

# 基于改进和积网络的双人交互行为识别

陈昌红, 刘 园

(南京邮电大学 通信与信息工程学院, 江苏 南京 210003)

**摘 要:**受到视角变化、相机移动、尺度、光线、遮挡等因素的影响, 双人交互行为识别的效果往往不太理想。有效地提取特征和合理地建立交互模型是双人交互行为识别与理解的两个重要研究内容。基于深度学习的思想, 直接在三维空间中构建多层神经网络, 使用两层卷积叠加独立子空间分析网络提取视频的时空特征。在此基础上, 提出了一种基于改进和积网络(sum product networks, SPNs)的双人行为识别算法。通过改进后的 LearnSPN 结构学习算法学习和积网络的结构和权重, 在训练过程中对数据集进行实例划分或者变量划分直至满足划分结束条件, 从而实现对双人交互行为的分类。该方法在 UT、BIT-Interaction 和 TV-human 交互数据库上进行测试, 实验结果证明了该方法对双人交互行为识别的有效性, 尤其对背景复杂的 TV-human 交互数据库效果更好。

**关键词:** 双人交互行为识别; 神经网络; 独立子空间分析; 和积网络; 结构学习算法

**中图分类号:** TN911.73

**文献标识码:** A

**文章编号:** 1673-629X(2019)10-0157-07

**doi:** 10.3969/j.issn.1673-629X.2019.10.031

## Human Interaction Recognition Based on Improved Sum Product Networks

CHEN Chang-hong, LIU Yuan

(School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Due to the influence of visual angle change, camera movement, scale, light and occlusion, the effect of human interaction recognition is always unsatisfactory. The most important research contents of human interaction recognition are effective feature extraction and reasonable interaction model establishment. Based on the idea of deep learning, multi-layer neural network is constructed in the three-dimensional space directly, and the spatial-temporal features of video are extracted by using two-layered convolution superimposed independent subspace analysis (ISA). Human interaction recognition algorithm is proposed based on improved sum product networks (SPNs). The improved LearnSPN structure learning algorithm is used to learn the structure and weight of SPNs. Instance partition or variable partition is implemented on the database in the training process until satisfying the ending condition. Therefore the human interaction recognition is realized. The method is tested on UT - Interaction, BIT-Interaction and TV-human Interaction database, and the results show its effectiveness, especially for the TV-human Interaction database with complex background.

**Key words:** human interaction recognition; neural network; independent subspace analysis; sum product networks; structure learning algorithm

### 1 概 述

利用信息技术自动地对视频中的人体行为进行识别, 是近年来计算机视觉领域的热点和重点问题。随着计算机视觉领域相关技术的发展和深度学习技术的兴起, 对理想环境下的人体行为识别取得了较高的识别准确率, 且对较复杂环境下人体行为的识别准确率也获得了一定程度的提升。但人体运动的高复杂性和

多变性使得识别的精准性和高效性很难满足实际的应用要求。

在视频中提取有效的人体行为特征, 以往的传统特征如基于人体几何特征<sup>[1]</sup>、运动信息特征<sup>[2]</sup>; 具有先验信息的多尺度 SIFT<sup>[3]</sup>、HOG<sup>[4]</sup> 特征、稀疏表征<sup>[5-6]</sup>; 基于时空兴趣点的视频序列信息的特征<sup>[7-10]</sup>; 在单人行为识别的基础上加入语义信息进行双人行为

收稿日期: 2018-09-27

修回日期: 2019-01-30

网络出版时间: 2019-04-24

基金项目: 国家自然科学基金(61471202)

作者简介: 陈昌红(1982-), 女, 副教授, 硕导, 研究方向为视频分析、模式识别; 刘 园(1992-), 女, 硕士, 研究方向为交互行为视频分析。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1004.010.html>

识别<sup>[11-12]</sup>;也有的文献将上述特征进行加权融合用于识别<sup>[13]</sup>。手工特征常常需要依靠研究者长时间的工作经验来人为设定,该特征对于不同视频的特征提取在一定程度上有局限。在目前的研究中,手工特征无法适用于不同复杂程度的数据库。在实际应用中,需要提出一个能从视频自身的信息中提取有效的、能够有效表征视频的特征。近年来,深度学习在目标识别方面表现优异<sup>[14-15]</sup>,可以从原始视频或者图像中直接自动学习到特征。神经网络学习的特征与传统的特征不同,神经网络特征不需要人为手动选定,且该特征在自适应性、不变性和通用性方面有良好的性能,在检测、分割和识别等多个计算机视觉方面应用广泛。独立子空间分析<sup>[16]</sup>(independent subspace analysis, ISA)是 2011 年提出的一种神经网络特征,ISA 结构能从视频数据中自动学习提取出具有不变性的特征。翟涛等<sup>[17]</sup>采用 ISA 构建两层卷积叠加网络,用于提取深层不变性特征,文中也采用了这种特征提取方法。

行为分类阶段训练性能优越、分类精准性高的分类器,能有效快速处理数据。图模型常常用来对大数据建模,对复杂的分布图模型可以很简洁地表达,但是它的难点在于模型结构的参数学习在一定程度上比较

复杂,模型的推导也具有一定的难度。这是由于图模型在做归一化运算时,其计算量会急剧增大。在实际应用中,图模型在做推理时,预估最差情况下的计算量会有指数级别。深度结构可以看成是一个有着多隐层变量的图模型,使用深度结构能够有效表达大部分的分布,但是非凸似然估计和高复杂度的推理使深度网络学习非常困难。稀疏连接树、混合模型等图模型在推理上可行,但是能表示的分布非常有限。Poon 和 Domingos<sup>[18]</sup>在 2011 年提出了和积网络结构(sum product networks, SPNs),SPNs 属于深度结构,但模型能进行有效的推理计算。SPNs 的结构属于有向无环图,主要由叶节点、sum 节点、product 节点及节点边上的权重值(权重值均为非负数)组成。SPNs 目前主要应用于图像复原、图像分类,并取得了一定的进展和成果,但先前的研究都是基于给定了初始的 SPNs 结构。为了使 SPNs 不需要预先定义的结构,并且学习更加灵活减少结构的冗余度,Gens 和 Domingos<sup>[19]</sup>在 2013 年提出了一种 SPNs 结构学习算法,该算法将实例集进行实例划分和变量划分并学习结构权重参数,得到自主学习的 SPNs 结构。

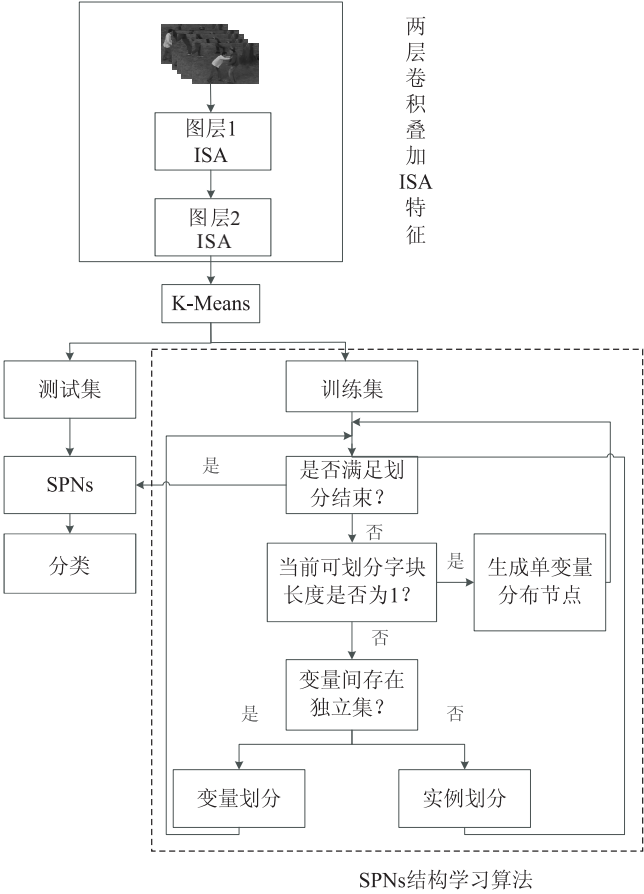


图 1 双人行行为识别算法框架

文中提出一种基于改进和积网络的交互行为分类算法。整个过程分为特征提取、训练 SPN 模型和 SPN

模型分类三个步骤。首先对视频进行密集采样,然后从每一帧中随机选取中心点的位置提取出相同大小的

视频块,得到视频时空样本。使用两层卷积叠加 ISA 网络提取视频特征。然后使用 K-means 方法进行特征聚类,映射为视觉单词。训练视频样本通过改进后的 LearnSPN 算法框架学习 SPN 的结构和权重,在训练过程中对数据集进行实例划分或者变量划分直至满足划分结束条件。在改进后的 LearnSPN 算法中,变量划分使用无向图结合并查集结构的 IndepGraph UnionFind 算法,实例划分使用更适合数据分布的 DBSCAN(density-based spatial clustering of application with noise)聚类算法。

双人行为识别算法框架如图 1 所示。

## 2 时空特征提取

### 2.1 ISA 网络结构

独立子空间分析(ISA)是一种无监督学习方法。ISA 的网络结构是一个两层网络,可以有效模拟人类视觉系统 V1 区简单细胞与复杂细胞感受的层次化响应模式。每个复杂细胞都接收一组简单细胞的输入数据,在此基础上构成子空间。网络结构如图 2 所示。层 1 对人脑视觉皮层简单细胞进行模拟,也称为简单单元。层 2 对人脑视觉皮层复杂细胞进行模拟,又称为合并单元。

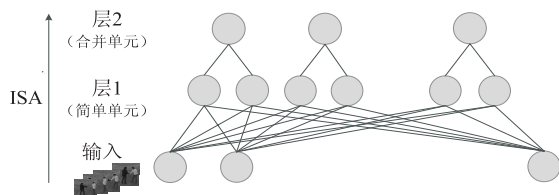


图 2 ISA 网络结构

对每一个输入样本  $x^t$ , ISA 获得特征进行变换的公式如下:

$$p_i(x^t; W, V) = \sqrt{\sum_{k=1}^m v_{ik} \left( \sum_{j=1}^n W_{kj} x_j^t \right)^2} \quad (1)$$

其中,  $x^t$  是 PCA 降维后均值为 0 的白化数据;  $W$  是输入数据和层 1 之间的权重,层 1 对输入进行加权平方,输出是对应的线性特征响应;  $V$  是层 2 和层 1 之间的权重,层 2 对层 1 输出的加权开方;  $p_i$  是同一个子空间特征的合并输出。

在训练阶段,权重  $V$  被设置为固定权重,在 ISA 的网络中只需优化权值参数  $W$ 。优化目标函数为:

$$\min_W \sum_{t=1}^T \sum_{i=1}^m p_i(x^t; W, V) \quad (2)$$

约束  $WW^T = 1$

其中,正交约束  $WW^T = 1$  保证了特征的不相关性,用于保证提取出的特征多样化。

### 2.2 两层卷积叠加 ISA 网络

在原始的 ISA 网络中,特征的概括能力与输入样

本的维度大小成正比,当样本维度越大时,提取出的特征表征能力越好。然而原始的 ISA 网络在训练过程所需的时间和输入数据维度成反比,当数据维度越大,花费的时间就越多。因为在每一步投影梯度下降时都要执行正交化,而正交化和输入维度的计算复杂度为  $O(n^3)$ 。在实际应用时,标准的 ISA 算法是针对二维图像数据而言的,当输入拓展到 3 维的视频数据中会遇到很大的困难。一个视频数据相当于一个长方体,展开成一个一维向量的维度非常大,时间复杂度呈 3 次方增加,因此计算速度非常慢。

为了处理高维数据,借助于卷积神经网络的结构,用卷积叠加提取视频的特征。把大的视频块拆分成不同的子块,以减少每次 ISA 计算维数,加快训练速度。图 3 用包含 ISA 和 PCA 的两层卷积叠加 ISA 网络对视频数据提取特征。图层 ISA1 的输入为通过密集采样获取的小视频块,通过训练学习得到 ISA 网络的参数,将该 ISA 网络进行平移复制,对多个小视频块提取出 ISA 特征,再将该特征通过卷积叠加融合成大视频块的表示,并输入到图层 ISA2 中得到视频的特征表示,在 ISA2 的输入前可以通过 PCA 降低输入数据的维度。把 ISA1 的输出通过 PCA 降维串联 ISA2 的输出作为最终视频的特征。

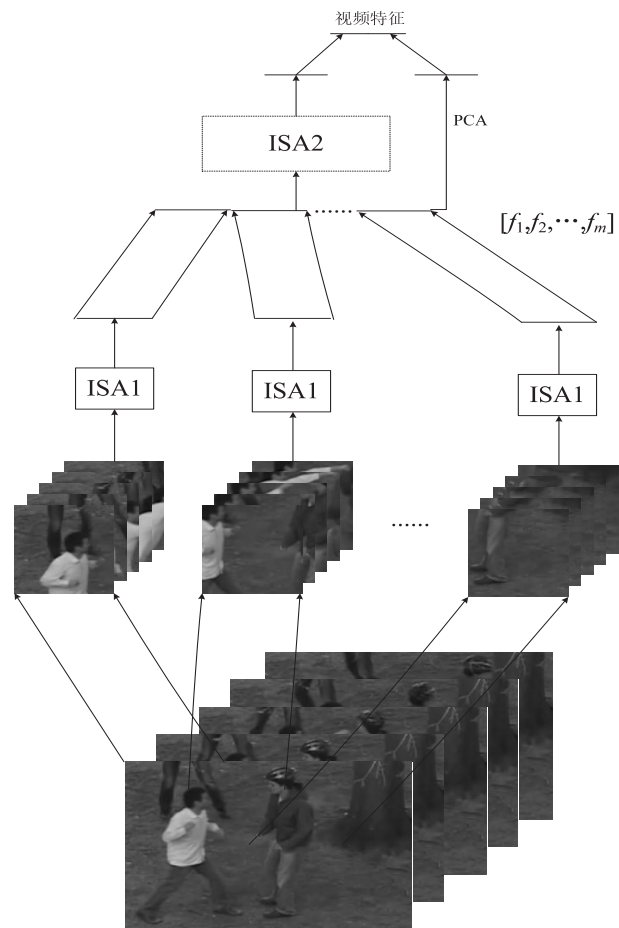


图 3 两层卷积叠加 ISA 网络

### 3 基于改进 SPNs 结构学习的视频分类

在 SPNs 的学习算法中主要有生成式和判别式学习算法,这两种算法都存在一个缺点,即需要预先定义 SPNs 的初始结构,再通过不同的学习算法学习 SPNs 的权重参数从而得到完整的 SPNs 结构。模型的层数越深,其表达能力也就越强。生成式的学习算法具有学习很深的 SPN 的能力,但预先定义的 SPNs 结构也拥有巨大的节点数量,在学习时需要较大的计算量和复杂的推导。判别式学习在分类问题上有最优的结果,但需要在灵活度和学习消耗上进行平衡。Gens 和 Domingos<sup>[19]</sup>充分利用 SPNs 的表达能力,在 2013 年提出了 SPNs 结构学习的第一个算法 LearnSPN。该算法不需要预先定义 SPNs 结构,根据输入的数据集使用该算法自动生成一个 SPN,生成的结构能够减少重复的节点,使计算和推理都更加灵活。

文中的研究内容是结合 SPNs 的结构和相关理论知识,在 SPNs 结构学习算法 LearnSPN 的基础上进行分析、理解和研究,从而进行改进,设计一种新的实例划分算法和变量划分算法,将学习到的 SPNs 在视频数据库上进行分类。

#### 3.1 SPNs 的结构学习算法

将数据集可视化为例证形式,其中行是实例列是变量,变量的样本是独立同分布的。根据算法横向或者纵向划分数据集。如果变量向量是单位长度,该算法返回相应的单变量分布,并使用 MPA 方法评估相应的参数。算法在划分行数或是划分列数上递归进行,直至变量向量的长度小于等于 1。数据集垂直划分的规则是,数据集中存在相互独立的子集,在数据集划分出多个互不相干的集合,形成 product 节点,且节点的值是所有子集的乘积。数据集水平划分的规则是,数据集中存在相似的实例,将其进行聚类划分把相似的实例划分到同一子集中,形成 sum 节点,且节点的值是在所有子集上的加权和,每个 SPN 的权值是相应子集的比例值。重复该过程直至达到递归停止条件。

LearnSPN 算法可以分为三个部分。第一部分为检测变量独立和判断是否划分变量  $V$  来创造新 Product 节点。如果它能够变量  $V$  拆分为相互独立的子集,算法在独立子集递归,并返回生成 SPNs 的乘积。第二部分为,如果变量  $V$  划分失败,必须从数据中找到可能聚类。它将实例聚类到相似的子集合中,算法在独立子集递归,并返回结果的 SPNs 的加权总和。SPN 的权重是对应相应子集中的实例的比例,也可以使用 Dirichlet 先验进行平滑处理。最后,如果向量范围是单一变量,LearnSPN 返回相应的单变量的概率分布。

可以使用不同的算法进行变量划分和实例划分。Gens 和 Domingos 提出了使用混合 EM 算法,通过将实例划分到它最后可能的聚类中来学习 Sum 节点;图模型的结构学习 Product 节点。

#### 3.2 改进后的 LearnSPN 算法

##### (1) 变量划分。

变量划分的核心是寻找变量间的独立子集,每一个子集中的元素独立于其他子集。假设给定变量  $V$ ,找到子集合  $P = \{P_1, P_2, \dots, P_n\}$  当  $P_i \cap P_j = \emptyset$  且  $P_1 \cup P_2 \cup \dots \cup P_n = V$ 。变量两两间的关系采用无向图来记录。无向图  $G = (V, E)$  存在若干顶点  $V$  和边  $E$ 。对于顶点  $i, j \in V$  且  $i$  不垂直于  $j$ , 则  $i, j$  间存在一条边  $e_{ij}$ 。因为当两个变量不独立时,图中两顶点间存在一条边,为了找到一个独立子图,在这个子图中的所有顶点没有路径到达另一子图。无向图使问题降低到了寻找连接子图问题。

并查集的数据结构常用于解决动态连通性问题。并查集是一种树型数据结构,用于处理不相交集合并以及查询。首先让所有元素独立成树,也就是根节点的树;然后根据需要将关联元素合并;合并方式是将一棵树最原始的节点的父亲索引指向另一棵树。使用按秩合并和路径压缩结合的并查集(union-find)结构寻找到连接子图。

文中使用并查集实现变量划分算法。初始化时每个顶点都在自己的不相交集合中,等价于无向图  $G$  的边是空的。遍历每个变量,验证一对变量  $X, Y \in V$  是否在同一集合。若  $u_x = u_y$ , 表示  $X, Y$  在同一集合,且  $X, Y$  间存在边  $e_{xy}$ ; 若  $u_x \neq u_y$ ,  $\Omega(X, Y)$  对变量间的独立性判断,采用的是 G-test 度量方法:

$$G(x_1, x_2) = 2 \sum_{x_1} \sum_{x_2} c(x_1, x_2) \cdot \log \frac{c(x_1, x_2) \cdot |T|}{c(x_1)c(x_2)} \quad (3)$$

其中和项是取每个变量上的取值,且  $c(\cdot)$  表示一对变量或是单个变量的某个设置的出现次数。改进后的 IndepGraphUnionFind 变量划分算法,将通过 Union-Find 寻找最小独立图替代寻找完全独立图。

##### (2) 实例划分。

对于实例划分,希望将相似的实例划分到同一子集中。LearnSPN 算法中的 EM 算法详细而准确,但是计算复杂、收敛速度慢,不适合大规模数据集和高维数据集。实践中,简单的 K-means 算法就能够进行实例划分。但是 K-means 算法存在缺陷,必须实现人工制定聚类的个数。DBSCAN 是基于密度的聚类算法,簇的个数由算法自动决定。DBSCAN 算法可以发现任意形状的聚类簇,并且可以在聚类时找出异常点。



文中对实例划分与变量划分均使用新的算法, DBSCAN 是基于密度聚类的,它拥有可以对抗噪声,能处理任意形状和大小的簇的优良特性。IndepGraphUnionFind 算法花费线性内存。

4 视频分类实验及结果分析

基于双层卷积叠加 ISA 的时空特征和 SPNs 结构学习的视频分类实验将在 UT - interaction、BIT - interaction、TV Human interaction 三个交互数据库上进行。在进行实验时,ISA1 网络提取的视频块大小为  $16 * 16 * 10$ ,子空间步幅为 2,提取 200 维的特征向量。ISA2 网络的视频块大小为  $20 * 20 * 14$ ,子空间步幅为 4,提取 50 维特征。将 IAS1 的特征进行 PCA 降维,联合 ISA2 的输出作为最终的 150 维特征向量。对三个数据库,文中的训练测试比为 9 : 1,从数据库中随机抽取 0.9 的视频作为训练集,剩下的作为测试集。在给定训练测试数量比的情况下,随机抽取 10 个训练集测试集,每次抽取彼此相互独立。最后的训练测试数量比对应的分类准确率是 10 次抽取结果的平均值。

(1) UT-interaction 数据集。

UT-interaction 数据集将两个集合的视频序列混合使用增大数据集的个数,以及增强数据集的多样性。数据集中包含有 20 组人的 5 种行为。从图 4 可以看出,文中方法在双人行为数据集 UT-interaction 上的平均分类准确率达到 94%,其中大部分的识别结果都是正确的,其中拳击和握手容易混淆。这是因为拳击和握手两种交互行为模糊性大,当握手行为中一人的行为被遮挡就与拳击具有非常高的相似性。握手和踢打、拳击之间之所以混淆,是因为行为在外观特征和动作特征上十分相似,因此也会被错分造成错误的识别结果。

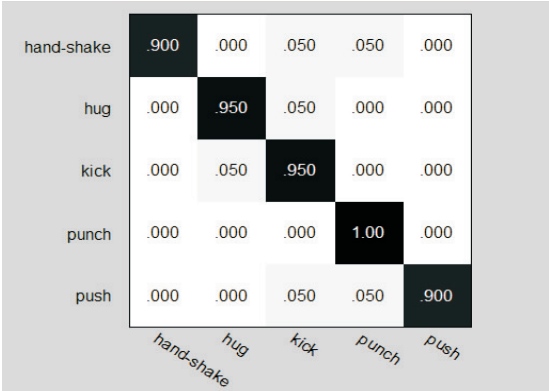


图 4 UT-interaction 数据集识别结果

在表 1 中,将文中算法与其他三种算法做比较。文献[20]中 Kong Yu 等主要考虑双人行为的交互,捕获行为各个部分的相互依赖关系;文献[21]中 Kong Yu 等主要提出一种新型的高层语义描述方法,用交互

式短语描述交互行为之间的关系;文献[22]中 Xu Wanru 等对于复杂的行为,考虑时间和空间特性,提出一种层次的时空模型来提取视频的高级特征表征。选取这三种算法是因为它们同时在 UT-interaction 数据集和 BIT-interaction 数据集上进行了实验,用于验证算法在不同复杂程度的双人交互行为视频上的识别结果。实验结果表明,文中方法在 UT-interaction 数据集上取得了不错的识别结果。

表 1 UT-Interaction 数据集对比结果

方法	准确率/%
文献[20]	85
文献[21]	91.67
文献[22]	94.17
文中算法	94

(2) BIT-interaction 数据集。

BIT-interaction 数据集在实验中包含 8 种双人交互行为,每种交互行为有 50 组视频。从图 5 中可以看出,平均分类准确率为 96.5%,取得了不错的识别结果。握手行为的错分率较高,是因为和击掌、推搡以及鞠躬之间有混淆,这是因为类类行为之间的相似性。

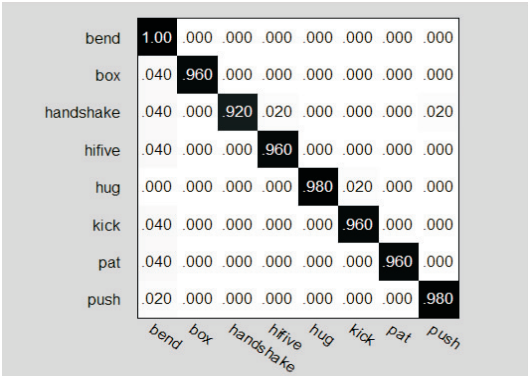


图 5 BIT-interaction 识别结果

在表 2 中,除了与文献[20-22]进行了比较以外,还与文献[23]的结果进行了比较。文献[23]采用的是深度学习算法 long-term residual recurrent network,结果表明文中算法有明显的优越性。

表 2 BIT-Interaction 数据集对比结果

方法	准确率/%
文献[20]	82
文献[21]	91
文献[22]	85
文献[23]	90
文中算法	96.5

(3) TV Human interaction 数据集。

TV Human interaction 数据集取自 23 个不同的电视节目,包含 4 种交互行为,每一组有 50 个视频片段。

从图 6 可以看出,文中方法在 TV Human interaction 数据集上的平均分类准确率为 80%。其中握手行为的识别率只有 78%,容易被错分为击掌;亲吻行为的识别率只有 80%,这是因为 TV Human interaction 数据集来自电视节目,视频中类间行为差异大,背景多样复杂。在视频的双人交互中会出现多个人,对识别有一定的干扰。

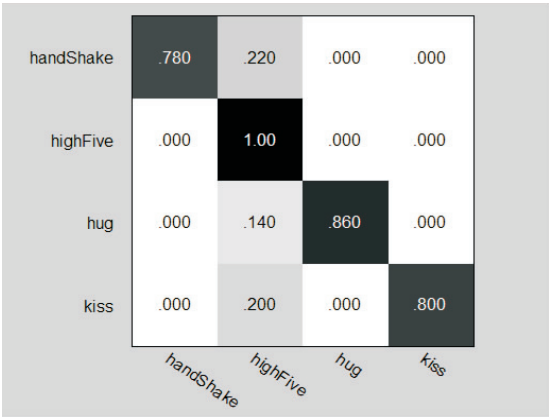


图 6 TV Human interaction 识别结果

文中算法与其他三种算法对比结果见表 3。由于 TV Human interaction 数据集比较复杂,一般算法都是只在这一个数据集上实验,虽然无法在 UT-interaction 数据集和 BIT-interaction 数据集上对比算法的性能,但对验证文中算法的普适性没有影响。在文献[24]中,Avgerinakis K 等提出一种运动补偿算法,准确定位活动空间;文献[25]中,Ke Qihong 等提出一种结合时间与空间信息的交互预测,在场景变化大的情况下有良好的表现;文献[26]中通过 Hough voting (HV) 检测对象与动作行为,对校训练样本也使用的算法。通过文中算法在识别结果上有较大的提升。

表 3 TV Human interaction 数据集对比结果

方法	准确率/%
文献[24]	46.6
文献[25]	64
文献[26]	66
文中算法	80

综合以上三个数据集上的实验结果表明,文中方法在不同复杂背景的双人行为视频中都取得了较好的识别结果。这是由于 ISA 网络提取的特征不仅能够反映全局的动作方向分布,同时也能探测到动作的改变以及检测移动边缘状态,这些信息对于视频有效信息的提取十分有帮助。SPNs 使用自动学习的改进后的 LearnSPN 算法结构,能正确表达视频信息,学习到的 SPNs 能很好地表达该视频类别。同时也验证了 SPNs 结构在视频分类上的效果,对 SPNs 结构学习改进算法,缩短了训练花费时间,同时减少了空间复杂度。

5 结束语

针对多复杂度的双人行为数据库识别问题,基于 ISA 特征和 SPNs 结构提出一种普适性的模型,提高对行为分类的鲁棒性。基于 ISA 网络所学习提取出的特征能基于数据自动提取特征,多层结构学习视频的高层特征,且具备传统特征不具备的鲁棒性。而 Sum-Product Networks 作为一个多隐层概率模型,对 SPN 学习算法进行了深入的理解和分析并提出适合文中数据库的改进,提出了一种 LearnSPN 学习算法,取得了不错的视频分类结果。但是学习得到的结构中有一些节点是相同的,整个 SPN 结构仍然有点冗杂。目前关于 SPNs 的研究较少,但 SPNs 在图像处理领域是非常具有价值的研究课题。

参考文献:

[1] FUJIYOSHI H, LIPTON A J. Real-time human motion analysis by image skeletonization[C]//Fourth IEEE workshop on applications of computer vision. Princeton, NJ, USA; IEEE, 1998: 15-21.

[2] CHAUDHRY R, RAVICHANDRAN A, HAGER G, et al. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions[C]//IEEE conference on computer vision and pattern recognition. Miami, Florida, USA; IEEE, 2009: 1932-1939.

[3] LOWE D G. Object recognition from local scale-invariant features[C]//Proceedings of the seventh IEEE international conference on computer vision. Kerkyra, Greece; IEEE, 1999: 1150-1157.

[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE conference on computer vision and pattern recognition. San Diego, CA, USA; IEEE, 2005: 886-893.

[5] 陈昌红, 张杰, 刘峰. 双人交互行为的稀疏表征方法[J]. 模式识别与人工智能, 2016, 29(5): 464-471.

[6] DOLLAR P, RABAUDE V, COTTRELL G, et al. Behavior recognition via sparse spatio-temporal features[C]//IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance. Beijing, China; IEEE, 2006: 65-72.

[7] 王世刚, 孙爱朦, 赵文婷, 等. 基于时空兴趣点的单人行为及交互行为识别[J]. 吉林大学学报: 工学版, 2015, 45(1): 304-308.

[8] RAPANTZIKOS K, AVRITHIS Y, KOLLIAS S. Dense saliency-based spatiotemporal feature points for action recognition[C]//IEEE conference on computer vision and pattern recognition. Miami, FL, USA; IEEE, 2009: 1454-1461.

[9] KNOPP J, PRASAD M, WILLEMS G, et al. Hough transform and 3D SURF for robust three dimensional classifica-

tion[ C ]//European conference on computer vision. Heraklion, Crete, Greece: Springer-Verlag, 2010: 589-602.

[ 10 ] 王佩瑶, 曹江涛, 姬晓飞. 基于改进时空兴趣点特征的双人交互行为识别[ J ]. 计算机应用, 2016, 36( 10 ): 2875-2879.

[ 11 ] 马 丽. 最大间隔方法及其在行为分析中的应用[ D ]. 南京: 南京邮电大学, 2016.

[ 12 ] 左鑫孟. 基于视频的双人交互行为识别与理解算法研究[ D ]. 沈阳: 沈阳航空航天大学, 2017.

[ 13 ] 金壮壮, 曹江涛, 姬晓飞. 多源信息融合的双人交互行为识别算法研究[ J ]. 计算机技术与发展, 2018, 28( 10 ): 32-36.

[ 14 ] WANG Xuanhan, GAO Lianli, SONG Jingkuan, et al. Beyond frame-level CNN: Saliency-aware 3D CNN with LSTM for video action recognition[ J ]. IEEE Signal Processing Letters, 2017, 24( 4 ): 510-514.

[ 15 ] JI Shuiwang, YANG Ming, YU Kai. 3D convolutional neural networks for human action recognition[ J ]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35( 1 ): 221-231.

[ 16 ] LE Q V, ZOU W Y, YEUNG S Y, et al. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis[ C ]//IEEE conference on computer vision and pattern recognition. Providence, RI, USA: IEEE, 2011: 3361-3368.

[ 17 ] 瞿 涛, 邓德祥, 刘 慧, 等. 多层独立子空间分析时空特征的人体行为识别方法[ J ]. 武汉大学学报: 信息科学版, 2016, 41( 4 ): 468-474.

[ 18 ] POON H, DOMINGOS P. Sum-product networks: a new deep architecture[ C ]//IEEE international conference on computer vision workshops. Barcelona, Spain: IEEE, 2012: 689-690.

[ 19 ] GENS R, DOMINGOS P. Learning the structure of sum-product networks[ C ]//International conference on machine learning. Atlanta, GA, USA: IEEE, 2013: 873-880.

[ 20 ] KONG Yu, LIANG Wei, DONG Zhen, et al. Recognising human interaction from videos by a discriminative model[ J ]. IET Computer Vision, 2014, 8( 4 ): 277-286.

[ 21 ] KONG Yu, JIA Yunde, FU Yun. Interactive phrases: semantic descriptions for human interaction recognition[ J ]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36( 9 ): 1775-1788.

[ 22 ] XU Wanru, MIAO Zhenjiang, ZHANG Xiaoping, et al. A hierarchical spatio-temporal model for human activity recognition[ J ]. IEEE Transactions on Multimedia, 2017, 19( 7 ): 1494-1509.

[ 23 ] ZHAO Yang, SUN Tanfeng, JIANG Xinghao, et al. Long-term residual recurrent network for human interaction recognition in videos[ C ]//International congress on image and signal processing, biomedical engineering and informatics. Datong, China: IEEE, 2017: 78-83.

[ 24 ] AVGERINAKIS K, ADAM K, BRIASSOULI A, et al. Moving camera human activity localization and recognition with motionplanes and multiple homographies[ C ]//IEEE international conference on image processing. Quebec City, Canada: IEEE, 2015: 2085-2089.

[ 25 ] KE Qihong, BENNAMOUN M, AN Senjian, et al. Leveraging structural context models and ranking score fusion for human interaction prediction[ J ]. IEEE Transactions on Multimedia, 2017, 20( 7 ): 1712-1723.

[ 26 ] YU Gang, YUAN Junsong, LIU Zicheng. Propagative Hough voting for human activity detection and recognition[ J ]. IEEE Transactions on Circuits & Systems for Video Technology, 2015, 25( 1 ): 87-98.

+++++

( 上接第 152 页 )

modelling of hand grip formation during reach to grasp[ J ]. Neurocomputing, 2007, 71( 1-3 ): 411-416.

[ 14 ] BULLOCK D, GROSSBERG S. Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation[ J ]. Psychological Review, 1988, 95( 1 ): 49-90.

[ 15 ] JEANNEROD M. The timing of natural prehension movements[ J ]. Journal of Motor Behaviour, 1984, 16( 3 ): 235-254.

[ 16 ] ULLOA A, BULLOCK D. A neural network simulating human reach grasp coordination by continuous updating of vector positioning commands[ J ]. Neural Networks, 2003, 16( 8 ): 1141-1160.

[ 17 ] FICUCIELLO F, PALLI G, MELCHIORRI C, et al. Postural synergies and neural network for autonomous grasping: a tool for dextrous prosthetic and robotic hands[ M ]. Berlin: Springer, 2013.

[ 18 ] VILAPLANA J M, CORONADO J L. A neural network model for coordination of hand gesture during reach to grasp[ J ]. Neural Networks, 2006, 19( 1 ): 12-30.

[ 19 ] KAWATO M, KURODA S, SCHWEIGHOFER N. Cerebellar internal models: implications for dexterous use of tools[ J ]. The Cerebellum, 2012, 11( 2 ): 325-335.

[ 20 ] JAEGER D. Cerebellar nuclei and cerebellar learning[ M ]//Handbook of the cerebellum and cerebellar disorders. Netherlands: Springer, 2013: 1111-1130.

[ 21 ] 阮晓钢, 张少白, 李欣源. 仿生机械臂的小脑控制模型和仿真[ J ]. 电子学报, 2007, 35( 5 ): 991-995.

[ 22 ] KAWATO M, GOMI H. A computational model of four regions of the cerebellum based on feedback-error learning[ J ]. Biological Cybernetics, 1992, 68( 2 ): 95-103.

[ 23 ] BULLOCK D, CISEK P, GROSSBERG S. Cortical networks for control of voluntary arm movements under variable force conditions[ J ]. Cerebral Cortex, 1998, 8( 1 ): 1-15.