



(12)发明专利

(10)授权公告号 CN 107194559 B

(45)授权公告日 2020.06.05

(21)申请号 201710335309.X

G06N 3/04(2006.01)

(22)申请日 2017.05.12

G06T 7/215(2017.01)

(65)同一申请的已公布的文献号
申请公布号 CN 107194559 A

(56)对比文件

(43)申请公布日 2017.09.22

CN 104217214 A,2014.12.17,

CN 106203283 A,2016.12.07,

(73)专利权人 杭州电子科技大学
地址 310018 浙江省杭州市下沙高教园区2号大街

CN 106407903 A,2017.02.15,

WO 2017031088 A1,2017.02.23,

审查员 王雪

(72)发明人 胡海洋 丁佳民 陈洁 胡华
程凯明

(74)专利代理机构 杭州君度专利代理事务所
(特殊普通合伙) 33240

代理人 杜军

(51)Int.Cl.

G06Q 10/06(2012.01)

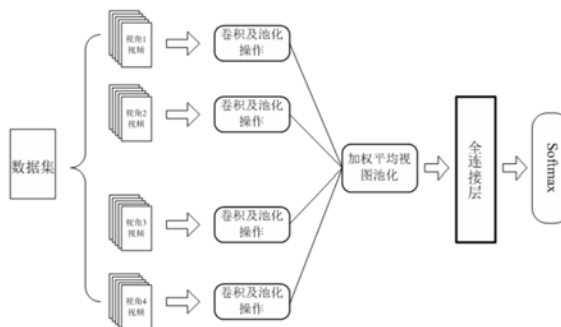
权利要求书2页 说明书7页 附图1页

(54)发明名称

一种基于三维卷积神经网络的工作流识别方法

(57)摘要

本发明公开了一种基于三维卷积神经网络的工作流识别方法。仅对不同工序任务进行事先划分并且在分析视频的过程中人为对不同动作为打上标签,这样不符合智能制造的自动化需求。本发明首先提出了一种带有自适应阈值的帧间差分法,该方法主要用于从复杂背景中分割出运动物体的区域,从而降低了后面特征提取与模型训练的时间复杂度;其次,对3D卷积神经网络进行了改进,使其能够充分适应具有多个监控设备的工厂环境,而对于不同视图,采用视图池化层对不同角度的视图按权重进行融合;最后,提出了一种新的动作划分方法,对视频中连续的生产动作进行自动划分,从而实现了自动化的工作流识别过程。



1. 一种基于三维卷积神经网络的工作流识别方法,其特征在於:该方法的具体步骤是:
步骤(1)、从数据集中导出包含多视角的工作流视频,获取各视角工作流视频的视频分辨率和帧数;

步骤(2)、初始化各视角工作流视频的帧间差分阈值;对各视角工作流视频分别进行步骤(3)~(11);

步骤(3)、设定 $t=2$;

步骤(4)、读取 $t-1$ 、 t 、 $t+1$ 三个连续的视频帧并将这三个视频帧进行灰度化和中值滤波处理;

步骤(5)、分别对前两帧和后两帧进行帧间差分运算得到两张帧间差分图;

步骤(6)、根据步骤(5)得到的两幅帧间差分图动态更新帧间差分阈值;帧间差分阈值动态更新方法如下:

6.1 设定 $l=1$,第 t 帧帧间差分阈值 $\tau_t^l = \frac{1}{2}(\max\{d_k\} + \min\{d_k\})$, d_k 为帧间差分图中第 k 个像素的像素值, $\max\{d_k\}$ 为帧间差分图中像素值的最大值, $\min\{d_k\}$ 为帧间差分图中像素值的最小值;

6.2 令 $\tau_t^{l+1} = \frac{1}{2}(\frac{\sum_{d_k \geq \tau_t^l} d_k}{N_1} + \frac{\sum_{d_k < \tau_t^l} d_k}{N_2})$, N_1 和 N_2 分别表示满足 $d_k \geq \tau_t^l$ 和 $d_k < \tau_t^l$ 的像素总数;

素总数;

6.3 若 $\tau_t^{l+1} = \tau_t^l$,则将 τ_t^{l+1} 赋值给 τ_t^l ,否则,令 $l=l+1$,重复步骤6.2;

步骤(7)、根据(6)得到的帧间差分阈值对当前帧进行二值化处理,大于帧间差分阈值的像素点设为1,小于帧间差分阈值的设为0;

步骤(8)、将前后两幅帧间差分图运行与操作,得到三帧差分图,并使用块提取方法获取兴趣点中心坐标;

步骤(9)、将提取到的兴趣点从当前帧原始图像中分割出来;

步骤(10)、 t 取值逐步加1,重复执行步骤(4)~(9),直到 t 取值比工作流视频最后一帧取值小1,重复过程中步骤(9)分割尺寸不变;将每次重复过程中步骤(9)中得到的兴趣点图像按先后顺序保存为兴趣点视频,并根据数据集中的分类规则对兴趣点视频进行分类;

步骤(11)、从步骤(10)得到的兴趣点视频中随机选取90%作为训练集,其余作为测试集;

步骤(12)、构建一个多视图三维卷积神经网络,初始化训练轮数为5000;多视图三维卷积神经网络构建方法如下:

12.1、卷积及池化操作如下:

①为第一卷积层初始化一个大小为 $9*9*9*10$ 的四维卷积核,激活函数为sigmoid,第一池化层窗口大小为2,步长为2;

②为第二卷积层初始化一个大小为 $9*9*7*30$ 的四维卷积核,激活函数为sigmoid,第二池化层窗口大小为2,步长为2;

③为第三卷积层初始化一个大小为 $9*8*5*50$ 的四维卷积核,激活函数为sigmoid,第三池化层窗口大小为2,步长为2;

④为第四卷积层初始化一个大小为 $4*3*3*150$ 的四维卷积核,激活函数为sigmoid,第四池化层窗口大小为2,步长为2;

12.2、初始化加权平均视图池化层中各特征图权重参数 α^{t_1} , α^{t_1} 为 $[0,1]$ 中随机值,且 $\sum_{t_1} \alpha^{t_1} = 1$;加权平均视图池化层中的加权平均视图池化操作如下:

$$a = \frac{\sum_{t_1} t_1 \exp^{\alpha^{t_1} \cdot z^{t_1}}}{\sum_{t_1} \exp^{\alpha^{t_1}}}$$

式中,a为加权平均视图池化操作后的加权平均特征图, t_1 为卷积及池化操作后的池化特征图序号, α^{t_1} 为序号 t_1 对应的池化特征图所占权重,exp表示e为底的指数函数, z^{t_1} 为序号 t_1 对应的池化特征图;

12.3、为前两层全连接层分别初始化一个 $3000*1500$ 和 $1500*750$ 的卷积核,并设置激活函数为Relu;加权平均视图池化操作后的加权平均特征图输入前两层全连接层;

12.4、为最后一层全连接层初始化一个 $750*14$ 的卷积核并设置Softmax分类函数;

步骤(13)、各视角 workflows 视频对应的训练集中均随机选取20个视频输入到步骤(12)中的多视图三维卷积神经网络中进行特征训练,并输出训练误差;

步骤(14)、各视角 workflows 视频对应的训练集中随机选取10个视频输入到多视图三维卷积神经网络中进行验证,得到多视图三维卷积神经网络分类识别的准确率;

步骤(15)、重复步骤(13)~(14),每重复一次训练轮数减1,直到训练轮数为0,得到一个训练好的多视图三维卷积神经网络;

步骤(16)、使用各视角 workflows 视频对应的测试集对步骤(15)中的多视图三维卷积神经网络进行测试;

步骤(17)、对新输入的工作流视频,获取视频分辨率和帧数,初始化帧间差分阈值;设定 $t=2$;

步骤(18)、根据步骤(4)~(8)提取相邻两帧兴趣点的中心坐标,并计算两个中心坐标间的距离,若距离大于设定的阈值T,则标记为运动状态 S_1 ,否则,标记为相对静止状态 S_0 ;

步骤(19)、t取值逐步加1,重复步骤(18)直到t取值比新输入的工作流视频最后一帧取值小1,统计连续的 S_0 和 S_1 的数量,当检测到的 S_0 或 S_1 数量大于或等于N时, $N>10$,分割出连续的 S_0 或 S_1 对应帧中目标兴趣点存储到帧队列中,否则丢弃连续的 S_0 或 S_1 对应帧;

步骤(20)、帧队列各个连续的 S_0 或 S_1 对应帧的集合中从第i帧开始提取连续的关键帧, $i>5$,使关键帧帧数与数据集中分类好的各段视频的帧数相同;

步骤(21)、将步骤(20)中关键帧按先后顺序组成的视频输入到步骤(15)中训练好的多视图三维卷积神经网络中对员工行为进行分类识别;

步骤(22)、将步骤(21)中得到的行为类别与预先定义的标准工作流进行比对。

一种基于三维卷积神经网络的工作流识别方法

技术领域

[0001] 本发明属于工作流识别技术领域,用于对生产制造流程的快速精确地识别与检测。

背景技术

[0002] 智能制造是制造自动化的进一步发展方向,它将人工智能技术广泛应用于工业制造过程的工程设计、工艺过程设计、生产调度、故障诊断等各个环节,从而实现制造过程智能化,并大幅度提高生产力。工作流识别(workflow recognition)作为智能制造的一个重要技术方向,目前已经引起产业界与科研界的重视。它利用制造车间里安装的摄像头,拍摄了生产线上生产调度的整个过程,然后对视频进行计算与处理,实现对工业生产流程进行快速精确的识别、检测,从而在保护员工人身安全,减少生产开销,保证产品质量,以及优化生产调度以及流程规范方面都将发挥重要的作用。

[0003] 然而,工作流识别技术有其复杂性与特殊性。首先,由于生产车间内各类机器、搬运车辆、辅助器械等物体较多,时常彼此遮挡,并且不同工序操作的相似性、车间内频繁的光线强弱变化,这些都给视频、图像的分析与识别带来了挑战。此外,动态的生产工作流程又使得识别过程相当复杂,很容易产生偏差:例如,工作流中不同的任务往往有不同的执行时间,并且在任务开始与结束之间没有一个明确的定义;这些任务甚至可能同时包含人和机器的动作,而其中一些与工作流无关的动作必须与真正的生产任务区分开来。这些方面使得传统的、依赖于目标物体检测与跟踪的动作/姿态识别方法很难适用于复杂的工厂制造环境。此外,目前一些研究者虽然对工作流识别技术展开了部分的研究,但如何对视频中的图像序列进行生产过程/动作的自动划分,这些研究工作并未给出明确的定义,它们大多仅是对不同工序任务进行事先划分并且在分析视频的过程中人为地对不同动作行为打上标签,这样显然不符合智能制造的自动化需求。

发明内容

[0004] 本发明针对目前的研究现状,提出了一个具有较强鲁棒性的工作流识别框架。在该框架中,首先提出了一种带有自适应阈值的帧间差分法,该方法主要用于从复杂背景中分割出运动物体的区域,从而降低了后面特征提取与模型训练的时间复杂度;其次,对3D卷积神经网络进行了改进,使其能够充分适应具有多个监控设备的工厂环境,而对于不同视图,采用视图池化层对不同角度的视图按权重进行融合;最后,提出了一种新的动作划分方法,对视频中连续的生产动作进行自动划分,从而实现了自动化的工作流识别过程。

[0005] 本发明方法的具体步骤是:

[0006] 步骤(1)、从数据集中导出包含多视角的工作流视频,获取各视角工作流视频的视分辨率和帧数;

[0007] 步骤(2)、初始化各视角工作流视频的帧间差分阈值;对各视角工作流视频分别进行步骤(3)~(11);

[0008] 步骤(3)、设定 $t=2$;

[0009] 步骤(4)、读取 $t-1$ 、 t 、 $t+1$ 三个连续的视频帧并将这三个视频帧进行灰度化和中值滤波处理;

[0010] 步骤(5)、分别对前两帧和后两帧进行帧间差分运算得到两张帧间差分图;

[0011] 步骤(6)、根据步骤(5)得到的两幅帧间差分图动态更新帧间差分阈值;帧间差分阈值动态更新方法如下:

[0012] 6.1 设定 $l=1$,第 t 帧帧间差分阈值 $\tau_t^l = \frac{1}{2}(\max\{d_k\} + \min\{d_k\})$, d_k 为帧间差分图中第 k 个像素的像素值, $\max\{d_k\}$ 为帧间差分图中像素值的最大值, $\min\{d_k\}$ 为帧间差分图中像素值的最小值;

[0013] 6.2 令 $\tau_t^{l+1} = \frac{1}{2}(\frac{\sum_{d_k \geq \tau_t^l} d_k}{N_1} + \frac{\sum_{d_k < \tau_t^l} d_k}{N_2})$, N_1 和 N_2 分别表示满足 $d_k \geq \tau_t^l$ 和

$d_k < \tau_t^l$ 的像素总数;

[0014] 6.3 若 $\tau_t^{l+1} = \tau_t^l$,则将 τ_t^{l+1} 赋值给 τ_t^l ,否则,令 $l=l+1$,重复步骤6.2;

[0015] 步骤(7)、根据(6)得到的帧间差分阈值对当前帧进行二值化处理,大于帧间差分阈值的像素点设为1,小于帧间差分阈值的设为0;

[0016] 步骤(8)、将前后两幅帧间差分图运行与操作,得到三帧差分图,并使用块提取方法获取兴趣点中心坐标;

[0017] 步骤(9)、将提取到的兴趣点从当前帧原始图像中分割出来;

[0018] 步骤(10)、 t 取值逐步加1,重复执行步骤(4)~(9),直到 t 取值比工作流视频最后一帧取值小1,重复过程中步骤(9)分割尺寸不变;将每次重复过程中步骤(9)中得到的兴趣点图像按先后顺序保存为兴趣点视频,并根据数据集中的分类规则对兴趣点视频进行分类;

[0019] 步骤(11)、从步骤(10)得到的兴趣点视频中随机选取90%作为训练集,其余作为测试集;

[0020] 步骤(12)、构建一个多视图三维卷积神经网络,初始化训练轮数为5000;多视图三维卷积神经网络构建方法如下:

[0021] 12.1、卷积及池化操作如下:

[0022] ①为第一卷积层初始化一个大小为 $9*9*9*10$ 的四维卷积核,激活函数为sigmoid,第一池化层窗口大小为2,步长为2;

[0023] ②为第二卷积层初始化一个大小为 $9*9*7*30$ 的四维卷积核,激活函数为sigmoid,第二池化层窗口大小为2,步长为2;

[0024] ③为第三卷积层初始化一个大小为 $9*8*5*50$ 的四维卷积核,激活函数为sigmoid,第三池化层窗口大小为2,步长为2;

[0025] ④为第四卷积层初始化一个大小为 $4*3*3*150$ 的四维卷积核,激活函数为sigmoid,第四池化层窗口大小为2,步长为2;

[0026] 12.2、初始化加权平均视图池化层中各特征图权重参数 α^{t_1} ， α^{t_1} 为[0,1]中随机值，且 $\sum_{t_1} \alpha^{t_1} = 1$ ；加权平均视图池化层中的加权平均视图池化操作如下：

$$[0027] \quad a = \frac{\sum_{t_1} \exp^{\alpha^{t_1}} \cdot z^{t_1}}{\sum_{t_1} \exp^{\alpha^{t_1}}}$$

[0028] 式中，a为加权平均视图池化操作后的加权平均特征图， t_1 为卷积及池化操作后的池化特征图序号， α^{t_1} 为序号 t_1 对应的池化特征图所占权重，exp表示e为底的指数函数， z^{t_1} 为序号 t_1 对应的池化特征图；

[0029] 12.3、为前两层全连接层分别初始化一个3000*1500和1500*750的卷积核，并设置激活函数为Relu；加权平均视图池化操作后的加权平均特征图输入前两层全连接层；

[0030] 12.4、为最后一层全连接层初始化一个750*14的卷积核并设置Softmax分类函数。

[0031] 步骤(13)、各视角 workflow 视频对应的训练集中均随机选取20个视频输入到(12)中的多视图三维卷积神经网络中进行特征训练，并输出训练误差；

[0032] 步骤(14)、各视角 workflow 视频对应的训练集中随机选取10个视频输入到多视图三维卷积神经网络中进行验证，得到多视图三维卷积神经网络分类识别的准确率；

[0033] 步骤(15)、重复步骤(13)~(14)，每重复一次训练轮数减1，直到训练轮数为0，得到一个训练好的多视图三维卷积神经网络；

[0034] 步骤(16)、使用各视角 workflow 视频对应的测试集对步骤(15)中的多视图三维卷积神经网络进行测试；

[0035] 步骤(17)、对新输入的工作流视频，获取视频分辨率和帧数，初始化帧间差分阈值；设定 $t=2$ ；

[0036] 步骤(18)、根据步骤(4)~(8)提取相邻两帧兴趣点的中心坐标，并计算两个中心坐标间的距离，若距离大于设定的阈值T，则标记为运动状态 S_1 ，否则，标记为相对静止状态 S_0 ；

[0037] 步骤(19)、t取值逐步加1，重复步骤(18)直到t取值比新输入的工作流视频最后一帧取值小1，统计连续的 S_0 和 S_1 的数量，当检测到的 S_0 或 S_1 数量大于或等于N时，分割出连续的 S_0 或 S_1 对应帧中目标兴趣点存储到帧队列中，否则丢弃连续的 S_0 或 S_1 对应帧。

[0038] 步骤(20)、帧队列各个连续的 S_0 或 S_1 对应帧的集合中从第i帧开始提取连续的关键帧， $i > 5$ ，使关键帧帧数与数据集中分类好的各段视频的帧数相同。

[0039] 步骤(21)、将步骤(20)中关键帧按先后顺序组成的视频输入到步骤(15)中训练好的多视图三维卷积神经网络中对员工行为进行分类识别；

[0040] 步骤(22)、将步骤(21)中得到的行为类别与预先定义的标准工作流进行比对。

[0041] 本发明有益效果如下：

[0042] 本发明所提供的基于三维卷积神经网络的工作流识别方法主要由以下功能模块组成：运动目标分割模块、行为识别模块和动作划分模块。

[0043] 运动目标分割模块主要实现从图像、视频序列中分割出目标兴趣点。由于工作流

视频序列中的目标运动相对较大,而背景基本处于静止状态,因此可以将前后两帧图像相减,得到帧间差分图,然后可根据像素之差与阈值的大小关系来对运动目标进行分割。而所采用的自适应三帧差分法是将三个视频帧中前两帧和后两帧得到的帧间差分图进行与运算,得到三帧差分图,并且其阈值的设定是根据前面的帧间差分图自动调整的,因此其可以有效避免噪声的影响;

[0044] 行为识别模块利用3D卷积神经网络以及多视图的学习能力对移动目标进行行为识别。为了实现多视图融合,我们采用了一个视图池化层(view-pooling layer)来融合这些全局视图信息。多视图3D-CNNs中涉及多个独立的3D-CNNs用于从不同视图的图像序列中提取特征;然后,在视图池化层中对来自不同视图提取到的特征描述符进行融合并学习视图相关特征;最后用一个带有softmax分类器的全连接神经网络(full connected neural network, FNN)进行最后的识别;

[0045] 动作划分模块定义了两种状态:运动状态和相对静止状态。针对每一帧的兴趣点,取其中心坐标,当兴趣点移动时,其中心坐标也会随之移动。这时,可以取相邻两帧的兴趣点中心坐标之差来表示当前兴趣点的状态。通过这样的方式实现动态和静态的划分;

[0046] 本发明提供的工作流识别方法可有效解决复杂环境下工作流识别需要解决的两个问题,第一个问题是生产车间内各类机器、搬运车辆、辅助器械等物体彼此遮挡,以及不同工序操作的相似性、车间内频繁的光线强弱变化对工作流识别带来的影响,第二个问题是如何对视频中的图像序列进行生产过程/动作的自动划分。

附图说明

[0047] 图1为多视图三维卷积神经网络构建示意图;

[0048] 图2为工作流动作划分示意图。

具体实施方式

[0049] 下面结合附图和实施例对本发明作进一步说明。

[0050] 首先进行概念定义及符号说明:

[0051] 帧间差分阈值 $\tau_t^l = \frac{1}{2}(\max\{d_k\} + \min\{d_k\})$, t 表示当前帧号, $l \geq 1$ 代表递归次序, d_k 为帧间差分图中第 k 个像素的像素值, $\max\{d_k\}$ 为帧间差分图中像素值的最大值, $\min\{d_k\}$ 为帧间差分图中像素值的最小值;

[0052] $\tau_t^{l+1}: \tau_t^{l+1} = \frac{1}{2} \left(\frac{\sum_{d_k \geq \tau_t^l} d_k}{N_1} + \frac{\sum_{d_k < \tau_t^l} d_k}{N_2} \right)$, N_1 和 N_2 分别表示满足 $d_k \geq \tau_t^l$ 和

$d_k < \tau_t^l$ 的像素总数。

[0053] a : 加权平均视图池化操作后的加权平均特征图。

[0054] t_1 : 卷积及池化操作后的池化特征图序号。

[0055] α^{t_1} : 序号 t_1 对应的池化特征图所占权重。

[0056] z^{t_1} : 序号 t_1 对应的池化特征图。

[0057] 其次,一种基于三维卷积神经网络的工作流识别方法,实现步骤如下:

[0058] (1) 运动目标分割: 生产线上的视频监控设备往往架设于较高的位置,导致监控画面中大部分区域都是与 workflows 识别无关的工厂背景,若直接从整个监控画面中提取特征向量,这将大幅增加特征提取的难度和计算的时间消耗。因此,使用自适应阈值的三帧差分法分割出视频中的运动目标(兴趣点)部分,从而减少后面步骤的工作量。具体的:

[0059] (1.1) 从数据集中导出多视角 workflows 视频,获取各视角 workflows 视频的视频分辨率和帧数;

[0060] (1.2) 初始化各视角 workflows 视频的帧间差分阈值; 设定 $t=2$,对各视角 workflows 视频分别进行步骤(1.3)~(1.9)

[0061] (1.3) 读取一个视频帧 t 及其相邻两帧 $t-1$ 和 $t+1$,并将这三个视频帧进行灰度化和中值滤波处理;

[0062] (1.4) 分别对前两帧和后两帧进行帧间差分运算得到两张帧间差分图;

[0063] (1.5) 根据步骤(1.4)得到的两幅帧间差分图动态更新帧间差分阈值,更新方法如下:

[0064] (1.5.1) 设定 l 取值为1,第 t 帧帧间差分阈值 $\tau_t^l = \frac{1}{2}(\max\{d_k\} + \min\{d_k\})$, d_k 为帧间差分图中第 k 个像素的像素值, $\max\{d_k\}$ 为帧间差分图中像素值的最大值, $\min\{d_k\}$ 为帧间差分图中像素值的最小值;

[0065] (1.5.2) 令 $\tau_t^{l+1} = \frac{1}{2}(\frac{\sum_{d_k \geq \tau_t^l} d_k}{N_1} + \frac{\sum_{d_k < \tau_t^l} d_k}{N_2})$, N_1 和 N_2 分别表示满足 $d_k \geq \tau_t^l$ 和

$d_k < \tau_t^l$ 的像素的总数;

[0066] (1.5.3) 若 $\tau_t^{l+1} = \tau_t^l$,则将 τ_t^{l+1} 赋值给 τ_t^1 ,否则,令 $l=l+1$,重复步骤(1.5.2);

[0067] (1.6) 根据(1.5)得到的帧间差分阈值对当前帧(即中间那帧)进行二值化处理,大于帧间差分阈值的像素点设为1,小于帧间差分阈值的设为0;

[0068] (1.7) 将前后两幅差分图进行与操作,得到三帧差分图,并使用 Blob Extraction (块提取)方法获取兴趣点中心坐标;

[0069] (1.8) 将提取到的兴趣点从当前帧原始图像中分割出来;

[0070] (1.9) t 取值逐步加1,重复执行步骤(1.3)-(1.8)直到 t 取值比 workflows 视频最后一帧取值小1,重复过程中,步骤(1.8)分割尺寸不变;将每次重复过程中步骤(1.8)中得到的兴趣点图像按先后顺序保存为兴趣点视频,并根据数据集中的分类规则对兴趣点视频进行分类;

[0071] (2) 基于多视图三维卷积神经网络的行为识别: 对当前制造业生产线进行考察后可发现,目前制造业生产线中往往会对同一个工作场景采用多个摄像头从不同角度进行同步实时监测,以此来保证产品的质量以及员工的安全。利用这一特点,我们使用多视图特征提取与融合的方法来有效降低了工厂复杂环境对行为识别的影响,提高行为识别的准确

率。具体执行步骤如下：

[0072] (2.1) 从(1)得到的兴趣点视频中选取90%作为训练集,其余作为测试集;

[0073] (2.2) 构建一个多视图三维卷积神经网络(见附图1)。初始化训练轮数为5000,多视图三维卷积神经网络构建方法如下:

[0074] 卷积及池化操作过程为(2.2.1)~(2.2.4):

[0075] (2.2.1) 为第一卷积层初始化一个大小为 $9*9*9*10$ 的四维卷积核,激活函数为sigmoid,第一池化层窗口大小为2,步长为2;

[0076] (2.2.2) 为第二卷积层初始化一个大小为 $9*9*7*30$ 的四维卷积核,激活函数为sigmoid,第二池化层窗口大小为2,步长为2;

[0077] (2.2.3) 为第三卷积层初始化一个大小为 $9*8*5*50$ 的四维卷积核,激活函数为sigmoid,第三池化层窗口大小为2,步长为2;

[0078] (2.2.4) 为第四卷积层初始化一个大小为 $4*3*3*150$ 的四维卷积核,激活函数为sigmoid,第四池化层窗口大小为2,步长为2;

[0079] (2.2.5) 初始化加权平均视图池化层中各特征图权重参数 α^{t_1} , α^{t_1} 为 $[0, 1]$ 中随机值,且 $\sum_{t_1} \alpha^{t_1} = 1$; 加权平均视图池化层(weighted average view-pooling layer, WAVP)

计算公式如下:

$$[0080] \quad a = \frac{\sum_{t_1} \exp \alpha^{t_1} \cdot z^{t_1}}{\sum_{t_1} \exp \alpha^{t_1}}$$

[0081] (2.2.6) 为前两层全连接层分别初始化一个 $3000*1500$ 和 $1500*750$ 的卷积核,并设置激活函数为Relu;加权平均视图池化操作后的加权平均特征图输入前两层全连接层;

[0082] (2.2.7) 为最后一层全连接层初始化一个 $750*14$ 的卷积核并设置Softmax分类函数,其中14为动作的种类。

[0083] (2.3) 各视角 workflow 视频的训练集中随机选取20个视频输入到(2.2)中的多视图三维卷积神经网络中进行特征训练,并输出训练误差;

[0084] (2.4) 各视角 workflow 视频的训练集中随机选取10个的视频输入到多视图三维卷积神经网络中进行验证,得到多视图三维卷积神经网络分类识别的准确率;

[0085] (2.5) 重复(2.3)~(2.4),每重复一次训练轮数减1,直到训练轮数为0,得到一个训练好的多视图三维卷积神经网络;

[0086] (2.6) 使用各视角 workflow 视频对应的测试集对(2.5)中的多视图三维卷积神经网络进行测试;

[0087] (3) 基于状态的动作划分方法:实际环境下,工人的动作往往是连续发生的,这种情况下,若要对动作进行识别需要先对动作进行分割,然后才能对每个动作分别进行识别。经观察发现,工人从取零件、搬运零件到放置零件以及工人从取焊接工具到进行零件焊接,这些行为中间都会发生一段位移(见附图2)。所以,可以根据工人的运动状态对动作进行划分。具体执行步骤如下:

[0088] (3.1) 对新输入的视频,获取视频分辨率和帧数,初始化帧间差分阈值;设定 $t=2$;

[0089] (3.2) 根据 (1.3) ~ (1.7) 提取相邻两帧兴趣点的中心坐标, 并计算两个中心坐标间的距离, 若距离大于我们设定的阈值 T , 则标记为运动状态 S_1 , 否则, 标记为相对静止状态 S_0 ;

[0090] (3.3) t 取值逐步加 1, 重复 (3.2) 直到 t 取值比新输入的视频最后一帧取值小 1, 统计连续的 S_0 和 S_1 数量, 当检测到连续的 S_0 或 S_1 数量大于或等于 N 时, $N > 10$, 用 (1.8) 的方法分割出连续的 S_0 或 S_1 对应帧中目标兴趣点存储到帧队列中, 否则丢弃连续的 S_0 或 S_1 对应帧。

[0091] (3.4) 帧队列各个连续的 S_0 或 S_1 对应帧的集合中从第 i 帧开始提取连续的关键帧, $i > 5$, 使其与数据集中分类好的各段视频的帧数相同。

[0092] (3.5) 将 (3.4) 中关键帧按先后顺序组成的视频输入到 (2.5) 中训练好的多视图三维卷积神经网络中对员工行为进行分类识别;

[0093] (3.6) 将 (3.5) 中得到的行为类别与预先定义的标准工作流程进行比对。

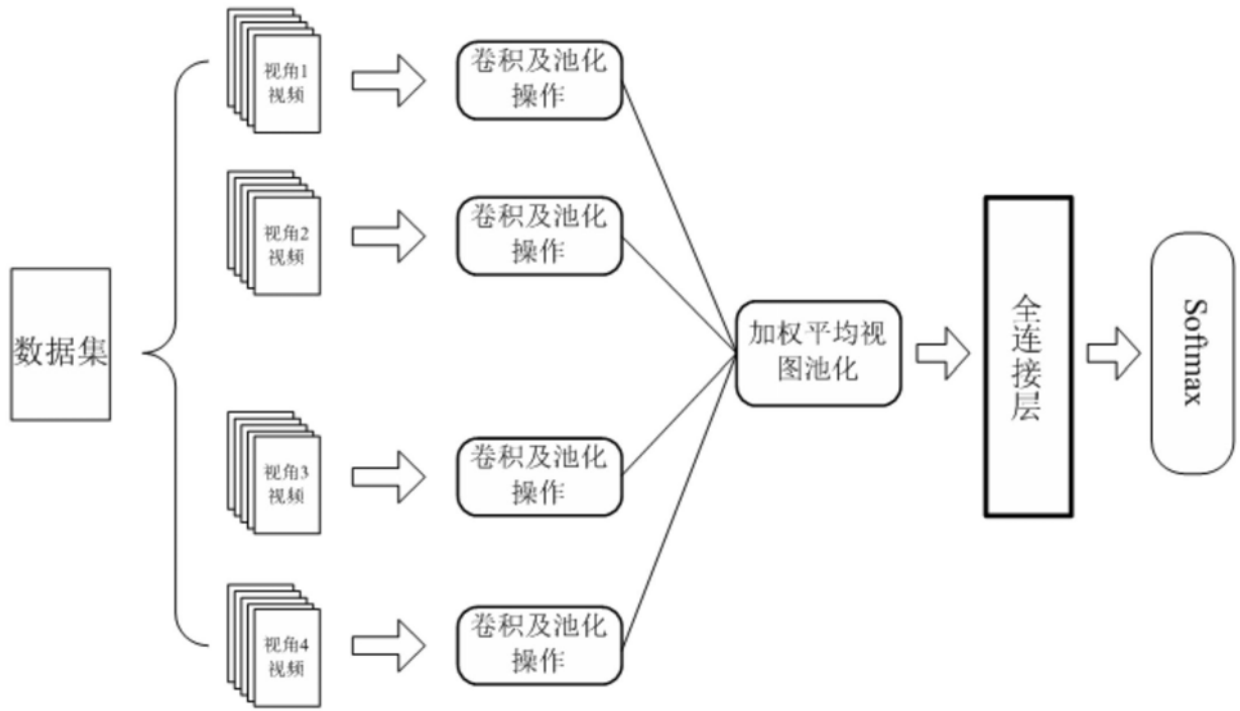


图1

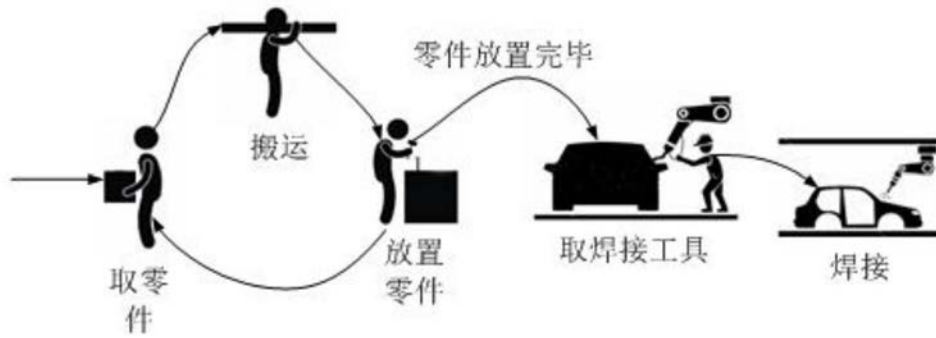


图2