基于 FMCW 雷达的双流融合神经网络手势识别方法

王 勇,王沙沙,田增山,周 牧,吴金君
(重庆邮电大学通信与信息工程学院,重庆 400065)

摘 要: 针对传统光学摄像头和无线技术的手势识别方法受光照环境影响和空间纵向、横向特征不全的问题,该文提出一种基于调频连续波(Frequency Modulated Continuous Wave,FMCW)雷达信号的双流融合神经网络(Two-Stream Fusion Neural Network,TS-FNN)手势识别方法. 首先,利用二维快速傅立叶变换(Fast Fourier Transform,FFT)求取中频信号的频谱,估计手势的距离和速度,并利用多重信号分类 (Multiple Signal Classification,MUSIC)方法计算角度. 其次,利用这三维参数在时间上的累积,将一个手势动作映射为 32 帧距离-速度矩阵图和角度时间图. 最后,建立 TS-FNN 进行手势特征提取和特征融合. 实验结果表明,TS-FNN 方法与传统卷积神经网络相比,手势的平均识别准确率提升了约 5%.

关键词: 人机交互; 手势识别; FMCW 雷达; 深度学习
中图分类号: TN961 **文献标识码:** A **文章编号:** 0372-2112 (2019)07-1408-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.07.003

Two-Stream Fusion Neural Network Approach for Hand Gesture Recognition Based on FMCW Radar

WANG Yong, WANG Sha-sha, TIAN Zeng-shan, ZHOU Mu, WU Jin-jun
(*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: To deal with the problem of easily being affected by illumination environment of the traditional optical camera based hand gesture recognition method and the incomplete spatial and lateral characteristics of the wireless based hand gesture recognition method, this paper proposes a frequency modulated continuous wave (FMCW) radar signal based two-stream fusion neural network (TS-FNN) for hand gesture recognition. Firstly, the spectrum of the IF signal is obtained by two-dimensional Fast Fourier Transform (2D-FFT), the range and speed of the gesture are estimated, and the angle is calculated by the Multiple Signal Classification (MUSIC) method. Secondly, using the accumulation of three-dimensional parameters in time, a gesture action is mapped to a 32-frame range-speed matrix diagram and an angular-time map. Finally, TS-FNN is established for gesture feature extraction and classification. The experimental results show that compared with the existing methods, the TS-FNN method improves the average recognition accuracy by about 5%.

Key words: human-machine interaction; gesture recognition; FMCW radar; deep learning

1 引言

手势识别作为人机交互的重要组成部分,其研究发展影响着人机交互的自然性和灵活性,并且在各个领域得到广泛应用. 在家庭娱乐方面,根据用户在游戏环境中左右挥动等动作来控制游戏中的角色,使用户体验效果更好. 在智能驾驶方面,由于司机在驾驶过程中可能被车载导航系统、电话系统分散注意力,可以通过识别驾驶员手势动作完成对导航系统以及车载娱乐系统的控制,提高驾驶的安全性.

手势识别技术根据数据来源主要分为:基于光学摄像头[1~3]的手势识别方法和基于无线技术[4~6]的手势识别方法. 在光学摄像头手势识别技术中,首先使用摄像头采集数据集,根据图像本身的变化以及手势的运动轨迹建立模型,进而提取手势特征,再利用神经网络[1]、模板匹配[2]和支持向量机(Support Vector Machine,SVM)[3]等机器学习方法进行识别. Coelho 等人[1]利用尺度不变特征变换匹配算法(Scale Invariant Feature Transform,SIFT)描述手势图像中的特征,但这种算法的实时性不高且不能准确提取边缘光滑的目标

收稿日期: 2018-07-03;修回日期:2018-12-10;责任编辑:孙瑶
基金项目: 国家自然科学基金(No. 61771083, No. 61704015);长江学者和创新团队发展计划(No. IRT1299);重庆市科委重点实验室专项经费—重庆市基础科学与前沿技术研究项目(No. cstc2017jcyjAX0380, No. cstc2015jcyjBX0065);重庆市高校优秀成果转化资助项目(No. KJZH17117)

特征. 文献[7]使用卷积神经网络(Convolutional Neural Networks,CNN)提取光学手势图片特征,可以实现手势图片特征的实时提取和分类. 但由于 CNN 只能提取手势单张瞬时图片的特征,忽略了手势的连续性信息. Molchanov 等人[8]使用三维卷积神经网络(3D Convolutional Neural Networks,3D-CNN)来提取手势的连续特征,但由于网络中的三维卷积只能提取短暂连续几张手势图片的时间信息,并不能完整地表示一个手势的时序性. 文献[9]中采用 3D-CNN 提取手势图片中的时空信息,再利用长短期记忆网络[10](Long Short-Term Memory,LSTM)获取手势中的时序信息,充分提取了手势动作特征. 然而,作者并没有在不同光照变化场景进行分析和测试.

为了克服光照影响,基于无线技术的手势识别方法采用了无线设备采集手势信号. 该方法的信号来源包括雷达信号和无线信道状态信息,在采集手势信号后,通过信号处理分析手势信号中的频域信息,再提取手势的运动参数,然后通过聚类[4]、动态时间规整[5]和隐马尔可夫模型[6]等方法进行识别. 文献[5]利用无线信道状态信息和太赫兹雷达信号获取数据源,并利用手势径向速度表征手势行为. 但作者在每一时刻直接计算一个距离标量值来表示手势特征信息,使得特征提取不全(缺乏速度和角度信息),从而降低了手势识别的准确率. Wang 等人在文献[11]中使用调频连续波(Frequency Modulated Continuous Wave,FMCW)雷达采集手势信号,并通过信号处理求取距离和速度,将对应的信号幅值映射为参数图. 最后,使用该参数图来表示每一时刻的手势,并将参数图输入到深度学习网络中进行特征提取和分类. 但该方法只对手势的径向变化比较敏感,限制了对横向变化敏感的角度特征提取,从而极大地限制了手势识别应用范围. 总之,现有基于无线电的手势识别方法,由于数据特征信息不全,使得手势识别精度较低[12].

基于上述分析,本文提出了一种基于 FMCW 雷达多参数图像的双流融合神经网络(Two-Stream Fusion Neural Network,TS-FNN)手势识别方法. 该方法包括三个阶段:第一阶段为雷达信号处理过程. 它主要包含 2 个步骤:(1)获取雷达发射天线和接收天线的混频信号,并求取手势的距离和速度,再根据多重信号分类算法(Multiple Signal Classification,MUSIC)[13]计算手势的角度;(2)将距离和速度映射到同一幅图中,即距离-速度图,并通过信号的时间顺序将距离-速度图和角度图生成序列,形成三维距离-速度时间图和二维角度时间图. 第二阶段搭建 TS-FNN. 设计卷积神经网络分别对距离-速度时间图和角度时间图进行特征提取,得到两个能够独立表示手势的特征向量. 再将特征向量进行并联融合,并利用 LSTM 进行时序特征提取后,采用归一化指数函数对提取到的手势特征进行分类. 第三阶段为训练和测试阶段. 本文将实测数据分为训练集和验证集,使用训练集训练 TS-FNN 网络,再利用验证集进行测试,并统计手势识别结果.

2　手势信号数据处理

2.1　中频信号提取

为了获取 FMCW 雷达的中频信号,需要将发射信号和接收信号输入到混频器,再通过低通滤波器滤除高频部分后得到中频信号.

雷达发射信号和接收信号皆为锯齿波,接收信号在时间上存在固定延时 t_d,具体形式如图 1 所示.

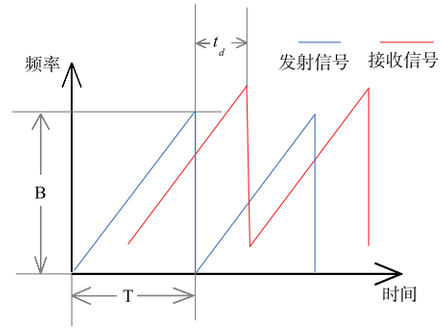

图1　发射信号与接收信号的时间相关曲线

由文献[14]可知,FMCW 雷达发射的锯齿波信号为可表示为

$$S_{\mathrm{TX}}(t) = A_{\mathrm{TX}}\cos\left(2\pi f_c t + 2\pi\int_0^t f_T(\tau)\,\mathrm{d}\tau\right)$$
$$= A_{\mathrm{TX}}\cos\left(2\pi f_c t + 2\pi\times\frac{1}{2}\times\frac{B}{T}t^2\right) \tag{1}$$

其中,$f_T(\tau)=\dfrac{B}{T}\times\tau$ 是发射信号频率随着时间变化的线性函数,f_c 是载波频率,B 是带宽,A_{TX} 是发射信号的幅值,T 是信号周期.

发射信号经过 $t_d=2\times\dfrac{R_0+vt}{c}$ 的时延和手势运动使得接收端的频移可表示为

$$\Delta\varphi=\frac{4\pi v\times t}{\lambda} \tag{2}$$

在接收端的信号由于延时和多普勒频移,频率变为 $f_R(t)=\dfrac{B}{T}(t-t_d)+\dfrac{\Delta\varphi}{2\pi}$,接收信号[14]可表示为

$$S_{\mathrm{RX}}(t)=A_{\mathrm{RX}}\cos\left(2\pi f_c t+2\pi\frac{B}{T}\times\frac{1}{2}(t^2+t_d^2-2t_d t)+\Delta\varphi\right) \tag{3}$$

其中,v 是手势相对雷达的径向运动速度,R_0 是在时间 $t=0$ 时手势与雷达的距离,c 为光速,A_{RX} 是接收信号的幅值.

为了获取发射信号和接收信号中的频移和相位差,将发射信号 $S_{TX}(t)$ 和接收信号 $S_{RX}(t)$ 输入到混频器,并经低通滤波器滤除高频后得到中频信号 $S_{IF}(t)$

$$S_{IF}(t)=\frac{1}{2}A_{IF}\cos\left\{2\pi\cdot\frac{B}{T}\left(\frac{1}{2}t_d{}^2-t_dt\right)-\Delta\varphi\right\} \tag{4}$$

由于实际测量中 t_d 非常小,$t_d{}^2$ 项可忽略不计,中频信号频率可近似为

$$f_{IF}=\frac{B}{T}\times 2\frac{R_0+vt}{c}+\frac{\Delta\varphi}{2\pi t} \tag{5}$$

2.2 基于 2D-FFT 的距离-速度图计算

由式(5)可知,距离 d 与中频信号频率 f_{IF} 成正比. 所以,手势到雷达的距离[15]为

$$d=\frac{c\times f_{IF}\times T}{2B} \tag{6}$$

当手势处于静止状态时,中频信号应为一个频率恒定的正弦信号. 而当手势相对雷达运动时,中频信号发生频移,信号的频率随着手势相对雷达的远近而变化.

对中频信号进行二维快速傅立叶变换[16],即可得到每个脉冲的多普勒频移 f_{FFT}. 由中频信号频率以及式(2)可得手势运动速度

$$v=\frac{\lambda\Delta\varphi}{4\pi}=\frac{\lambda f_{FFT}}{4T} \tag{7}$$

由式(6)和式(7)可知,距离与中频信号频率 f_{IF} 成正比,速度和脉冲的多普勒频移 f_{FFT} 成正比. 雷达每次发送 32 帧数据,为了获取中频信号频率 f_{IF},对中频信号每个脉冲周期采样 64 个点进行快速傅里叶变换. 计算出每个脉冲中频率为 f_{IF} 的频移后,将结果中同一频点的复信号生成一个新的频移信号,进而再对此信号进行 FFT 得到多普勒频移 f_{FFT}.

每一帧信号可以得到一个距离-速度矩阵图,本文每次发送 32 帧信号,进而可以得到 32 个距离-速度序列图来表示一个手势动作,从推拉手势的 32 帧距离-速度图中每隔 4 帧挑选 1 帧,选出的 8 帧距离-速度图如图 2 所示.

由图 2 第 1、5、9 和 13 帧可知,在前 16 帧中,手势离雷达的距离越来越近,速度变快再变慢;由第 17、21、25 和 29 帧可知,在后 16 帧中手势离雷达的距离越来越远,速度变快再变慢. 所以,前 16 帧表示推的动作,后 16 帧表示拉的动作.

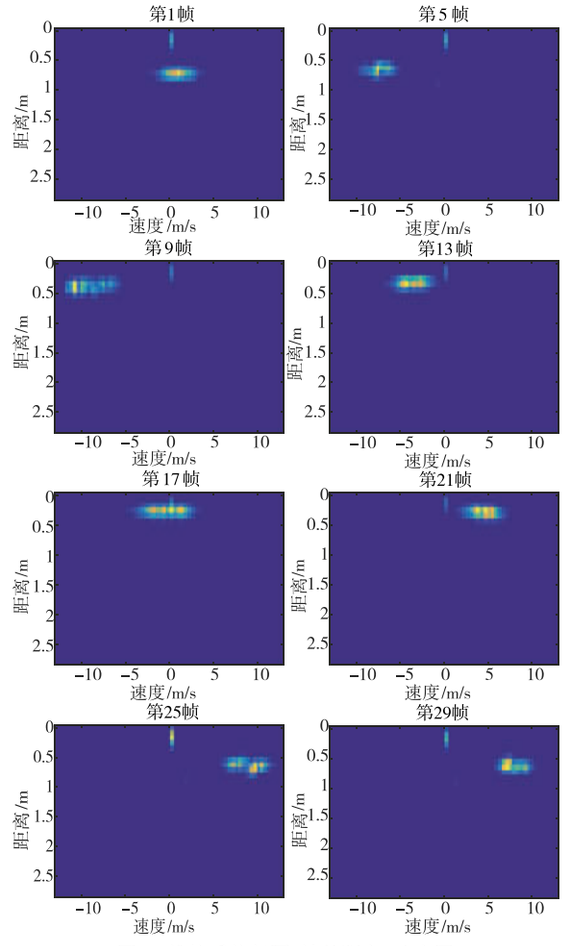

图2 推拉手势每隔4帧的距离-速度图

2.3 基于 MUSIC 算法的角度时间图计算

为了计算手势的横向信息,本文采用 MUSIC 算法[13]估计手势目标的角度.

当前方有 K 个目标时,雷达发射一帧信号后第一个目标的接收信号为 $S_1(t)$,则 K 个目标的接收信号为 $\boldsymbol{S}(t)=[S_1(t)\quad S_2(t)\quad\cdots\quad S_K(t)]^{\mathrm{T}}$. 由于阵元之间的间隔为 d,设第 K 个目标的角度为 θ_K,则所有 K 个目标接收信号的导向矢量阵如式(8)所示. 其中,M 表示阵元数.

$$\boldsymbol{A}=[\boldsymbol{a}(\theta_1)\quad\cdots\quad\boldsymbol{a}(\theta_K)]=\begin{bmatrix}1&\cdots&1\\ \exp\left(-\mathrm{j}2\pi d\sin\theta_1\cdot\frac{f_0}{c}\right)&\cdots&\exp\left(-\mathrm{j}2\pi d\sin\theta_K\cdot\frac{f_0}{c}\right)\\ \cdots&\cdots&\cdots.\\ \exp\left(-\mathrm{j}2\pi(M-1)d\sin\theta_1\cdot\frac{f_0}{c}\right)&\cdots&\exp\left(-\mathrm{j}2\pi(M-1)d\sin\theta_K\cdot\frac{f_0}{c}\right)\end{bmatrix} \tag{8}$$

所以,最终的接收信号为:

$$\boldsymbol{X}(t)=\boldsymbol{A}\boldsymbol{S}(t)+\boldsymbol{N}(t) \tag{9}$$

其中,$\boldsymbol{N}(t)=[n_1(t)\quad n_2(t)\quad\cdots\quad n_M(t)]^{\mathrm{T}}$ 为每个阵元的噪声向量.

计算 $\boldsymbol{X}(t)$ 的协方差矩阵 $\boldsymbol{R}=E\{\boldsymbol{X}(t)\boldsymbol{X}^{\mathrm{H}}(t)\}$,对其进行特征分解,得到特征向量 $\boldsymbol{v}_i(i=1,2,\cdots,M)$. 其中 $M-K$ 个特征值为 σ^2,即 $\boldsymbol{R}$ 是 $M-K$ 重的. 本文使用这 $M-K$ 个特征值对应的特征向量来定义噪声子空间,并使用其他的特征向量来定义信号子空间[17]. 令 $\boldsymbol{E}_N=[\boldsymbol{v}_1\quad\boldsymbol{v}_2\quad\cdots\quad\boldsymbol{v}_M]^{\mathrm{T}}$,根据信号子空间和噪声子空间的正交性可知 $\boldsymbol{E}_N^{\mathrm{H}}\boldsymbol{a}(\theta_k)=\boldsymbol{0}$. 由此构造空间谱函数如下:

$$\boldsymbol{P}_{\mathrm{MUSIC}}(\theta)=\frac{1}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{E}_N\boldsymbol{E}_N^{\mathrm{H}}\boldsymbol{a}(\theta)} \tag{10}$$

对式(10)进行谱峰搜索,其中 K 个极大值所对应的 θ 即为信号源的方向. 由式(10)可知,每一帧数据可以进行角度搜索,并获取每个角度下信号的峰值.

本文实验中每次发送 32 帧信号,对每一帧信号估计角度,将计算结果按时间顺序构成一个角度时间图. 因此,相对于雷达横向变化的手势如左右划、左划和右划对应的角度时间图变化十分明显,如图 3 所示.

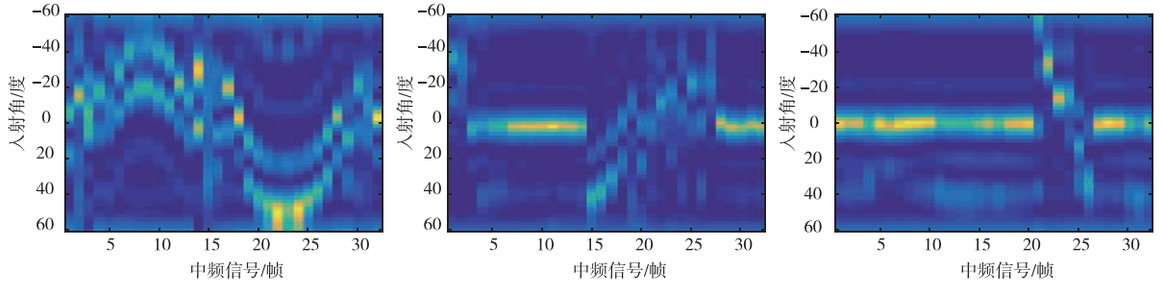

图3　角度时间图

3　双流融合神经网络

本文将两种参数图的特征进行了融合,使得融合特征能够结合横向和纵向信息共同表示每种手势的特点. 本文提出的 TS-FNN 主要包含 4 个步骤:(1)设计 3D-CNN 对距离-速度图进行特征提取;(2)设计 CNN 网络对角度时间图进行特征提取;(3)将从距离-速度图和角度时间图提取到的特征进行并联融合,并使用 LSTM 提取并联融合特征的时序信息;(4)添加 2 个全连接层和 1 个输出层进行特征分类. TS-FNN 的实现流程如图 4 所示.

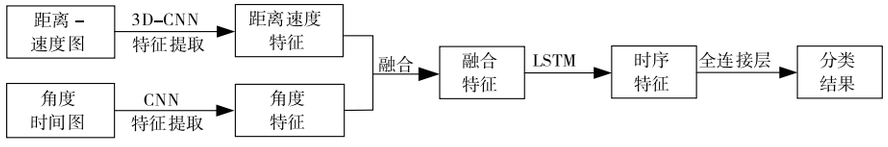

图4　TS-FNN框架

3.1　距离-速度特征提取

在距离-速度图求解过程中,旨在提取序列中手势连续位置的变化信息. 根据序列的维度信息,本文采用 3D-CNN 进行特征提取. 根据 VGG-16[18] 的网络结构,设置网络为 5 层卷积池化层. 由于手势运动范围较小,在距离-速度图中的运动幅度不大,所以在网络的后 3 层设置两次卷积操作,整个网络包括 5 个卷积池化层,2 个全连接层. 在第一和第二个卷积池化层中,对输入进行了一次卷积和一次池化. 在后三个卷积池化层中,进行两次卷积和一次池化,全连接层连接最后一个池化,生成 1024 个神经元. 因此,网络包含了 8 次卷积,5 次池化,具体结构如表 1 所示.

通过对距离-速度图进行三维卷积和池化获取连续帧的特征图,然后通过两层全连接层获取 1024×1 维的距离-速度特征向量.

表 1　距离-速度特征提取网络配置

类型	卷积核/步长	输出大小
卷积 1-1	64×3×3×3	64×32×64×64
池化 1	1×2×2	64×32×32×32
卷积 2-1	128×3×3×3	128×32×32×32
池化 2	2×2×2	128×16×16×16
卷积 3-1	256×3×3×3	256×16×16×16
卷积 3-2	256×3×3×3	256×16×16×16
池化 3	2×2×2	256×8×8×8
卷积 4-1	512×3×3×3	512×8×8×8
卷积 4-2	512×3×3×3	512×8×8×8
池化 4	2×2×2	512×4×4×4
卷积 5-1	512×3×3×3	512×4×4×4
卷积 5-2	512×3×3×3	512×4×4×4
池化 5	2×2×2	512×2×2×2

3.2 角度特征提取

本文以 VGG-16[18] 作为角度特征提取的基础结构. 由于图像维度较小,相较于光学的 RGB 手势图,角度时间图存在大片的空白信息,可以将 VGG-16 网络卷积层的两次卷积简化为一次. 根据卷积后特征图的变化,将卷积层由 5 层简化为 4 层,并采用 Relu 激活函数对卷积结果进行非线性化处理. 因此,提取角度时间图的卷积神经网络包括 4 个卷积池化层和 1 个全连接层,具体结构如表 2 所示.

表 2 角度特征提取网络配置

类型	卷积核/步长	输出大小
卷积 1	64 ×3 ×3	64 ×64 ×64
池化 1	2 ×2	64 ×32 ×32
卷积 2	128 ×3 ×3	128 ×32 ×32
池化 2	2 ×2	128 ×16 ×16
卷积 3	256 ×3 ×3	256 ×16 ×16
池化 3	2 ×2	256 ×8 ×8
卷积 4	512 ×3 ×3	512 ×8 ×8
池化 4	2 ×2	512 ×4 ×4

通过对角度时间图卷积和池化获取不同手势角度时间图的特征,再通过一层全连接层获取 1024 ×1 维的角度特征向量.

3.3 特征融合时序信息提取

由距离-速度图和角度时间图各自生成 1024 ×1 维特征,将两组向量并联组成步长为 1024、每一步维度为 2 的融合特征. 由此生成的融合特征在每一步包含了手势的距离、速度和角度信息. 融合特征中包含手势的时间顺序,所以每一步的信息并非相互独立. 为了有效利用融合特征中的每步之间的联系,本文采用 LSTM[10] 网络进行特征提取. LSTM 网络由 LSTM 单元构成,而 LSTM 单元包含记忆单元、忘记门、输入门以及输出门. 将融合特征的每一步输入到 LSTM 单元,其中细胞状态存储了前几步的信息,并且决定了当前这一步的输出,由此保留了融合特征在每一步之间的联系,生成最终的时序特征向量.

首先,每一步的特征信息通过忘记门 f_t 决定从单元状态中去除 C_{t-1} 的部分信息,再通过输入门 i_t 决定在单元状态中存储输入的信息 $\tilde{C}_t$,如式(11):

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \end{cases} \tag{11}$$

其中,$\sigma(\cdot)$ 表示 sigmoid 函数,$\sigma(x) = \dfrac{1}{1+e^{-x}}$,$W_f$、$W_i$ 和 W_C 是 LSTM 单元中的权重,b_f、b_i 和 b_C 是对应的偏置.

最后,将输入门的结果通过输出门得到隐层状态 h_t 和输出信息 o_t,如式(12):

$$\begin{cases} o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \tag{12}$$

此外,LSTM 可以将误差保留在更为恒定的水平,从而避免梯度消失造成记忆单元中状态信息失效的问题. 因此,本文将两个向量并联获取一个步长为 1024 的矩阵,再通过 LSTM 将特征中累积的时序信息提取出来,起具体处理过程如图 5 所示.

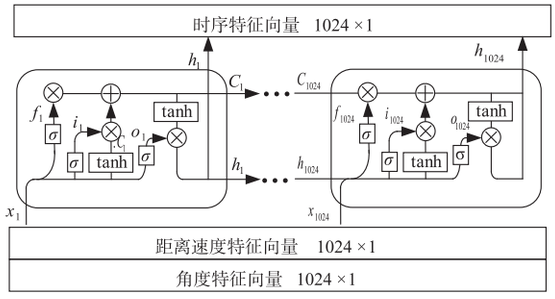

图5 融合特征处理过程

3.4 模型训练

将手势特征提取完毕后,添加全连接层将手势特征映射到标本标记空间. 由于融合特征由不同的网络生成,在全连接层前需要进行归一化处理. 全连接层生成的特征向量输入到归一化指数函数(13)中进行分类.

$$\mathrm{softmax}(z) = \frac{\exp(\theta_i^{\mathrm{T}} z_i)}{\sum_{j=1}^{k} \exp(\theta_j^{\mathrm{T}} z_j)} \tag{13}$$

其中,i 为第 i 类手势,k 为手势种类,本文中 $k=6$,z_i 为特征向量的第 i 个元素,θ_i 为 z_i 对应的权重.

在 3D-CNN 网络中,由于网络过深会造成过拟合,因此本文采用 L2 正则化来惩罚特征权重. 在网络训练中,本文使用交叉熵损失函数[10] 作为目标函数. 此外,本文采用指数衰减法,在训练的初步阶段选择较大的学习率使网络的损失值快速地收敛到较小值,并随着学习率的指数级减小,网络模型的损失值逐渐趋于平稳.

4 实验分析与讨论

4.1 实验平台

本文雷达平台为德州仪器 AWR1642 单芯片 FMCW 雷达传感器. 在实验中,从雷达传感器采集到信号源数据后传输到 PC,使用 Matlab 软件进行信号处理,然后在配置为 Intel-6700K 处理器和 NVIDIA-GTX1080 显卡的服务器上利用 Tensorflow 深度学习框架进行训

练,如图 6 所示.

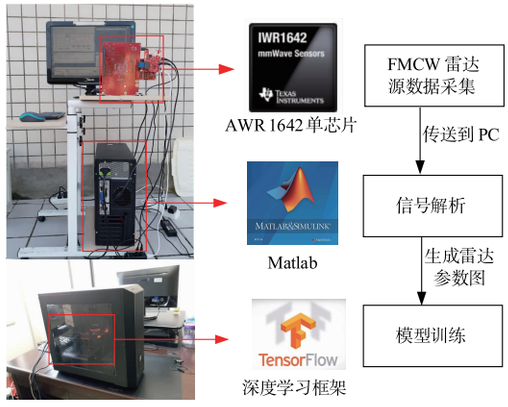

图6　实验平台

4.2　实验数据

本文在建立特征提取的深度学习网络中,自建手势信号数据集进行验证. 本文共设计 10 类手势,具体手势动作如图 7 所示,每一类手势重复 400 次,共 4000 个手势雷达数据.

图7　手势种类

在数据处理后,将数据集分为训练集和验证集,输入到 TS-FNN 中进行训练和测试.

4.3　实验结果

4.3.1　网络训练过程

为比较 TS-FNN 网络的性能,本文在不同的初始学习率下进行模型训练,如图 8 所示. 初始学习率为 0.05 时,网络权重每次更新过大,误差逐渐增大,导致模型不收敛. 在初始学习率为 0.008 时,由于每次更新幅度太大而陷入局部最小值,无法得到全局最优解. 当初始学习率太小,如 0.001 时,网络权重更新太慢. 在学习率为 0.005 时,识别的效果最好,准确率最高时达到 93.75%. 因此,本文选取学习率为 0.005.

本文采用了指数衰减法对 TS-FNN 模型的学习率进行更新,设置的学习衰减率越小,学习率每次衰减越大. 本文对比了不同衰减率在不同迭代步数下的准确率,如图 9 所示. 当学习率每次衰减过大如 0.79 时,迭代到一定步数后,学习率衰减到接近于零. 此时,模型权重不再继续更新,识别准确率也不再提升. 因此,学习衰减率为 0.89 和 0.99 时,准确率都有明显的提升. 当学习率衰减率进一步增加为 0.995 时,学习率衰减太慢,准确率相较于 0.89 和 0.99 时有所下降. 当衰减率

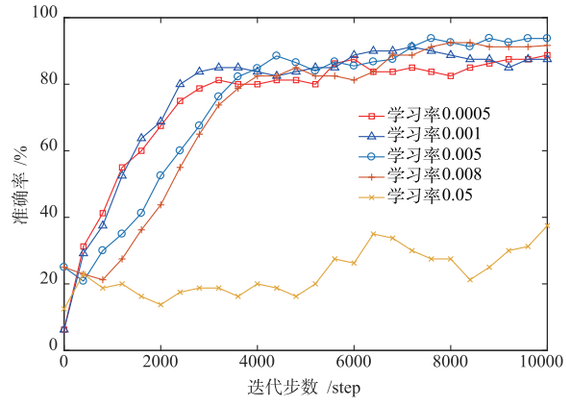

图8　不同初始学习率下模型的准确率

为 0.99 时,既不会因为学习率衰减太快使损失函数迭代到一定步数就不再下降,也不会因为衰减太慢而增加训练时间. 因此,衰减率为 0.99 时,识别效果最好,本文选取该值作为学习衰减率.

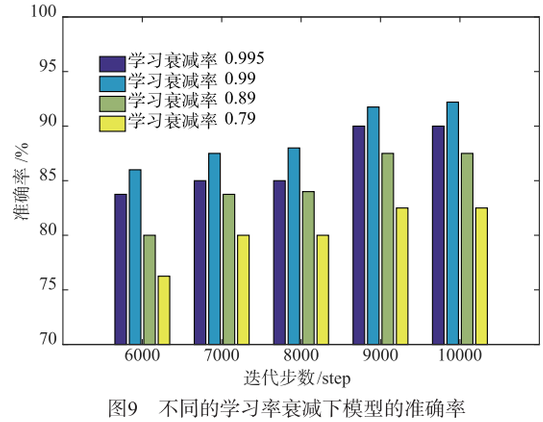

图9　不同的学习率衰减下模型的准确率

4.3.2　手势识别性能分析

为了验证 3D-CNN 和 CNN 能够有效提取距离-速度图和角度时间图的特征,将距离-速度参数图输入到文献[10]中的 3D-CNN 网络和角度时间图直接输入到文献[18]中的 VGG-16 网络,能够单独进行特征提取和分类,本文将这两个网络与提出的 TS-FNN 网络进行了对比实验.

图 10 对比了各个网络训练过程的损失函数值. 由图可以看出,随着迭代步数的增加,误差逐渐减小. 当训练到第 6000 步时,TS-FNN 已收敛,误差值接近于 0. 而文献[10]和文献[18]的误差值较大且不稳定,仍然需要继续训练才能进一步减小损失函数值. 因此,本文提出的 TS-FNN 比文献[10]和文献[18]收敛速度更快.

将训练集输入到 TS-FNN 模型中迭代训练,图 11 为 TS-FNN、文献[10]和文献[18]在不同迭代步数时所存储模型的识别准确率. 由于文献[10]模型较为复杂,在前 6000 步迭代中模型收敛较慢,其准确率比文献[18]低.

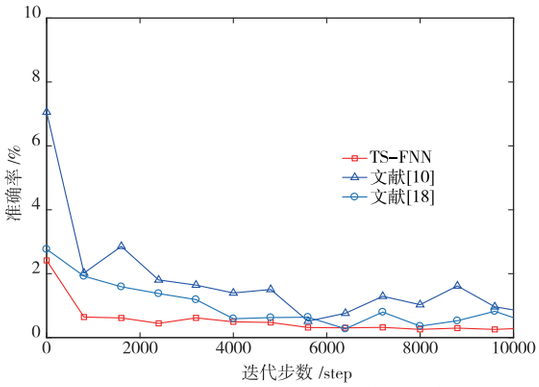

图10 TS-FNN、文献[10]和文献[18]的损失函数

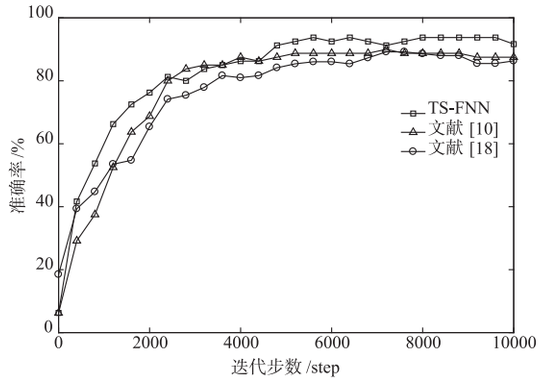

图11 TS-FNN、文献[10]和文献[18]的准确率

由于文献[10]结合距离和速度两个参数,在迭代8000步后其准确率略高于文献[18].在整个迭代过程中,文献[10]的准确率可达89.83%,而文献[18]的准确率可达到91.33%,表明了本文生成的距离-速度图和角度时间图都可以从运动参数上表示每种手势.本文提出的TS-FNN准确率可达到93.75%,相比于文献[10]和文献[18]网络的准确率分别提高了4.4%和2.6%.

本文CNN deep和3D-CNN shallow分别与本文中TS-FNN网络的CNN和3D-CNN架构进行组合,共生成了四种组合方式,分别为:CNN+3D-CNN shallow、CNN deep+3D-CNN shallow、CNN+3D-CNN(原始TS-FNN)以及CNN deep+3D-CNN.表3为文献[10]、文献[18]和不同组合下形成的TS-FNN对每种手势识别的准确率.

表3 手势识别准确率/%

网络结构	左划	右划	左右划	右左划	前推	后拉	推拉	拉推	推左	推右	平均
CNN+3D-CNN shallow	69.57	74.95	71.68	77.05	61.98	70.83	68.80	69.17	77.74	67.50	70.93
CNN deep+3D-CNN shallow	88.47	89.08	79.73	89.02	89.48	58.39	90.13	66.46	72.63	79.11	80.25
CNN+3D-CNN	92.19	89.54	92.96	91.57	94.12	95.73	93.45	90.21	89.71	91.12	92.06
CNN deep+3D-CNN	92.66	91.18	94.97	89.43	93.76	96.88	92.68	92.70	90.59	92.31	92.72
文献[10]	84.66	82.39	86.72	90.26	93.67	92.61	89.02	89.54	86.43	82.78	87.81
文献[18]	93.16	90.48	91.97	89.15	86.85	82.22	81.81	88.51	87.55	82.88	87.46

由表3可知,在不同组合的TS-FNN中,网络越深识别效果越好.但是由于角度时间图较为简单,CNN+3D-CNN生成的TS-FNN准确率已经达到了92.06%.而使用加深的CNN deep时,CNN deep+3D-CNN的平均准确率(92.72%)相比于CNN+3D-CNN只有很小的提升(约0.7%),所以本文采用较浅的CNN网络.另一方面,由于距离-速度图包含了32帧图像,包含的图像较多,使用CNN提取角度特征时,利用3D-CNN shallow提取距离-速度特征的平均准确率只有80.25%,而利用3D-CNN的平均准确率有大幅度提升(达92.06%).所以本文采取网络较浅的CNN和较深的3D-CNN网络组成TS-FNN.

在文献[10]训练距离-速度图的模型中,对只有横向变化的手势——左划、右划和左右划三个手势的识别效果较差,所有手势的平均识别准确率为87.81%.同时,在文献[18]训练角度时间图的模型中,对只有径向变化的手势——前推、后拉和推拉动作识别也不理想,对10种手势识别的平均识别准确率为87.46%.对于一些较为复杂和相似的手势,如左右划和右左划,推拉和拉推,以及推左和推右,在文献[10]和文献[18]中识别效果较差,均只有80%左右.而TS-FNN将手势动作径向的距离和速度信息以及横向的角度信息提取相应参数图特征,并且进行了特征融合和时序信息提取.因此,相比于文献[10]和文献[18]方法,每一类手势识别的准确率均在89%以上,十类手势的平均识别准确率为92.06%(提升了5%).

5 结束语

本文提出了一种基于FMCW雷达信号的TS-FNN手势识别方法.该方法利用手势信号生成距离-速度参数图和角度时间参数图并建立TS-FNN进行特征提取和融合,保留了手势横向和纵向运动参数的时序特征.实测结果表明,所提TS-FNN方法的平均手势识别准确率达92%以上,从而验证了融合径向和横向参数的深

度特征能够有效提高手势识别精度.

参考文献

[1] COELHO Y L,SALOMAO J M,KULITZ H R. Intelligent hand posture recognition system integrated to process control[J]. IEEE Latin America Transactions,2017,15(6):1144 – 1153.

[2] 冯志全,杨学文,徐涛,等. 结合手势二进制编码和类-Hausdorff 距离的手势识别[J]. 电子学报,2017,45(9):2281 – 2291.
FENG Zhi-quan,YANG Xue-wen,XU Tao,et al. Gesture recognition combining gesture binary coding and class-Hausdorff distance[J]. Acta Electronica Sinica,2017,45(9):2281 – 2291. (in Chinese)

[3] PAI N S,HONG J H,CHEN P Y,et al. Application of design of image tracking by combining SURF and TLD and SVM-based posture recognition system in robbery pre-alert system[J]. Multimedia Tools and Applications,2017,76(23):25321 – 25342.

[4] KHAN F,LEEM S K,CHO S H. Hand-based gesture recognition for vehicular applications using IR-UWB radar[J]. Sensors,2017,17(4):833.

[5] ZHOU Z,CAO Z,PI Y. Dynamic gesture recognition with a terahertz radar based on range profile sequences and doppler signatures[J]. Sensors,2017,18(1):10.

[6] WANG W,LIU A X,SHAHZAD M,et al. Device-free human activity recognition using commercial WiFi devices[J]. IEEE Journal on Selected Areas in Communications,2017,35(5):1118 – 1131.

[7] LI Y,WANG X,LIU W,et al. Deep attention network for joint hand gesture localization and recognition using static RGB-D images[J]. Information Sciences,2018,441:66 – 78.

[8] MOLCHANOV P,GUPTA S,KIM K,et al. Hand gesture recognition with 3D convolutional neural networks[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)[C]. Boston:IEEE Computer Society,2015. 1 – 7.

[9] ZHU G,ZHANG L,SHEN P,et al. Multimodal gesture recognition using 3D convolution and convolutional LSTM[J]. IEEE Access,2017,99:1 – 1.

[10] TRAN D,BOURDEV L,FERGUS R,et al. Learning spatiotemporal features with 3D convolutional networks[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. Washington:IEEE Computer Society,2015. 4489 – 4497.

[11] WANG S,SONG J,LIEN J,et al. Interacting with Soli:exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum[A]. Proceedings of the 29th Annual Symposium on User Interface Software and Technology[C]. Tokyo:ACM,2016. 851 – 860.

[12] LI G,ZHANG R,RITCHIE M,et al. Sparsity-based dynamic hand gesture recognition using micro-doppler signatures[A]. Proceedings of the 2017 IEEE Radar Conference[C]. Seattle:IEEE,2017. 0928 – 0931.

[13] 郭跃,王宏迺,周陬. 阵元间距对 MUSIC 算法的影响[J]. 电子学报,2007,35(9):1675 – 1679.
GUO Yue,WANG Hong-xun,ZHOU Zou. The influence of element spacing on MUSIC algorithm[J]. Acta Electronica Sinica,2007,35(9):1675 – 1679. (in Chinese)

[14] LIN J J,LI Y P,HSU W C,et al. Design of an FMCW radar baseband signal processing system for automotive application[J]. SpringerPlus,2016,5(1):42.

[15] 王元恺,肖泽,许建中,等. 一种改进的 FMCW 雷达线性调频序列波形[J]. 电子学报,2017,45(6):1288 – 1293.
WANG Yuan-kai,XIAO Ze,XU Jian-zhong,et al. An improved FMCW radar linear frequency modulation sequence waveform[J]. Acta Electronica Sinica,2017,45(6):1288 – 1293. (in Chinese)

[16] JARDAK S,AHMED S,ALOUINI M S. Low complexity moving target parameter estimation for MIMO radar using 2D-FFT[J]. IEEE Transactions on Signal Processing,2017,65(18):4745 – 4755.

[17] PAN H,ZHANG F,SHI C,et al. High-precision frequency estimation for frequency modulated continuous wave laser ranging using the multiple signal classification method[J]. Applied Optics,2017,56(24):6956 – 6961.

[18] SIMONYAN K,ZISSERMAN A. Very Deep Convolutional Networksfor Large-Scale Image Recognition[OL]. https://arxiv.org/pdf/1409.1556.pdf,2014.

作者简介

王　勇　男,1987 年 9 月出生,云南昭通人. 重庆邮电大学讲师,IEEE 会员. 2010 年、2012 年和 2018 年分别获哈尔滨工业大学学士、硕士和博士学位. 主要研究方向为无线通信、能效优化、深度学习等.
E-mail:yongwang@cqupt.edu.cn

王沙沙　女,1992 年 9 月出生,重庆沙坪坝人. 2016 年在重庆邮电大学获工学学士学位. 现为重庆邮电大学硕士生,研究方向为手势识别技术.
E-mail:1259657562@qq.com