

基于人体骨骼信息的列车司机行为智能识别

王志毅¹, 刘昕武²

(1. 朔黄铁路发展有限责任公司, 河北 肃宁 062350;
2. 株洲中车时代电气股份有限公司, 湖南 株洲 412001)

摘要: 文章提出了一种基于骨架识别的人体姿态捕捉预测算法, 用于识别列车驾驶员违章行为。该方法是将司机作业监控视频拆分成一组连续的图片组合, 然后通过构建深度学习模型网络预测每一帧图片中的人体姿态, 并使用逻辑回归模型对司机行为进行分类, 最终, 通过将识别出的司机行为与LKJ数据相关联, 裁定司机行为是否违章。该方法比目前机务段靠人工抽检的方式更加高效合理, 但其识别效果受限于车载摄像头清晰度。

关键词: 人体骨骼; 深度学习; 姿态识别; LKJ; 列车

中图分类号: U268.4

文献标识码: A

doi: 10.13890/j.issn.1000-128x.2020.04.018

Intelligent Recognition of Train Driver Behavior Based on Human Skeleton Information

WANG Zhiyi¹, LIU Xinwu²

(1. Shuohuang Railway Development Co., Ltd., Suning, Hebei 062350, China;

2. Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou, Hunan 412001, China)

Abstract: A human posture capture and prediction recognition was proposed, which was used to identify violated driver. Firstly, driver behavior or action captured by camera was split into a set of successive frames with human posture. Then each posture in a frame was recognized by deep learning network and a type of behaviors composed by a set of posture was classified via logistic regression model. Finally, the association between detected driver behaviors and locomotive LKJ data was applied in the recognition of violated driver. Compared with the daily random-check approach by inspectors in locomotive depots, this way was more reasonable and efficient. However, the accuracy of it was limited by the resolution of surveillance cameras in the driving cab of locomotive.

Keywords: human skeleton; deep learning; posture recognition; LKJ; train

0 引言

主、副司机的驾驶作业行为是否规范对机车的安全运行及运维起到十分关键的作用^[1], 为此, 有关部门制定了一系列的司机行为作业守则和配套的视频监控设备来确保司机的规范操作^[2-3]。目前, 各机务段主要采用人工审阅从机车下载的监控视频来判别司机在机车上的作业是否合规。然而由于各机务段每天出车数量过大、每辆车视频记录时间过长, 传统靠人工审阅视频的方法无法及时有效地分析完成当天所有视频

的审阅。因此, 迫切需要一种有效、自动、快速的针对机车司机行为识别的方法。

随着深度学习技术在图像处理领域取得的进步与发展^[4-5], 目前轨道交通领域存在一些关于驾驶室司机行为识别的理论方法^[6], 但由于驾驶室情况复杂, 不同类型机车驾驶室内部设施、摄像头安装角度等也均不相同, 对于司机行为违规判定的自动化产品设计规范一直没有统一制定。

本文提出了一种基于人体骨架信息的列车司机行为识别方法, 其基于列车驾驶室视频数据和列车LKJ数据, 采用人体骨架识别方法获取主副司机在视频中

收稿日期: 2019-07-15; 修回日期: 2020-07-02

人体骨架的坐标信息, 构建人体行为识别模型, 最后根据制定的规定判断司机作业是否违规, 从而有效规范司机作业, 确保列车安全、正常运行。

1 算法概述

1.1 人体骨架识别

人体骨架识别是指在图像中获取相关人物各重要关节点(头、肩、手、脚蹬)的二维坐标信息, 主要有2个途径获得: 一是采用深度学习方法分析传统RGB图像, 构建模型获得; 二是通过深度摄像机(如Kinect)直接获得。

在深度学习识别模型构建领域, 文献[7]提出了一种基于部分亲和域(Part Affinity Fields, PAF)的实时二维人体骨骼识别方法。该方法通过使用正反馈网络预测人体关节位置信息置信图(也可称为热点图)以及由关节之间亲和度构成的二维向量域PAF。算法首先计算人体所有关节位置, 然后通过关节构建人体躯干。由于人体每段躯干由2个关节组成, 这2个关节之间的亲和度可以用1个二维向量PAF表示。PAF的关键作用也就是采用亲和度计算实现人体一段躯干两端关节的快速匹配。式(1)用于描述一张图内位置为 p 的第 k 个人的关节 j 置信度分布情况; 式(2)用于计算一张图内第 k 个人的第 c 条躯干的PAF。

$$S_j^*(p) = \max_k S_{j,k}^*(p) \quad (1)$$

$$L_c^*(p) = \frac{1}{n_c(p)} \sum_k L_{c,k}^*(p) \quad (2)$$

式中: p 为图内关节点位置二维坐标信息; k 为图中人体编号; j 为关节点编号。

因为人体每个关节部位由图片当中特定区域的像素值组成, 越靠近关节点中心位置的像素对应该关节的置信度越高(在热点图中表示为越热); 反之, 离该关节点越远的位置的像素对应该关节的置信度越低。因此图片中每个像素位置关于该关节的置信度分布 $S_{j,k}^*(p)$ 符合高斯分布:

$$S_{j,k}^*(p) = \exp\left(-\frac{\|p - x_{j,k}\|_2^2}{\sigma^2}\right) \quad (3)$$

式中: σ 为控制峰值的传播范围; $x_{j,k}$ 为第 k 个人的第 j 个关节的实际坐标位置; $S_{j,k}^*(p)$ 为图中位置 p 的第 k 个人关节的置信度。

由于每个人体有多个关节点, 图中任意一个像素位置会对应多个关节, 整张图的关节位置置信度 $S_j^*(p)$ 呈现为多峰的高斯分布形式。因此, 式(1)选用 \max 函数筛选对每个像素点对应置信度最高的关节, 从而过滤掉同一个位置对应2个关节的情况。

式(2)当中的 $L_{c,k}^*(p)$ 定义了图中任意像素位置 p 对应第 k 个人的第 c 条躯干的矢量, 其计算公式为

$$L_{c,k}^*(p) = \frac{x_{j_2,k} - x_{j_1,k}}{\|x_{j_2,k} - x_{j_1,k}\|_2} \quad (4)$$

式中: c 为人体躯干编号, 躯干 c 由关节 j_1 和关节 j_2 组成; $x_{j_1,k}$ 和 $x_{j_2,k}$ 分别为第 k 个人的第 j_1 个和第 j_2 个关节的实际坐标位置。

当位置 p 确实属于躯干 c 的一部分时, $L_{c,k}^*(p)$ 描述了一条从关节 j_1 指向 j_2 的单位矢量; 反之, $L_{c,k}^*(p)$ 等于0。由于人体之间会存在遮挡的情况, 图片当中的某些位置可能处在不同人的不同躯干上, 从而产生多个单位矢量, 所以引入 $n_c(p)$ 描述位置 p 在 k 个人中所有非零矢量总数。最后, 将像素位置 p 的所有躯干单位矢量进行求和, 并除以位置 p 的非零向量数目 $n_c(p)$ 所得到的平均值即为位置 p 的PAF。

1.2 逻辑回归二分类^[8]

根据在线性回归理论, 可使用梯度下降法得到一个映射函数 $h_\theta(X)$ 去贴近样本点, 并且该函数是对连续值的一个预测。逻辑回归是解决分类问题的一种算法, 可以通过这个算法得到一个映射函数 $f: \mathbf{X} \rightarrow y$ 。其中, \mathbf{X} 为特征向量, 由众多特征 x 组成, $\mathbf{X} = \{x_0, x_1, x_2, \dots, x_n\}$, y 为预测的结果。在逻辑回归问题当中, 标签 y 为一个离散值, 通常用来表示目标的类别(即标签)。

分类问题的核心思想是寻找一个判定边界, 对不同标签的数据样本进行划分。由于样本数据的特征值通常可以映射到一个超平面, 因此可以通过学习训练得到一个方程 $E_\theta(\mathbf{X})=0$ 来表示判定边界:

$$E_\theta(\mathbf{X}) = \mathbf{X}^T \boldsymbol{\theta} \quad (5)$$

式中: $\boldsymbol{\theta} = \{\theta_0, \theta_1, \theta_2, \dots, \theta_n\}$ 为特征向量 \mathbf{X} 的系数矩阵; \mathbf{X} 向量当中的每一个特征 x_i 对应 $\boldsymbol{\theta}$ 向量中的系数 θ_i , 而且每一个 θ_i 均是未知的, 需要从样本数据中学习得来。

为保留 $E_\theta(\mathbf{X})=0$ 中的常数项, 令特征向量 $\mathbf{X} = \{x_0, x_1, x_2, \dots, x_n\}$, 对于位于判定边界上的样本, 其特征向量 \mathbf{X} 可使得 $E_\theta(\mathbf{X})=0$ 。因此, 判定边界是满足 $E_\theta(\mathbf{X})=0$ 的特征向量 \mathbf{X} 所代表的点集合, 也可以看作是特征向量的等高超平面。

对于二分类问题, 即样本标签的类型只有2种类型。其中, 一类的样本点在判定边界的一边, $E_\theta(\mathbf{X}) > 0$; 而另一类的样本会在判定边界的另一边, $E_\theta(\mathbf{X}) < 0$ 。当样本点离判定边界越远时, $E_\theta(\mathbf{X})$ 的绝对值越大, 这时样本的标签是某种类型的概率会很大; 当样本点离判定边界越近时, $E_\theta(\mathbf{X})$ 的绝对值接近0, 样本的标签是某种类型的概率会在0.5左右。因此可以将 $E_\theta(\mathbf{X})$ 函数转换成一种概率函数, 通过概率判断样本的标签是某一种类型的概率, 这种转换可以使用sigmoid函数 $g(z)$ 来实现^[9]:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (6)$$

式中: z 为函数自变量; e 为自然常数。

$g(z)$ 的取值范围为(0, 1), 它的作用是将一个实数

z 映射到 $(0, 1)$ 的区间。

当 z 为 0 时, 函数 $g(z)$ 的值为 0.5; z 大于 0 且 z 值越大时, 函数值大于 0.5 且收敛于 1; z 小于 0 且 z 值越小时, 函数值小于 0.5 且收敛于 0。因此, sigmoid 函数可适用于在二分类问题中将 $E_\theta(\mathbf{X})$ 函数转换为概率函数。当 $E_\theta(\mathbf{X}) > 0$ 时, 样本标记的类型为某一类型的概率会大于 0.5; 当 $E_\theta(\mathbf{X}) < 0$ 时, 样本标记的类型为某一类型的概率会小于 0.5; 当 $E_\theta(\mathbf{X})$ 约等于 0 时, 样本标记的类型为某一类型的概率会在 0.5 左右。

在二分类问题中, 可以找到逻辑回归函数 $h_\theta(\mathbf{X}) = g[E_\theta(\mathbf{X})]$, 判定边界可看作 $h_\theta(\mathbf{X}) = 0.5$ 时的等高超平面。将式 (5) 代入式 (6) 可得

$$h_\theta(\mathbf{X}) = g(E_\theta(\mathbf{X})) = g(\mathbf{X}^T \theta) = \frac{1}{1 + e^{-\mathbf{X}^T \theta}} \quad (7)$$

2 基于人体骨架的行为识别

基于人体骨架识别的列车司机行为识别方法流程如图 1 所示。

①视频分类。由于受驾驶室布局及内设影响, 不同车型的驾驶室摄像头安装位置有差异, 因此需要按摄像头的位置对视频源进行初步划分。

②行为姿态分解。将需要分析的行为进行姿态分解, 例如将手指前方这个动作分解成抬手、伸直手臂和指向前方 3 个姿态。

③关键帧提取。从机车驾驶室视频中将司机姿态符合步骤②描述的帧提取出来作为正样本数据。此外, 为满足后续二分类模型训练需要, 还需要从视频中提取包含司机正常姿态的帧作为负样本数据, 例如司机正常坐立无任何违章行为的姿态。

④识别人体骨架。通过人体骨架识别算法获得关键帧的司机人体骨架二维坐标信息。

⑤提取关键坐标。从识别出的人体骨架二维坐标中提取每种姿态对应关键部位的二维坐标。例如, 分析手指前方这个动作, 需要将人体的手臂、肩膀及上身坐标提取出来, 因为这几个关节坐标决定司机“抬手”“伸直手臂”“指向前方”这 3 种姿态。其余与该行为无关的坐标 (如腰部、臀部、腿部等) 将被滤除, 不进入后续模型。

⑥数据标注。将每一种姿态对应的关键坐标数据进行标注。每一种姿态只能对应一种标签 (如“抬手”), 并且司机正常姿态的负样本坐标数据需要和违章姿态的正数据区分开, 也需要对应一种标签 (如“正常”)。

⑦构建二分类模型训练。对于每一种姿态, 除了关键坐标信息外, 还可从坐标信息中提取特征扩充数据维度。例如关键部位坐标之间的相对位置、欧式距离、部位之间的夹角 (例如胳膊与身体躯干、身体躯干与腿部之间的夹角) 等都可作为该姿态的数据特征。将每一种姿态的正负样本坐标数据及其特征值合并后即

可进行逻辑回归二分类模型训练。如前文所述, 模型的目的是寻找一种映射 $f: \mathbf{X} \rightarrow y$, 其中, \mathbf{X} 由该姿态的关键坐标和其特征值组成, y 为该姿态对应的标签。模型根据式 (7) 寻找正负样本数据之间的差异, 通过不断迭代训练的方式, 调整与 \mathbf{X} 匹配的系数矩阵 θ 直到最优。在最优状态下, 系数矩阵 θ 将被固定, 对于任意给定的特征向量 \mathbf{X} , 模型通过式 (7) 计算 $h_\theta(\mathbf{X})$ 的值是否大于 0.5 来判断 \mathbf{X} 是正常姿态还是违章姿态。

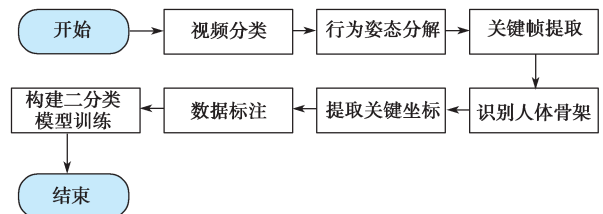


图 1 行为识别流程图

3 试验结果

为验证本文所提方法的有效性, 基于某机务段的视频数据, 进行了“司机手指前方”和“司机偏头”2 种动作的识别试验。每个动作有效图片为 10 000 张, 其中训练图片为 9 000 张, 测试图片为 1 000 张。由于采用的是二分类识别模型, 因此图片又分为做了动作和没做动作 2 种。

测试行为: 手指前方, 偏头

测试环境: Ubuntu 16.04, CUDA 8.0, cudnn 5.1

测试软件: OpenPose v1.3.0

软件参数:

net_resolution = -1x368

model_pose = BODY_25

alpha-pose = 0.6

scale_gap = 0.3

render_threshold = 0.05

测试数据集示例如图 2 所示。



图 2 数据集示例

试验结果如表 1 和表 2 所示。其中, 司机实际动作代表 1 张图片当中显示司机真实的动作; 算法识别结果为计算机算法模型识别出的动作。若 1 张图片当中司机做了某个动作 (如偏头), 并且算法也识别出偏头行为, 则可认为算法识别正确; 若图片当中司机

没做某个动作（如没有偏头），并且算法也没有识别出偏头行为，则也认为算法识别正确。反之，若图片当中司机做了某个动作，但算法没有识别出该动作，或者司机没有做某个动作，但算法识别出了某个动作，则认为算法识别错误。表中记录的数字为针对 2 个动作在各自 1 000 张的测试图片上的识别命中情况。从表 1 和表 2 可知，在测试数据中，司机实际手指前方动作共有 666 次，其中算法正确识别出手指前方动作为 657 次，9 次判断错误；司机实际未手指前方为 334 次，其中算法正确识别出没有手指前方 304 次，30 次错误识别成司机有手指前方动作。司机手指前方动作识别准确率为

$$\frac{\text{识别正确次数}}{\text{样本总数}} \times 100\% = \frac{657 + 304}{1000} \times 100\% = 96.1\%$$

在测试数据中，司机实际偏头动作共有 595 次，其中算法正确识别出偏头动作为 582 次，13 次判断错误；司机实际未偏头动作有 405 次，其中算法正确识别出没有偏头 344 次，61 次错误识别成有偏头动作。司机偏头动作识别准确率为

$$\frac{\text{识别正确次数}}{\text{样本总数}} \times 100\% = \frac{582 + 344}{1000} \times 100\% = 92.6\%$$

手指前方识别的准确率略高于偏头动作的识别是因为手势行为的识别用到的人体关节坐标要多于偏头动作行为，这说明越多的人体坐标信息对动作识别的准确率提升有促进作用。

表 1 司机手指前方识别命中率试验结果

司机实际动作	正确	错误
有手指前方	657	9
未手指前方	304	30

表 2 司机偏头识别命中率试验结果

司机实际动作	正确	错误
有偏头	582	13
未偏头	344	61

4 结论

采用人体骨架识别方法的司机违规行为判别，具有以下特点：

①有效且实际地解决了机务段靠人工抽检海量驾驶室视频数据导致的识别效率低和识别准确率不高的

问题。

②采用按不同车型和不同摄像头安装位置构建分析模型，有效提高模型预测准确率。

③采用人体骨架识别方法，对司机姿态判定更具有准确性。

④通过姿态组合构建司机行为的方式，使对司机行为的判定更加严格准确。

由于目前车载摄像头精度与视角限制，导致识别准确率有待提高；同时，由于计算资源的要求，导致整套方法成本偏高。后续将针对这 2 个问题继续进行研究。

参考文献:

[1] KECKLUND L, INGRE M, KECKLUND G, et al. The train-project: Railway safety and the train driver information environment and work situation-A summary of the main results[J]. Computers in Railways VII, 2000, 50. [2020-06-30]. <https://www.witpress.com/elibrary/wit-transactions-on-the-built-environment/50/4182>. DOI: 10.2495/CR001021.

[2] MCLEOD R W, WALKER G H, MORAY N. Analysing and modelling train driver performance[J]. Applied Ergonomics, 2005, 36(6): 671-680.

[3] MCLEOD R W, WALKER G H, MORAY N. Assessing the human factors risks in extending the use of AWS[J]. Applied Ergonomics, 2005, 36(5): 671-680.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//ACM. Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2012, 1: 1097-1105.

[5] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展 [J]. 计算机科学, 2016, 43(2): 1-8.

[6] 彭军, 何伟锋, 杨迎泽. 嵌入式列车司机驾驶疲劳检测系统设计 [J]. 计算机工程与应用, 2009, 45(33): 57-59.

[7] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1302-1310. DOI: 10.1109/CVPR.2017.143.

[8] 何晓群, 刘文卿. 应用回归分析 [M]. 4 版. 北京: 中国人民大学出版社, 2015.

[9] WikiMili. Sigmoid function[EB/OL]. 2020-06-21[2020-07-02]. https://wikimili.com/en/Sigmoid_function.

作者简介: 王志毅 (1970—), 男, 工程师, 主要从事机务运用管理工作。