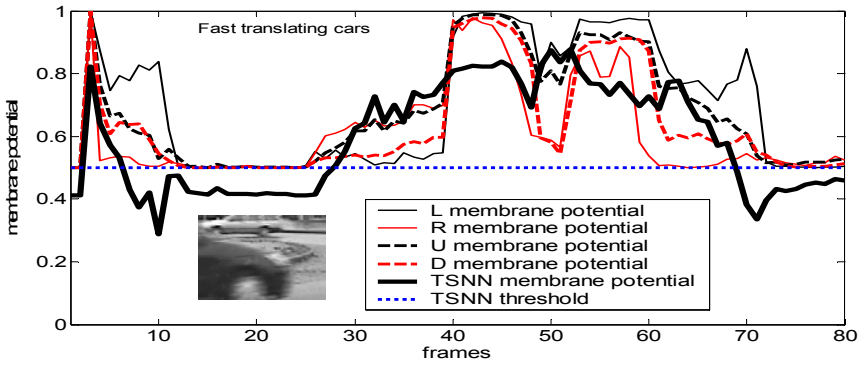
Figure 5. The images and excitation maps in the S layer of the LGMD network, in SL, SR, SU and SD layers of the DSNNs. The white parts of the excitation maps indicate excitation caused by the visual events. The differences in excitation are obvious in response to the same input images. Columns: (a) Collision scene from video sequence no.1; (b) pedestrian scene from video sequence no.4; (c) fast translating car from video sequence no.7; (d) translating van from video sequence no.17.



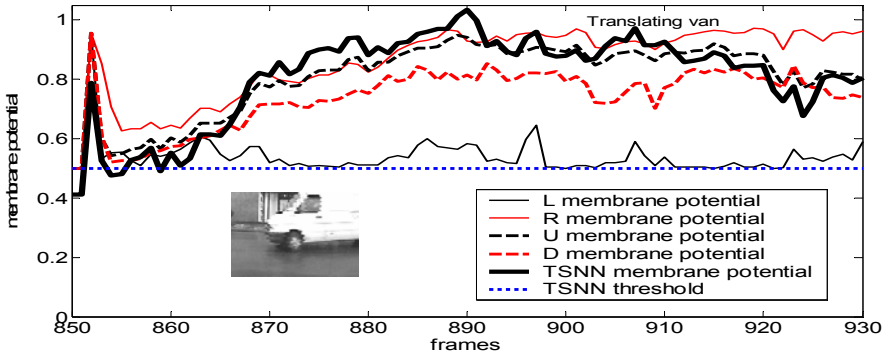
(a), processing video sequence no.1 (snapshot taken at frame 170)



(b), processing video sequence no.4 (snapshot taken at frame 58)

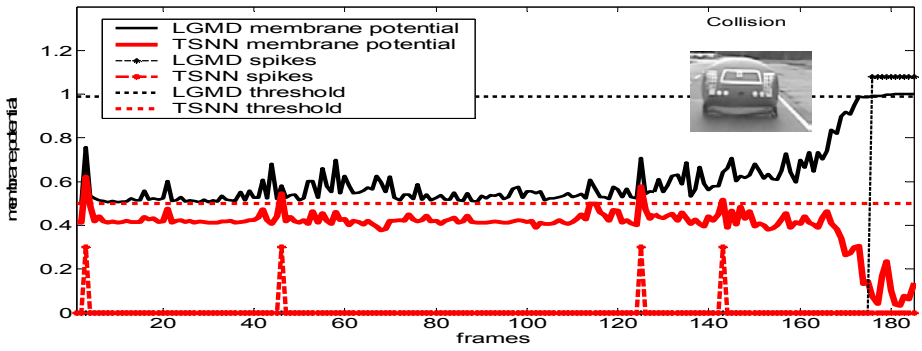


(c), processing video sequence no.7 (snapshot taken at frame 44)

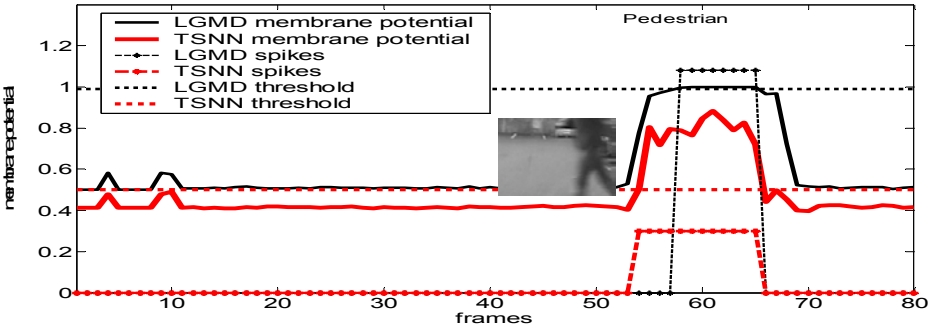


(d), processing video sequence no.17 (snapshot taken at frame 880)

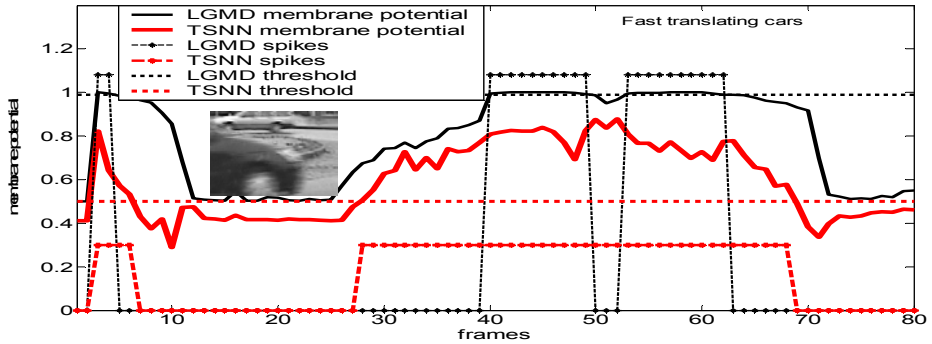
Figure 6. The membrane potential of the L, R, U, D and the TSNN in processing some of the video sequences. The membrane potentials of the four DSNNs are fused in an unequal way into the TSNN (Table 2). The frame numbers of the images showed in the plot are indicated in the subtitles in brackets.



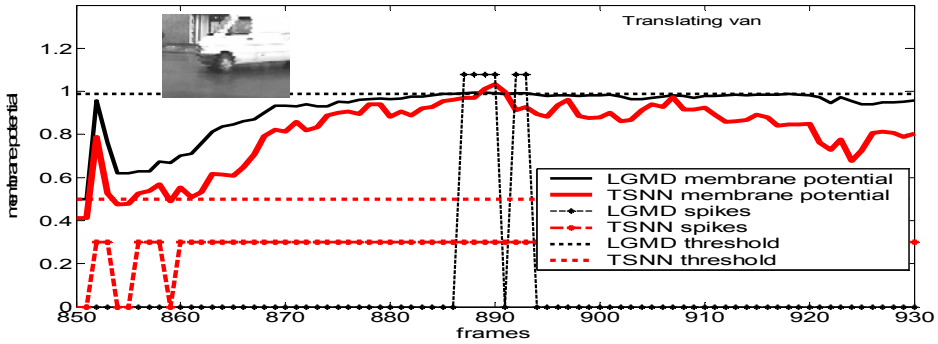
(a). processing video sequence no.1 (snapshot taken at frame 170)



(b). processing video sequence no.4 (snapshot taken at frame 58)



(c). processing video sequence no.7 (snapshot taken at frame 44)



(d). processing video sequence no.17 (snapshot taken at frame 880)

Figure 7. The membrane potential of the LGMD and TSNN while processing some of the video sequences. The translating movements have been detected by the TSNN. (a) Processing video sequence no.1; (b) processing video sequence no.4; (c) processing video sequence no.7; (d) processing video sequence no.17. The frame number of the image showed in the plot is indicated in the subtitles in brackets.