时域感兴趣区域精确定位与膜电位多核调整的动态视觉传感器数据分类

黄庆坤, 陈云华†, 张 灵, 兰浩鑫

(广东工业大学 计算机学院, 广东 广州 510006)

摘要: 动态视觉传感器(DVS)因其在获取视觉信息时具有低功耗, 低延迟等特性, 本质上十分适用于便携式设备上的实时动作识别. 在对DVS事件流时域感兴趣区域(ROI)进行定位与分割时, 现有方法往往不能根据不同物体运动自适应地设定最佳检测阈值、无法对静态场景中少量背景噪声进行过滤, 为此, 提出基于LIF神经元模型和脉冲最大值监测单元的运动符号检测(MSD), 以实现在多种不同物体运动下事件流时域ROI关键时间点的自适应精确定位; 在对分类器进行训练时, 对不同的脉冲输入模式, 使用不同的核函数调整突触后神经元膜电位, 使训练得到的突触权重朝着正确发放的方向改变, 提出一种具有抗噪性的脉冲神经网络学习算法MK–Tempotron. 实验结果表明, 与同类方法相比, 本文方法在DVS数据集上的识别精度能获得高达14.61%的提升.

关键词: 动态视觉传感器DVS; DVS数据分类; 目标识别; 时域感兴趣区域ROI; 神经网络; MK–Tempotron

引用格式: 黄庆坤, 陈云华, 张灵, 等. 时域感兴趣区域精确定位与膜电位多核调整的动态视觉传感器数据分类. 控制理论与应用, 2020, 37(8): 1837 – 1845

DOI: 10.7641/CTA.2020.91001

Dynamic vision sensors data classification based on precise temporal region of interest locating and multi-kernel membrane potential adjusting

HUANG Qing-kun, CHEN Yun-hua†, ZHANG Ling, LAN Hao-xin

(College of Computer Science, Guangdong University of Technology, Guangzhou Guangdong 510006, China)

Abstract: Dynamic vision sensors (DVS), due to their low power consumption and low latency when acquiring visual information, are essentially suitable for real-time motion recognition on portable devices. When locating and segmenting the temporal region of interest (ROI) of the DVS event stream, the existing methods often cannot adaptively set the optimal detection threshold according to the motion of different objects, and they cannot eliminate a small amount of background noise in a static scene either. To solve these problems, a motion symbol detection (MSD) method based on the leaky integrate-and-fire (LIF) neuron model and a peak spiking monitoring unit is proposed, to precisely locate the critical time point for the temporal ROI in the event stream containing a variety of different moving objects. And an anti-noise learning algorithm based on the tempotron rule, which is named as MK–Tempotron, is proposed to deal with background noise in static scenes. In MK–Tempotron, different kernels are applied according to different input spiking patterns to adjust the post-synaptic membrane potential of the neuron during training, so that the synaptic weights can be changed in the direction of correct firing of spikes, thus the anti-noise performance of the classification algorithm is improved. Experimental results show that compared with some similar methods, the recognition accuracy of the proposed method on several DVS datasets can be improved by as much as 14.61%.

Key words: dynamic vision sensors; DVS data classification; objects recognition; temporal region of interest; neural networks; MK–Tempotron

Citation: HUANG Qingkun, CHEN Yunhua, ZHANG Ling, et al. Dynamic vision sensors data classification based on precise temporal region of interest locating and multi-kernel membrane potential adjusting. *Control Theory & Applications*, 2020, 37(8): 1837 – 1845

收稿日期: 2019−12−11; 录用日期: 2020−04−19.

†通信作者. E-mail: yhchen@gdut.edu.cn; Tel.: +86 13660188815.

本文责任编委: 苏剑波.

广东省自然科学基金项目(2016A030313713), 广东省交通运输厅科技项目(科技–2016–02–030)资助.

Supported by the Natural Science Foundation of Guangdong Province, China (No.2016A030313713) and the Science and Technology Project of the Department of Transportation of Guangdong Province (Science and Technology–2016–02–030).

1 引言

动态视觉传感器(dynamic vision sensors, DVS)[1], 在智能机器人、无人驾驶等领域有着广阔的应用前景. 与传统的视觉传感器相比, DVS丢弃帧和曝光时间的概念, 通过监测每个像素点的光强变化输出事件流, 从而解决了传统视觉传感器所带来的数据冗余问题. 其十分类似于人类视网膜功能, 在获取视觉信息时具有低功耗、低延迟等卓越特性, 因此本质上十分适合用于便携式设备上的实时动作识别任务.

由于DVS是一种基于事件的传感器, 单个独立事件是无意义的. 目前大多数的方法是将事件流分割为多个片段后, 再进行特征的提取和分类, 因此, 如何定位与分割事件流的时域感兴趣区域(region of interest, ROI), 对特征提取和分类效果的影响至关重要. 目前分割事件流时域ROI的方法主要分为两大类, 即硬事件分割(hard events segmentation, HES)和软事件分割(soft events segmentation, SES).

在HES方法中, Anna Baby等人[2]是用一个固定的时间窗口, 把整个事件流分割成时间大小相等的若干虚拟帧. 但是由于DVS的工作机制, 移动速度越快的物体所产生的激活像素越多, 因此在固定的时间窗口下, 所捕获物体的形状将取决于它的运动速度. 而Ghosh等人[3]是用一个事件数量固定的动态时间窗口对事件流进行分割, 此方法在很大程度上消除运动速度对物体形状的影响. 但对于多种不同物体运动的场景, 存在着不同物体的最佳检测阈值差异较大的问题. 同时, 上述两种HES方法, 认为每个输出事件(即使其为噪声事件)都具有同等的重要性, 因此抗噪性能较差.

与HES不同的是, SES可根据事件的输出特性自适应地分割事件流片段. 因此, SES比HES更能准确地对时域ROI进行定位与分割. Peng等人[4]利用LIF(leaky integrate-and-fire)神经元模型[5]的阈值响应机制进行运动符号检测(motion symbol detection, MSD), 以实现SES. 由于LIF神经元的泄漏机制以及连续的增量集成, 能够有效降低噪声事件的干扰. 但由于LIF神经元模型采用硬阈值, 因此该方法在处理不同物体运动时, 同样会存在不同物体的最佳检测阈值差异较大的问题.

为了解决上述问题, 对此, 本文提出一种基于LIF神经元模型和脉冲最大值监测单元的MSD, 以实现对不同物体运动所产生的事件流时域ROI关键时间点的自适应定位, 从而解决分割片段受不同物体的最佳检测阈值差异较大以及背景噪声事件影响的问题.

在已有的一些对事件流进行表征的方法中, 是直接基于卷积神经网络(convolutional neural networks, CNN)进行权重训练, 然后将训练好的CNN转换成脉冲神经网络(spiking neuron networks, SNN), 以实现对DVS数据的特征表示与识别[6–11]. 该类方法需要对网络中的大量参数进行优化, 才能获得较高的识别率. 为了保证DVS数据处理的低功耗、低延迟性, 本文首先基于所提出的MSD进行事件流自适应定位与分割, 然后基于Gabor滤波器进行空域特征提取, 最后再采用直接训练得到的SNN来实现DVS数据的分类.

在SNN学习算法中, Tempotron[12]学习算法由于只需标记发放状态, 而不需要标记发放时间, 更适用于真实环境刺激下的分类任务. 但由于Tempotron学习算法在接收到脉冲并使神经元突触后膜电位(post-synaptic potential, PSP)达到阈值后, 突触后神经元只发放一个脉冲, 之后将会忽略后续该神经元接收到的所有脉冲, 此时如果脉冲数据中存在着噪声, 很容易造成突触后神经元错误发放. 因此, Tempotron学习算法对于存在噪声干扰的脉冲数据, 识别精度并不高.

为此, 本文对Tempotron学习算法作如下改进: 在训练过程中对不同的脉冲输入模式(P^+或P^-)使用不同的核函数调整神经元PSP, 使得训练后的神经元PSP在输入脉冲为P^+模式时更容易(P^-模式时更难)达到脉冲发放阈值, 从而使得输出神经元在受到噪声干扰时的响应发生改变, 使本来错误发放的突触后神经元被调整为正确的发放, 形成一种多核SNN分类算法MK–Tempotron(multi kernel tempotron), 以提高分类算法的抗噪性能.

在MNIST–DVS[13], Poker–DVS[13]和Posture–DVS[14]等常用DVS数据集上的实验结果表明, 与同类方法[4, 15]相比, 本文所提出方法的识别精度可获得高达14.61%的提升.

2 本文DVS数据特征提取及分类方法

本文所采用的DVS数据特征提取和分类流程如图1所示.

2.1 MSD精确定位与DVS事件流空域特征提取

2.1.1 MSD精确定位

由于DVS在获取静态场景时, 只输出少量的噪声事件, 为解决DVS数据中的背景噪声事件以及产生时间久远的旧事件影响, 本文的MSD将由一个LIF神经元组成, 其阈值机制可以有效实现对背景噪声事件的过滤, 而其泄漏机制可有效降低旧事件对当前片段的干扰. 此外, 为了解决在多种物体运动场景中, 不同物体的最佳检测阈值差异较大的问题, 本文的MSD还将引入一个脉冲最大值监测单元, 以实现针对不同物体检测阈值的自适应变化. 最终, 本文基于LIF神经元模型[5]和脉冲最大值监测单元的MSD, 实现了对不同运动物体下事件流时域ROI关键时间点的自适应精确定位.

MSD如图1左下虚线矩形框所示. 每个输入的DVS事件都能激活一个PSP, 在t_i时刻接收输入事件的PSP

计算如式(1)所示:

$$V(t)=\sum_{t_i}K(t-t_i)+V_{\rm rest}, \tag{1}$$

其中: $V_{\rm rest}$为静息电位, 核函数K的定义如式(2)所示:

$$K(t-t_i)=V_0\times(\exp(\frac{-(t-t_i)}{\tau_{\rm m}})-\exp(\frac{-(t-t_i)}{\tau_{\rm s}})), \tag{2}$$

其中: $\tau_{\rm m}$和$\tau_{\rm s}$分别表示膜和突触电流延迟时间常数, V_0作用是进行归一化.

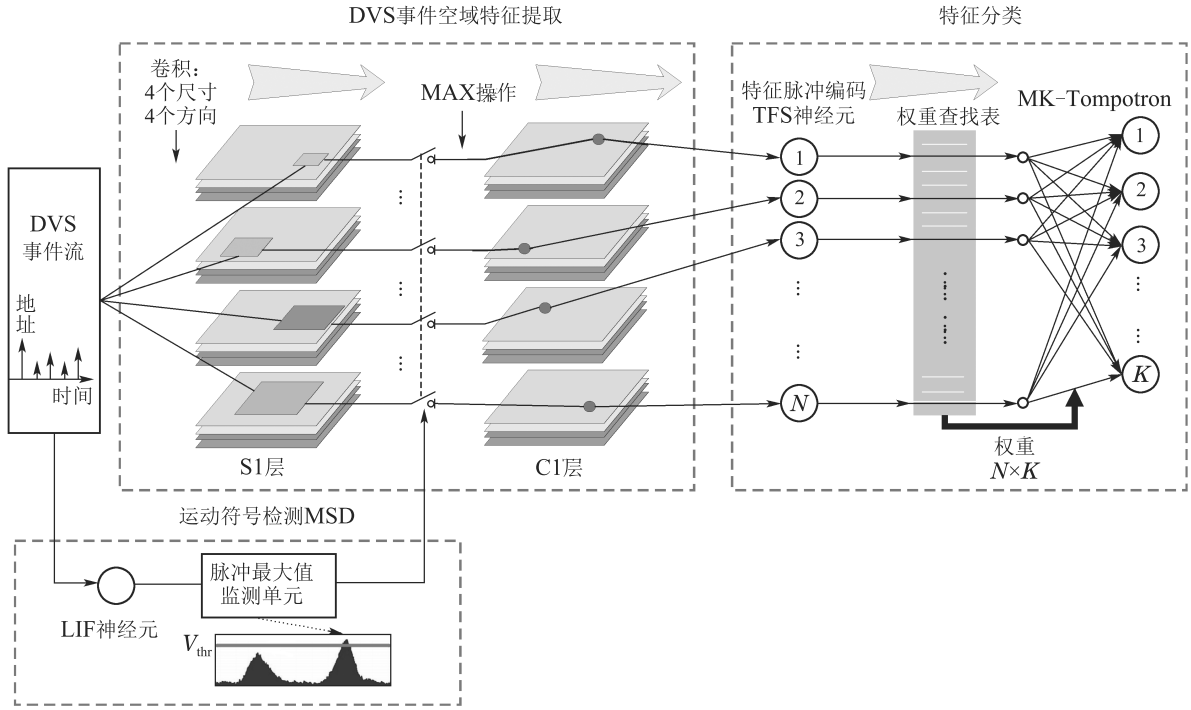


图 1 DVS特征提取与分类流程

Fig. 1 DVS feature extraction and classification process

神经元总电位由脉冲最大值监测单元连续监测, 具体过程为对于某一特定时间t_0, 当满足如式(3)中的条件时, 则t_0为该时间窗内的最大值.

$$V(t_0)\geqslant V_{\rm thr}\wedge V(t_0)\geqslant V(t),\ \forall t\in[t_0-\frac{t_k}{2},t_0+\frac{t_k}{2}], \tag{3}$$

其中: t_k为时间窗大小, $V_{\rm thr}$为预设阈值. 当脉冲最大值监测单元监测到某时刻为该段时间窗内的最大值并且超过预设阈值$V_{\rm thr}$时, 该定位处时间窗内的事件流才会被累积起来作为一个分割片段, 并使得图1中的开路闭合, 此时得到的分割片段才会被输入到后续层中.

由于不同物体的形状大小不一, DVS的工作机制会导致大物体比小物体在运动时产生更多的事件, 若对LIF神经元仅采用硬阈值响应的方式来进行定位, 将会使得不同物体的最佳检测阈值差异很大, 无法达到对不同物体的自适应定位. 而采用脉冲最大值监测单元所监测的最大值可进行动态变化, 不再受硬阈值的影响, 从而对不同物体运动具有自适应性, 解决了不同物体下最佳阈值设置差异大的问题. 此外, 当DVS在捕获物体运动最强烈时, 产生的事件流是最活跃的, 同时也是物体运动特征最为丰富的时刻, 该时刻输出的事件流将使得LIF神经元膜电位达到最大值, 即对应着该事件流的时域ROI关键时间点. 因此, 脉冲最大值监测单元可以持续监测膜电位并对其最大值处进行定位, 该定位处即事件流时域ROI关键时间点.

2.1.2 DVS事件流空域特征提取

为了减少网络中的参数以及提高生物真实性, 本文采用一种由人类视觉皮层启发的预定义权值(Gabor滤波器权值)层次化模型[16]对DVS事件流中的空域特征进行提取.

本文将DVS输出的每个地址事件投影到一组4个不同尺寸(3×3, 5×5, 7×7, 9×9)和4个不同方向为(0°, 45°, 90°, 135°)的Gabor滤波器组, Gabor滤波器[17]的定义如式(4)所示:

$$\begin{cases} g(x,y)=\exp(-\dfrac{X^2+\gamma^2Y^2}{2\sigma^2})\times\cos(\dfrac{2\pi}{\lambda}X),\\ X=x\cos\theta+y\sin\theta,\\ Y=-x\sin\theta+y\cos\theta, \end{cases} \tag{4}$$

其中: θ为Gabor核函数方向, σ为高斯函数标准差, γ为空间长宽比, λ为正弦函数波长, X为卷积核横坐标, Y为卷积核纵坐标, 所使用参数值如表1所示.

表 1 Gabor滤波器参数值

Table 1 Gabor filter parameter values

滤波器尺寸	θ	σ	λ	γ
3	0	1.2	1.5	0.3
5	$\frac{\pi}{4}$	2.0	2.5	0.3
7	$\frac{\pi}{2}$	2.8	3.5	0.3
9	$\frac{3\pi}{4}$	3.6	4.6	0.3

每个滤波器对特定尺寸感受野的神经元细胞进行建模, 从而对特定方向的特征作出最佳响应, 最终得到S1层特征图. 由于S1层中的每个神经元的卷积操作都是动态进行, 为了避免旧的事件对特征提取造成影响, 采用了具有遗忘机制的动态卷积, 即随着时间推移, 响应值将会缓慢恢复至初始值. 卷积后的神经元将对其特定的特征作出一个响应, 当满足MSD的条件时, S1层中的神经元才会与其感受野内的邻近神经元竞争, 只有当它是这个感受野内的响应值最大时(即MAX操作)[16], 该神经元才能在C1层特征图中被保留下来. 而MAX操作后在C1层中被保留下来的每个神经元将表示特定大小和方向的线段特征.

2.2 DVS数据分类算法

保留下来的神经元接着被输入到一组TFS(time-to-first spike)神经元中, 对特征图的每个特征编码成时域脉冲[18], 然后输入到MK–Tempotron中进行突触权重的学习并分类. 此外, 为了使得SNN训练算法更高效, 在编码后每个特征脉冲都与一个相对应的地址关联, 该地址可以用来通过访问权重查找表直接获取其相应的突触权重, 如图1中的权重查找表所示. 下面将对本文提出的SNN学习算法MK–Tempotron进行介绍.

2.2.1 Tempotron算法

Tempotron学习算法[12]以LIF[5]作为神经元模型, 由全部输入该神经元PSP加权和得到突触后神经元膜电位, 如式(5)所示:

$$V(t)=\sum_i w_i\sum_{t_i^{\rm f}}K(t-t_i^{\rm f})+V_{\rm rest},\tag{5}$$

其中: w_i为第i个输入神经元的突触权重, $t_i^{\rm f}$为第i个输入神经元的发放时间, $V_{\rm rest}$为静息电位, K为核函数, 其表达式见式(2).

如果膜电位高于阈值, 神经元会进行发放, 发放脉冲后神经元将会忽略后续的脉冲输入, 并让膜电位恢复到静息电位, 即在发放时间之后到达的脉冲将不再对神经元的膜电位产生影响.

Tempotron学习算法作用是训练突触权重, 使得突触后神经元能够根据样本标签类别决定其是否发放. 当样本标签类别与实际发放情况不符时, 将会对突触权重进行修正. 修正的最终目的是降低损失函数L, 其定义如式(6)所示:

$$L=\begin{cases}\vartheta-V(t_{\max}),\ \text{若为}P^+\text{模式时},\\ V(t_{\max})-\vartheta,\ \text{若为}P^-\text{模式时},\end{cases}\tag{6}$$

其中: $t_{\max}$表示神经元膜电位达到最大值的时间, ϑ表示发放阈值, P^+和P^-分别表示两种不同的输入脉冲模式. 突触权重修正如式(7)所示:

$$\Delta w_i=\begin{cases}\beta\sum\limits_{t_i^{\rm f}<t_{\max}}K(t_{\max}-t_i^{\rm f}), & \text{若}P^+\text{模式错误},\\ -\beta\sum\limits_{t_i^{\rm f}<t_{\max}}K(t_{\max}-t_i^{\rm f}), & \text{若}P^-\text{模式错误},\\ 0, & \text{正确},\end{cases}\tag{7}$$

其中β为学习率.

2.2.2 MK–Tempotron算法

在Tempotron算法中, 噪声干扰突触后神经元的输出响应, 主要通过两种方式:

1) 在脉冲模式为P^+情况下, 存在的噪声可能令神经元膜电位在$t_{\max}$时小于发放阈值, 使得突触后神经元本该发放, 实际却没有发放.

2) 在脉冲模式为P^-情况下, 存在的噪声可能令神经元膜电位在$t_{\max}$时大于发放阈值, 使得突触后神经元本不该发放, 实际却进行发放.

因此在权值训练时, 在输入脉冲模式为P^+(或P^-)情况下, 要使得存在噪声时也能够让突触后神经元发放(或不发放), 则需要使神经元膜电位在$t_{\max}$时变得更高(或更低). 为此, 本文提出MK–Tempotron算法, 该算法在训练权值时, 对两种不同的输入模式分别采用不同的核函数K_1和K_2来计算神经元膜电位, 算法步骤如下:

步骤 1 初始化权重并输入样本脉冲;

步骤 2 神经元响应状态与标签响应状态进行比较, 判断应选择K_1或K_2求膜电位;

步骤 3 通过选择相应核函数计算神经元膜电位, 得到神经元的响应状态;

步骤 4 响应状态与标签相符则结束, 否则使用式(7)调整权重并跳至步骤2.

其中$\boldsymbol{K}_1$和$\boldsymbol{K}_2$的定义分别如式(8)–(9)所示:

$$K_1(t-t_i^{\rm f})=$$
$$(V_0-a)(\exp(\frac{-(t-t_i^{\rm f})}{\tau_{\rm m}})-\exp(\frac{-(t-t_i^{\rm f})}{\tau_{\rm s}})),\quad(8)$$
$$K_2(t-t_i^{\rm f})=$$
$$(V_0+b)(\exp(\frac{-(t-t_i^{\rm f})}{\tau_{\rm m}})-\exp(\frac{-(t-t_i^{\rm f})}{\tau_{\rm s}})),\quad(9)$$

其中: a和b为变化系数, 使得在输入脉冲模式为P^+(或P^-)时神经元膜电位更低(或更高), 其中P^+和P^-模式时神经元膜电位的计算分别如下式(10)–(11)所示:

$$V^{P^+}(t_{\max})=\sum_i w_i\sum_{t_i^{\rm f}}K_1(t_{\max}-t_i^{\rm f})+V_{\rm rest},\quad(10)$$

$$V^{P^-}(t_{\max})=\sum_i w_i\sum_{t_i^{\rm f}}K_2(t_{\max}-t_i^{\rm f})+V_{\rm rest}.\quad(11)$$

其损失函数如式(12)所示:

$$L=\begin{cases}\vartheta-V^{P^+}(t_{\max}), & \text{若为}P^+\text{模式时},\\ V^{P^-}(t_{\max})-\vartheta, & \text{若为}P^-\text{模式时},\end{cases}\quad(12)$$

其中: $V^{P^+}(t_{\max})$是神经元在P^+时的最大膜电位, $V^{P^-}(t_{\max})$是神经元在P^-时的最大膜电位. 通过此方法进行训练的神经元突触权重最终将会朝着正确发放的方向发生改变.

对于多类别分类任务, 训练时, 本文将采用One-hot编码[19]对N个MK–tempotron神经元进行标记. 若属于第1类, 那么第1个MK–tempotron神经元的输出标记为1(神经元应发放), 其他神经元的输出标记为0(不应发放). 在测试时, 只需观察哪些神经元是否发放, 即可判断其类别.

3　实验结果与分析

本文方法的实验都在MATLAB(版本为2015a)上进行软件仿真, 硬件环境为CPU i7–6700, 显卡RT–X2–080ti, 内存32 G的环境下进行实验. 实验中DVS数据集参数设置都严格参照Peng[4]中使用的值.

3.1　DVS数据集

本文采用了MNIST–DVS[13], Poker–DVS[13], Posture–DVS[14]3个常用的DVS数据集对MK–Tempoton的抗噪性能与Tempoton进行比较, 并与同类方法[4, 15]进行对比评估.

MNIST–DVS[13]数据集中包含了0–9共10类手写数字, 是由10,000张原始的MNIST[20]手写数字图像, 通过放大而得到3个不同规格的图像以缓慢移动的方式显示在显示器上, 使用分辨率为128×128的DVS记录得到, 每个样本的记录时长为100 ms, 像素分辨率为28×28. 由于记录到的MNIST–DVS数据集存在着由动态背景引起的噪声, 其识别难度比标准MNIST数据集更高.

Poker–DVS[13]数据集包含了分别为梅花、方块、红桃和黑桃4种不同花色的扑克牌, 其分辨率为32×32. 该数据集是通过在DVS摄像机记录特制的扑克牌组2∼4 s, 每张卡片可在屏幕上显示20∼30 ms, 最终获得131个包含着4种花色样本的DVS数据集.

Posture–DVS[14]数据集包含了分别为弯腰、坐下和站立、行走3种不同的人类活动姿态共484个样本, 其分辨率为32×32.

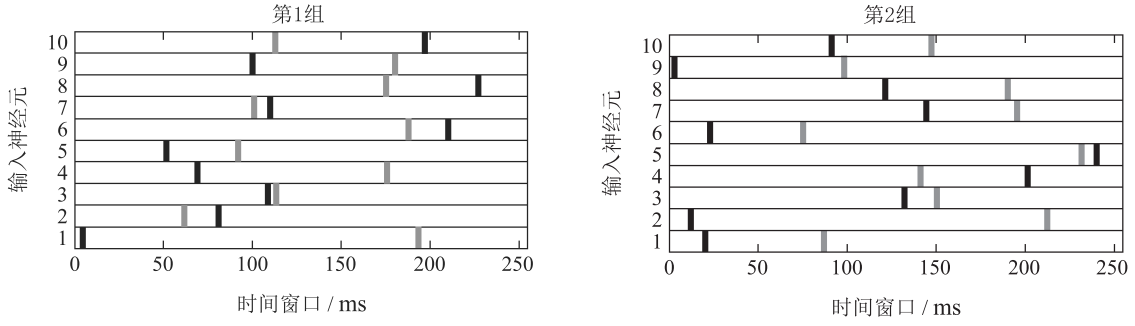
这些数据集均是由DVS记录得到的, 会造成颜色信息的缺失和受到动态环境引起的噪声事件影响, 人眼识别这些样本也是存在一定的难度.

3.2　MK–Tempoton的抗噪性及其与Tempoton的比较

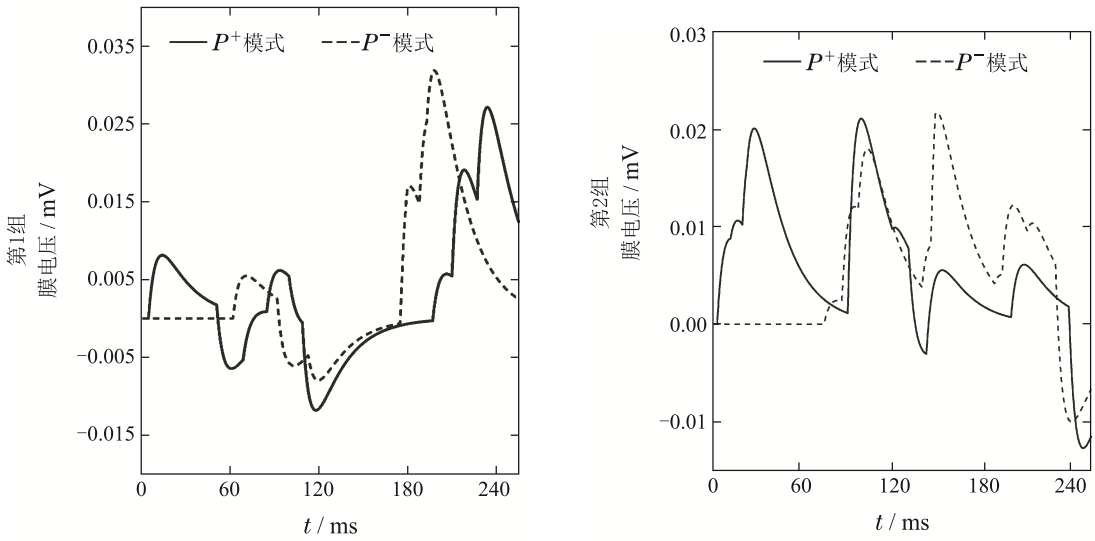
3.2.1　MK–Tempoton的抗噪性

本部分将使用两种脉冲输入模式P^+和P^-, 分别对Tempotron和MK–Tempotron训练后得到的膜电位进行对比实验, 给出其中的两组实验结果如下图2所示.

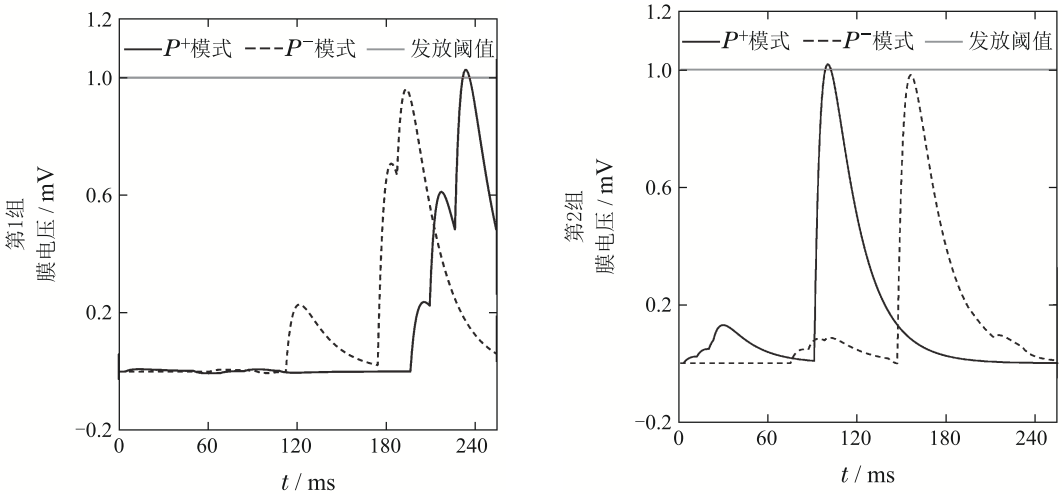
两组不同的输入脉冲P^+和P^-模式如下图2(a)所示, 其中: 黑色代表P^+模式, 灰色代表P^-模式. 对于P^+和P^-模式都由10个输入神经元组成, 时间窗口大小为255 ms, 输入神经元将在时间窗口内随机地发放脉冲.



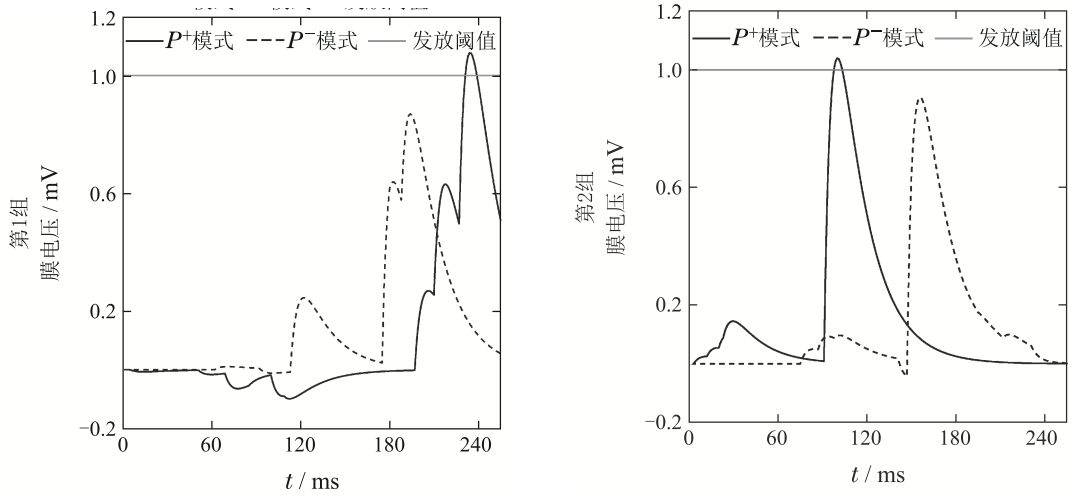
(a) 输入脉冲模式



(b) 训练前电压



(c) Tempotron训练后电压



(d) MK–Tempotron训练后电压

图 2 Tempotron和MK–Tempotron抗噪性实验的两组对比数据

Fig. 2 Comparison of Tempotron and MK–Tempotron in anti-noise perfomance

使用两组不同的P^+和P^-模式输入得到初始权重时的神经元膜电位分别如图2(b)的黑色实线和灰色虚线所示, 由图2(b)可看出, 此突触后神经元在接收到P^+和P^-模式的输入脉冲时, P^+和P^-模式的膜电位均没超过阈值$V_{\rm thr}=1$, 即突触后神经元在接收到输入的两种脉冲模式后均没有发放, 因此在初始化权重的情况下无法正确区分P^+和P^-模式.

对此, 下面将分别使用Tempotron和MK–Tempotron算法对突触权重进行训练, 以使得其能在P^+模式下发放, 而在P^-模式下不发放. Tempotron算法在两组P^+和P^-模式下训练后的神经元膜电压如图2(c)所示, MK–Tempotron算法在两组P^+和P^-模式下训练后的神经元膜电压如图2(d)所示.

由图2(c)–2(d)可看出使用Tempotron和MK–Tempotron训练后, 突触后神经元都能够正确分类. 在P^+模式下使用了MK–Tempotron训练得到的神经元膜电压在最大值处到$V_{\rm thr}=1$的距离更远, 在P^-模式下得到的神经元膜电压在最大值处到$V_{\rm thr}=1$的距离同样更远, 而使用了Tempotron训练后, 在P^+和P^-模式下得到的神经元膜电压最大值处到$V_{\rm thr}=1$的距离非常近. 因此, 使用MK–Tempotron算法具有更强的抗噪性能, 但是在有噪声存在时, MK–Tempotron 算法更能使输出神经元作出正确的响应.

3.2.2 MK–Tempotron与Tempotron的比较

使用MK–Tempotron算法和Tempotron算法应用于本文的DVS数据分类方法, 在上述3个DVS数据集上进行实验, 每种方法重复实验10次取其平均值, 每次实验将随机抽取90%的样本作为训练集, 剩余的10%样本作为测试集, 结果如表2所示.

表 2 MK–Tempotron与Tempotron实验比较

Table 2 MK–Tempotron and Tempotron experimental

算法	MNIST–DVS 准确率/%	Poker–DVS 准确率/%	Posture–DVS 准确率/%
MSD+Gabor+Tempotron	75.52	88.33	95.61
MSD+Gabor+MK–Tempotron	78.11	91.66	99.74

如表2实验结果所示, MK–Tempotron算法使用在本文的方法中, 在DVS数据集MNIST–DVS, Poker–DVS, Posture–DVS上的识别精度比使用Tempotron算法的识别精度分别提高了2.59%, 3.33%, 4.13%, 实验结果表明MK–Tempotron算法在存在背景噪声的DVS数据集中也能达到较好的抗噪性能, 从而使识别精度有所提升.

其中: 实验中各个数据集所使用的膜延迟时间常数$\tau_{\rm m}$, 突触电流延迟时间常数$\tau_{\rm s}$, MSD的时间窗口大小$t_{\rm k}$, 事件流卷积的泄漏率μ, 核函数K_1和K_2的变化量系数a和b的值如表3所示.

表 3 实验参数

Table 3 Experimental parameters

参数	MNIST–DVS	Poker–DVS	Posture–DVS
$\tau_{\rm m}$	20 ms	20 ms	20 ms
$\tau_{\rm s}$	5 ms	5 ms	5 ms
$t_{\rm k}$	30 ms	30 ms	1 s
μ	10 $\rm s^{-1}$	10 $\rm s^{-1}$	50 $\rm s^{-1}$
a	0.2	0.2	0.3
b	0.7	0.4	0.5

3.3 本文方法(MSD+Gabor+MK–Tempotron)与同类方法的比较

为进一步验证本文方法(MSD+Gabor+MK–Tempotron)的性能, 本文将其与事件包+支持向量机[4](bag of events+support vector machines, BOE+SVM)和Gabor+Hausdorff[15]方法进行识别精度和分类效率的比较. 在BOE+SVM方法中, 利用连续事件的联合概率分布对每个输入事件进行表征, 然后使用SVM对特征进行分类, 该基于概率统计的方法有着良好的识别精度和分类效率. 在Gabor+Hausdorff的方法中, 使用Gabor滤波器能够较好地提取目标尺度和位移不变性的线段特征, 然后采用结合动态聚类的改进Hausdorff距离分类器进行分类, 该方法能对DVS数据集有着良好的识别效果.

本部分仍使用MNIST–DVS, Poker–DVS和Posture–DVS数据集进行实验, 每种方法重复实验十次取其平均值, 每次实验将随机抽取90%的样本作为训练集, 剩余的10%样本作为测试集, 实验中本文方法所使用的参数值如表3所示. 实验结果如表4所示.

由表4可知, 本文方法在MNIST–DVS数据集的识别精度比Gabor+Hausdorff方法提高了14.61%, 相比BOE+SVM方法识别精度提高了3.29%. 由于BOE+SVM方法使用了SVM[21]进行分类, 不需要训练分类器, 因此分类耗时最短. 而本文的方法与Gabor+Hausdorff方法相比较, 由于在特征提取过程中使用到MAX操作以及在分类中采用可以直接根据地址访问权重查找表的SNN算法, 因此使得分类效率有很大幅度的提升. 本文的方法在MNIST–DVS数据集上的识别精度比BOE+SVM和Gabor+Haus-

dorff方法更高, 而且有着很高的分类效率, 从而证明了本文所使用的特征提取方法以及MK–Tempotron分类算法是十分有效的.

在Poker–DVS数据集中, 3种方法的识别精度非常接近, 尽管本文的方法在此数据集上的识别精度没有超过另外两种方法, 但与识别精度最高的方法相比仅有1.34%的差距, 也能达到91.66%的高识别精度水准, 其差距在只有131个样本的数据集中并不大. 与此同时, 本文的方法与Gabor+Hausdorff方法相比仍然保持着较高的分类效率水准, 因此, 本文方法在Poker–DVS数据集的实验中与另外两个方法一样能达到较高的分类性能. 而在Posture–DVS数据集的实验中, 本文方法的识别精度最高可达到100%, 经过十次实验取其平均值最终能达到99.74%的识别精度, 比BOE+SVM和Gabor+Hausdorff方法分别提高了7.86%和1.08%. 在分类效率方面, 与Gabor+Hausdorff方法相比有着绝对的优势. 如表4所示, 实验结果表明本文方法, 在能达到理想的识别精度同时, 也能保持着较高的分类效率, 从而证明本文方法是有效可行的.

表 4 本文方法与同类方法实验对比

Table 4 Comparison of methods in this paper with similar methodss

算法	MNIST–DVS		Poker–DVS		Posture–DVS	
	准确率/%	分类效率/s	准确率/%	分类效率/s	准确率/%	分类效率/s
BOE+SVM [4]	74.82	3.75	93.00	0.01	98.66	0.8
Gabor+Hausdorff [15]	63.50	7691.26	92.53	23.34	91.88	11430.28
MSD+Gabor+MK–Tempotron(本文方法)	78.11	537.05	91.66	16.61	99.74	92.54

4 结束语

本文主要针对DVS数据分类系统中, 事件流的时域ROI定位与分割问题, 提出一种精确时序MSD进行事件流时域ROI的精确定位, 解决了现有方法不能根据不同物体运动自适应地设定最佳检测阈值、无法对静态场景中少量背景噪声进行过滤等问题. 此外, 针对已有SNN学习算法抗噪性差的问题, 提出了一种抗噪性能好的SNN学习算法MK–Tempotron, 该算法通过在训练过程中对两种不同的输入脉冲模式分别采用两个不同的核函数调整神经元膜电位, 使得即使在存在背景噪声的DVS数据中, 输出神经元也能作出正确的响应. 本文所提出的方法, 与基于事件的BOE+SVM和Gabor+Hausdorff的同类方法相比, 识别率能获得高达14.61%的提升.

参考文献:

[1] GUO M, HUANG J, CHEN S. Live demonstration: A 768 × 640 pixels 200 Meps dynamic vision sensor. *The 2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. New York: IEEE, 2017: 1.

[2] BABY S A, VINOD B, CHINNI C, et al. Dynamic vision sensors for human activity recognition. *2017 the 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. New York: IEEE, 2017: 316 – 321.

[3] GHOSH R, MISHRA A, ORCHARD G, et al. Real-time object recognition and orientation estimation using an event-based camera and CNN. *The 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*. New York: IEEE, 2014: 544 – 547.

[4] PENG X, ZHAO B, YAN R, et al. Bag of events: An efficient probability-based feature extraction method for AER image sensors. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 28(4): 791 – 803.

[5] QI Y, SHEN J, WANG Y, et al. Jointly learning network connections and link weights in spiking neural networks. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence Main Track (IJCAI)*. Stockholm: IJCAI, 2018: 1597 – 1603.

[6] STROMATIAS E, SOTO M, SERRANO-GOTARREDONA T, et al. An event-driven classifier for spiking neural networks fed with synthetic or dynamic vision sensor data. *Frontiers in Neuroscience*, 2017, 11: 350.

[7] NEIL D, PFEIFFER M, LIU S C. Learning to be efficient: Algorithms for training low-latency, low-compute deep spiking neural networks. *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. New York: ACM, 2016: 293 – 298.

[8] CAO Y, CHEN Y, KHOSLA D. Spiking deep convolutional neural networks for energy-efficient object recognition. *International Journal of Computer Vision*, 2015, 113(1): 54 – 66.

[9] DIEHL P U, NEIL D, BINAS J, et al. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. *The 2015 International Joint Conference on Neural Networks (IJCNN)*. New York: IEEE, 2015: 1 – 8.

[10] PEREZ-CARRASCO J A, ZHAO B, SERRANO C, et al. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing–application to feedforward ConvNets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2706 – 2719.

[11] RUECKAUER B, LUNGU I A, HU Y, et al. Theory and tools for the conversion of analog to spiking convolutional neural networks. *arXiv preprint arXiv: 1612.04052*, 2016.

[12] TANG H, YU Q, TAN K C. Learning real-world stimuli by single-spike coding and tempotron rule. *The 2012 International Joint Conference on Neural Networks (IJCNN)*. New York: IEEE, 2012: 1 – 6.

[13] SERRANO-GOTARREDONA T, LINARES-BARRANCO B. Poker-DVS and MNIST–DVS. Their history, how they were made, and other details. *Frontiers in Neuroscience*, 2015, 9: 481.

[14] SHI C, LI J, WANG Y, et al. Exploiting lightweight statistical learning for event-based vision processing. *IEEE Access*, 2018, 6: 19396 – 19406.

[15] CHEN S, AKSELROD P, ZHAO B, et al. Efficient feedforward categorization of objects and human postures with address-event image sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 34(2): 302 – 314.

[16] ORCHARD G, MEYER C, ETIENNE-CUMMINGS R, et al. HFirst: A temporal approach to object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(10): 2028 – 2040.

[17] GDYCZYNSKI C M, MANBACHI A, HASHEMI S M, et al. On estimating the directionality distribution in pedicle trabecular bone from micro-CT images. *Physiological Measurement*, 2014, 35(12): 2415.

[18] KRAUSE T U, WURTZ P D D R. *Rate coding and temporal coding in a neural network*. Bochum, Germany: University of Bochum, 2014.

[19] YU Q, TANG H, TAN K C, et al. Rapid feedforward computation by temporal encoding and learning with spiking neurons. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 24(10): 1539 – 1552.

[20] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278 – 2324.

[21] SHALEV-SHWARTZ S, SINGER Y, SREBRO N, et al. Pegasos: Primal estimated sub-gradient solver for svm. *Mathematical Programming*, 2011, 127(1): 3 – 30.

作者简介:

黄庆坤　硕士研究生, 主要研究方向为计算机视觉、神经形态计算等, E-mail: 517262103@qq.com;

陈云华　博士, 副教授, 主要研究方向为深度学习、神经形态计算、计算机视觉, E-mail: yhchen@gdut.edu.cn;

张　灵　博士, 教授, 主要研究方向为模式识别、智能化信息处理、人工智能, E-mail: june4567@126.com;

兰浩鑫　硕士研究生, 主要研究方向为计算机视觉、神经形态计算等, E-mail: 412217176@qq.com.