国家计量技术规范

《行人重识别（ReID）算法测评方法》

编制说明

《行人重识别（ReID）算法测评方法》编制组

2023年10月

**《行人重识别（ReID）算法测评方法》编制说明**

# 一、任务来源及计划要求

《行人重识别（ReID）算法测评方法》国家计量技术规范的编写任务，是由国家市场监督管理总局关于印发计量函[2021]50号《市场监管总局办公厅关于下达<2021年国家计量技术规范制定、修订及宣贯计划>的通知》批准立项。根据国家市场监管总局的批复，由之江实验室、中国计量科学研究院、浙江省计量科学研究院、广东省计量科学研究院、福建省计量科学研究院、江苏省计量科学研究院、浙江大华技术股份有限公司、浙江省智能技术标准创新促进会负责《行人重识别（ReID）算法测评方法》国家计量技术规范制定的起草工作。

# 二、项目意义

计算机视觉技术作为人工智能技术的重要领域之一，广泛应用在安防、医疗、自动驾驶等领域。而人工智能算法作为技术核心，如何可信的评价其性能成为计量科学关注的新课题。美国NIST于2020年开展了针对人脸识别算法的质量评估研究，通过对成像属性（聚焦、照明、失真等）和对象属性（头部姿势、面部表情等）的量值化，建立了标准数据库，并立项了三项ISO/IEC的相关标准，为人工智能算法的计量评估奠定了基础。

行人重识别算法（ReID）是利用计算机视觉技术，跨摄像头、跨场景下在图像集或视频序列中检索特定行人的技术。该技术解决了在实际监控场景下，由于拍摄角度、相机清晰度、行人运动、照明、遮挡等原因，导致目标面部信息不全，人脸识别失效的缺点，根据行人的穿着、体态、发型等信息认知行人，对目标进行跨摄像头连续跟踪。这对于在复杂非配合环境下，尤其是疫情期间佩戴口罩的场景中，获取目标跟踪具有重要意义。

行人重识别算法的准确性和一致性直接决定了行人重识别技术的应用效果，尤其应用于安防领域时，算法的优劣将对社会安定、人身安全产生直接的影响。目前市场上应用行人重识别算法的产品有很多，由于缺乏科学合理的评价手段，导致算法的准确性和一致性比较差。因此有必要从计量角度对行人重识别算法评估建立规范，从而规范和统一行人重识别算法的评价方法。该规范的研究和制定，能够形成明确清晰的行人重识别算法测评体系，客观反应算法在不同复杂场景中的性能，从计量层面保证行人重识别算法性能的准确性、可靠性与一致性，以助推行人重识别及计算机视觉技术的发展。

# 三、编制依据

本规范依据JJF 1001-2011《通用计量术语及定义》、JJF 1071-2010《国家计量校准规范编写规则》以及JJF 1059-2012《测量不确定度评定与表示》为基础性系列规范进行制定。

本规范主要参考了GB/T 29268.2-2012《信息技术 生物特征识别性能测试和报告 第2部分：技术与场景评价的测试方法》、GB/T 33767.5-2018《信息技术 生物特征样本质量 第5部分：人脸图像数据》、GB/T 41864-2022《信息技术 计算机视觉 术语》的技术要求编制而成。

本规范为首次发布。

# 四、编制过程

## 4.1 预研阶段

2021年1月：由之江实验室、中国计量科学研究院、浙江大华技术股份有限公司等主要单位组成预研组，对目前行人重识别算法的应用现状、存在的问题、算法性能的测试开展了调研。通过算法评测的相关标准、行人重识别算法评测的数据库调研，起草了国家计量规范项目申报书。

## 4.2 立项阶段

2021年8月23日，全国法制计量管理计量技术委员会MTC1[2021]5号通知下达了《行人重识别（ReID）算法评测方法）》的国家计量技术规范制修订项目计划，项目周期为1年。

## 4.3 起草阶段

2021年9月，编制组于之江实验室召开启动会，针对技术规范的技术路线进行了讨论，确定行人重识别算法的测评参数、测试数据集构建框架。

2021年10月-2022年2月，针对本规范的测试数据集的构建开展广泛文献调研，同时与行人重识别领域的研究人员、测试集构建者以及华为、电信等算法使用商、研发商开展广泛研讨，征求产学研用各界对数据集构建建议。

2022年3月-2022年7月，开展数据集采集与数据标注工作。共采集标注2000ID、20个摄像头约40w张图像数据，同时对采集后数据集进行整理分类，构建为规范的行人重识别数据集。

2022年9月-2022年10月，研制算法测试接口、流程，针对不同的算法模型开展测试，同时撰写规范草案、编制说明与试验报告。

2022年11月-2023年2月，针对数据集开展第二轮数字量具特性标注，并对标注后的数据进行实验测试，进一步修改规范草案、编制说明与试验报告 。

2023年3月，全国数字计量技术委员会首次会议召开，并对该规范内容进行评审讨论，委员专家对规范中存在的问题提出相应的修改建议。

2023年4月，就委员会所收到的意见及修改情况，编制组于之江实验室联合大华、海康等单位及浙江省计量院召开研讨会议，对修改后的稿件进行进一步讨论完善。

2023年5月-2023年9月，为保障不确定度的评定的顺利进行，对数据集进行进一步重构优化，确定每个数据集的ID数量及query与gallery分布情况，并进行实验测试与不确定度分析，形成征求意见稿。

2023年10月编制组对该规范于之江实验室召开评审会议，并参加于西安举办的全国数字计量委员会会议对规范内容进行评审，编制组依据两次评审会议意见对规范进行修改完善。

# 五、主要技术内容的论据

本规范通过文献调研、与产学研各界的技术交流，掌握了行人重识别算法现有通用的评测方法和评测参数，确定了将平均精度、处理效率作为算法的主要计量特性。

在构建算法评测用数字量具的过程中，广泛调研了影响算法性能的关键因素，以及现有测试数据库的特点与不足。目前常用的测试数据集如表1所示。目前共同存在的问题是图像场景简单、行人数据量有限、摄像头风格变动。同一算法在不同数据集中的测试结果不同，同时算法在测试集中测试结果良好，而在实际应用中效果极差，因此无法真实反应算法的性能。因此在构建数字量具的过程中，重点考虑了图像的多样性、真实性、可量化的计量特性等重要因素。

在整个规范的研究、起草过程中，本着科学、合理、规范的原则，以适用性和可操作性作为前提，明确行人重识别算法的计量参数、测评要求，解决其可信度测评方法，制订本计量规范。规范的主要研究内容包括数字量具的构建、评测条件、评测项目、评测方法等内容。下面对数字量具的构建、评测项目和评测方法的论据进行分析。

表1 现有ReID算法测试数据集情况

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **数据集** | **年份** | **ID数量** | **图片数量** | **相机数量** |
| VIPeR | 2007 | 632 | 1264 | 2 |
| iLIDS | 2009 | 119 | 476 | 2 |
| GRID | 2009 | 250 | 1275 | 8 |
| PRID2011 | 2011 | 200 | 1134 | 2 |
| CUHK01 | 2012 | 971 | 3884 | 2 |
| CUHK02 | 2013 | 1816 | 7264 | 10 |
| CUHK03 | 2014 | 1467 | 13164 | 2 |
| Market1501 | 2015 | 1501 | 32668 | 6 |
| DukeMTMC | 2017 | 1404 | 36441 | 8 |
| Airport | 2017 | 9651 | 39902 | 6 |
| MSMT17 | 2018 | 4101 | 126441 | 15 |

## 5.1 数字量具构建

5.1.1 数字量具的分组

通过调研及实验分析得到，影响算法性能的主要因素包括光照、行人遮挡、拍摄角度、图像分辨率。为评测算法在各种影响因素下的性能，进而为算法提升提出针对性的建议，需将以上四个主要影响因素进行解耦，分析算法在不同影响因素下的性能。因此本规范将数字量具分为五组：全特性数字量具、拍摄角度数字量具、光照数字量具、遮挡数字量具以及分辨率数字量具。

5.1.2 数字量具的通用要求

数字量具的通用要求包含了行人和图像的要求两部分。

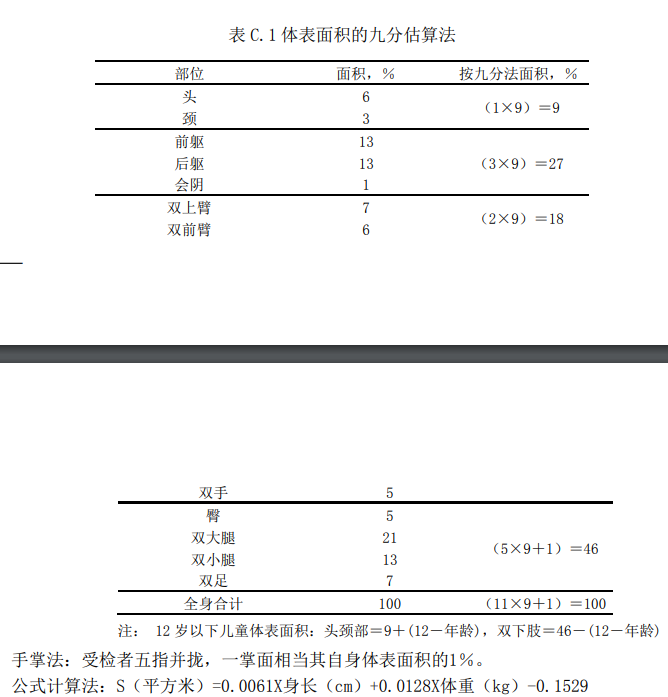
对于所给出的行人性别和年龄分布要求，其主要参考我国“第七次全国人口普查公报第四号与第五号文件”所公布的数据制定，以构建丰富多样、性别均衡、年龄段全覆盖的符合实际应用场景的数字量具。

对于图像的要求，主要规定了图像格式、标注信息、图像质量、图像数量和数据分布。其中图像格式要求为“jpg、.bmp、.png、.jpeg”等，符合数码摄像机输出的最典型图像格式。图像标注信息包含行人编号、图像采集相机编号、拍摄角度、平均亮度、遮挡比例、分辨率，其中行人编号、图像采集相机编号为了区分不同行人与所拍摄的相机信息，属于基础标注信息；拍摄角度、平均亮度、遮挡比例、分辨率的标注为数字量具的计量特性，必须标注。图像质量规定了采集设备数量、图像分辨率、采集设备的高度和角度，设备数量不少于5台的要求依据主要行人重识别现有数据集及实际应用的需求；图像分辨率的要求依据主要是算法的实际测试与与算法研发商的经验，普遍认为对于分辨率低于的图像，算法将失效，为此将最低分辨率定位成像设备设置中的采集设备的高度和角度主要以户外监控摄像设备设置为参考设置。图像数量提出的综合考虑了数据多样性的最低要求和实际构建的难度。

5.1.3 数字量具的计量特性

拍摄角度、光照、遮挡、分辨率是公认的影响行人重识别算法精度的主要因素，也是数字量具分类的重要依据，因此将拍摄角度、平均亮度、遮挡率和分辨率作为数字量具的计量特性。其中拍摄角度通过与拍摄角度参考图像（每30度为间隔拍摄）对比赋值，分辨率通过程序读取图像二维像素值表示。

遮挡率参照《人体损伤程度鉴定标准》定义的体表面积九分估算法进行赋值（九分估算法：成人体表面积视为100％，将总体表面积划分为11个9％等面积区域，即头（面）颈部 占一个9％，双上肢占二个9％，躯干前后及会阴部占三个9％，臀部及双下肢占五个9％＋1％）。



平均亮度则采用整张图片YUV色度空间的亮度值Y表示，平均亮度参考ITU-R BT.601、ITU-R BT.709中的转化公式计算得到：

式中：

——图像平均亮度；

——图像红色（R）通道的平均灰度值；

——图像绿色（G）通道的平均灰度值；

——图像蓝色（B）通道的平均灰度值。

下面分别对每个数字量具的计量特性做分析。

（1）拍摄角度数字量具。拍摄角度以30°为区间划分，主要原因为对比赋值对于30°的判断精度相对较高，同时能反映对算法的影响。本规范将拍摄角度划分为正面、侧面、背面三大类，其中正面与背面各覆盖120°角度范围，侧面各覆盖120°。以所覆盖的角度范围为基准，将正面、侧面、背面所包含的图像数据量占比设置为30%、40%、30%，与覆盖角度范围占比保持基本一致。将平均亮度[90-130]是对照明条件较好的图片进行统计分析的结果，目的是使平均亮度基本一致且光照良好，以便减少光照对算法精度的影响。遮挡率及图像分辨率的计量要求基于实验结果，当遮挡率不高于20%时，对算法的影响可以忽略，图像分辨率大于数时，算法性能稳定。

（2）光照数字量具。对于平均亮度，本规范划分了[30,70]、、共3个亮度区间，总体覆盖低亮度、正常亮度和高亮度的场景，其所包含图像数占比依据数据亮度实际分布情况及各区间覆盖范围分别设置为20%、60%、20%，符合对算法在不同环境照明情况下的性能评测需求。其中最小值30的设置原因为，经实验测试，平均亮度低于30时，算法性能下降，且低于该值的亮度区间不属于行人重识别常用场景。拍摄角度计量要求为行人角度处于或区间范围内，该区间图像内行人保持正面姿态，对算法性能影响较低。遮挡比例和分辨率的计量要求依据同上。

（3）分辨率数字量具。分辨率如5.1.2所述，最小值设置为对上则不设限。此外，对于分辨率设置了[]数、数、像素数以及大于像素数四大区间。其所包含图像数占比依据数据亮度实际分布情况及各区间覆盖范围分别设置为20%、30%、30%、20%，符合分辨率多样性的需求。其他计量特性的要求依据同上。

（4）遮挡数字量具。由实验测试可知，遮挡率高于80%时，大部分算法失效，因此遮挡数字量具中，遮挡率最小值不设限，最大值设置为80%。此外，对于遮挡设备了、(20%,40%]、(40%,60%]、(60%,80%]四大区间。其所包含图像数占比与各区间覆盖范围比例保持一致，设置为25%、25%、25%、25%。其他计量特性的要求依据同上。

数字量具除以上计量特性外，同时也考虑了场景、季节的多样性，采集包含了工业园区、马路、校园的室内和室外场景，同时在季节上包含了春夏，因此在服装特点上也做到了多样性。

## 5.2 评测项目的选择

现有ReID算法的评测通常包含以下指标：

*Rank-K*：算法返回的排序列表中，前*K*位存在检索目标则称为*Rank-K*命中；

：平均精度，反应检索的人在数据库中所有正确的图片排在排序列表前面的程度。

*Rank-K*无法全面反映算法的精度，而则反映了算法的综合精度水平，因此将作为评测项目之一。

此外算法的处理效率也是反映算法性能的关键参数，因此提出时处理效率*pe*，通常算法的时间效率用处理单张图像的时间表示，但单张图像的大小并非一致，难以横向对比，因此提出将处理每亿像素的平均时间作为时间效率的衡量参数，更科学合理。

# 六、工作小结

此次《行人重识别（ReID）算法计量评估方法》国家计量技术规范的编写，起草组对行人重识别算法的测试评价做了大量的评测工作，同时对算法的量值溯源技术开展了研究，提升了起草组技术人员在数字计量领域的计量技术能力，对人工智能算法的评价和不确定度评价方法奠定了基础。但由于人工智能算法的复杂性和不可解释性，所提出的评测方法难免存在一定的局限性。在之后将不断的根据专家的建议、行业的应用情况，不断完善数字量具和规范内容。

《行人重识别（ReID）算法测评方法》编制组

2023年10月