



(12) 发明专利

(10) 授权公告号 CN 114399552 B

(45) 授权公告日 2022.06.14

(21) 申请号 202210296702.3

(22) 申请日 2022.03.24

(65) 同一申请的已公布的文献号  
申请公布号 CN 114399552 A

(43) 申请公布日 2022.04.26

(73) 专利权人 武汉视合远方科技有限公司  
地址 430000 湖北省武汉市东湖新技术开发区流芳园南路9号光谷电子工业园4号厂房295号

(72) 发明人 何泽仪 李珍 洪勇 闫科 罗书培

(74) 专利代理机构 武汉蓝宝石专利代理事务所  
(特殊普通合伙) 42242  
专利代理师 范三霞

(51) Int.Cl.

G06T 7/73 (2017.01)

G06V 20/40 (2022.01)

G06V 40/10 (2022.01)

G06V 10/82 (2022.01)

G06N 3/04 (2006.01)

G08B 21/04 (2006.01)

(56) 对比文件

CN 112001308 A, 2020.11.27

CN 112633057 A, 2021.04.09

CN 107045623 A, 2017.08.15

CN 113256699 A, 2021.08.13

US 2021173045 A1, 2021.06.10

审查员 周锦

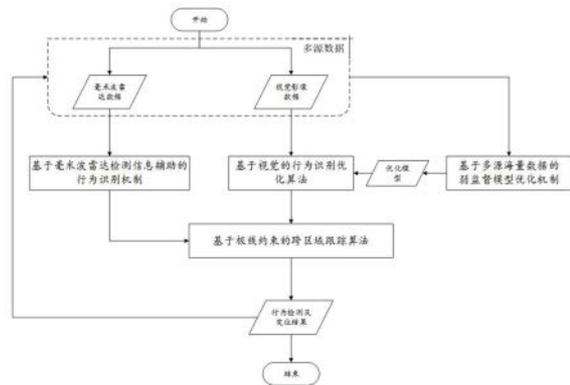
权利要求书2页 说明书10页 附图4页

(54) 发明名称

一种室内监护环境行为识别及定位方法

(57) 摘要

本发明涉及一种室内监护环境行为识别及定位方法,包括:获取多源视频,对每一源的视频进行预处理以获取样本集;将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息,将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息;通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号;根据行为识别信息、人头定位信息与ID号得到对应检测对象的行为分类信息。该方法通过多角度多源数据对检测对象进行行为识别以及人头定位,从而判断每个检测对象的准确的行为类别,以实现室内监护环境当中对于受照护者行为的有效智能化监护。



1. 一种室内环境行为识别及定位方法,其特征在于,包括:

获取多源视频,对每一源的视频进行预处理以获取样本集,包括:将每一源的视频按照相同时间间隔分割成若干段视频片段,在每段视频片段中随机采样一帧图像,随机采样时间机制采用离散时间傅里叶变换,每次采样得到的全部帧图像组成一个样本集;

将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息,包括:将样本集输入训练好的第一2D卷积神经网络,以分别获取样本集中每帧图像的特征图,将全部特征图进行堆叠,得到特征块;将所述特征块同时输入训练好的第二2D卷积神经网络以及3D卷积神经网络,在第二2D卷积神经网络中提取各帧的静态空间特征并进行堆叠,沿着帧维度对各帧的静态空间特征求平均值;在3D卷积神经网络中按照预设时间间隔跨帧提取特征块中的动态特征;将所述静态空间特征的平均值与所述动态特征进行拼接以得到行为识别信息,根据行为识别信息输出对应动作类别数目的一维向量;

将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息;

通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号;

根据行为识别信息、人头定位信息与ID号得到对应检测对象的行为分类信息;

获取毫米波雷达成像图,对毫米波雷达成像图提取特征以获取检测对象的行为识别辅助信息;

当通过视频检测的行为识别结果置信度较低时,通过所述行为识别辅助信息判断对应检测对象的行为分类信息,并采用得到的行为分类信息对视频检测的行为识别结果进行修正,修正后得到的数据用于优化行为识别模型。

2. 根据权利要求1所述一种室内环境行为识别及定位方法,其特征在于,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,包括:

将样本集输入训练好的目标检测模型,依次提取样本集中的每帧图像的人头特征信息并记录每个人头对应的像素位置,根据人头特征信息为每个人头分配唯一的ID号,输出连续帧中对应ID号的人头的像素位置变化记录,所述像素位置变化记录即人头定位信息。

3. 根据权利要求2所述一种室内环境行为识别及定位方法,其特征在于,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

为每个ID号分别构建跟踪模块,所述跟踪模块用于追踪连续帧中对应ID号的人头定位信息,通过人头定位信息计算当前人头在视频中的消失时间;

所述跟踪模块中设有消失时间阈值,当判定对应ID号人头的消失时间达到消失时间阈值,则从内存中清除该ID号对应人头的实例;当判定对应ID号人头的消失时间未达到消失时间阈值,则历史记录中保留该ID号对应人头的实例。

4. 根据权利要求3所述一种室内环境行为识别及定位方法,其特征在于,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

将相邻帧中各跟踪模块的人头定位信息进行交并比运算,将交并比值最大的人头定位信息进行同一ID迁移。

5. 根据权利要求3或4所述一种室内环境行为识别及定位方法,其特征在于,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

基于跟踪模块中同一ID号的人头定位信息,得到对应人头的高程信息,根据所述高程

信息计算连续帧中对应人头的高程变化信息,所述高程变化信息用于行为识别辅助判别。

6. 根据权利要求1所述一种室内环境行为识别及定位方法,其特征在于,所述的通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号,包括:

基于获取的多源视频的任意两个相机共视区域的位置关系构造对极线,求解两个相机图像的坐标归一化关系,通过坐标归一化关系获得相机之间点到极线一致性约束的基础矩阵;

根据所述基础矩阵以及检测对象的人头定位信息,将一个相机图像中任一检测对象映射到另一个相机图像中同一检测对象上,并将该检测对象的ID号赋值给另一个相机图像中对应的同一检测对象。

7. 根据权利要求1所述一种室内环境行为识别及定位方法,其特征在于,还包括:

当通过视频检测以及毫米波雷达检测得到的行为识别结果置信度均较低时,对行为识别结果进行人工校准标注,修正后得到的数据用于优化行为识别模型。

## 一种室内监护环境行为识别及定位方法

### 技术领域

[0001] 本发明涉及基于图像的行为识别及定位技术领域，具体涉及一种室内监护环境行为识别及定位方法。

### 背景技术

[0002] 由于老龄化和空心化社会的来临，独居老人、纯老人家庭、独居年轻人的数量也日益上升。而当前情势下，整个居家照护服务面临着社工短缺、人工监护成本过高及家人彼此之间照护困难等等问题。先看老年人所面对的困境：老人独自在家或者在养老机构日常活动时，容易发生意外，及时的监护与报警不可或缺。而在目前主流的以养老院、护工上门服务等养老模式下，老人在室内的日常活动能力和活动强度得不到有效的监护和评估，经常会导致健康情况监护的不细致、不准确，从而错失相关疾病的最佳检查、治疗周期。与此同时，近年来年轻人因工作压力大，睡眠不足等原因带来的健康隐患在独居环境下也有越演越烈的趋势。年轻人半夜猝死，晨发脑溢血等等新闻更是层出不穷。

[0003] 有鉴于此，有必要研究用于满足监护要求的室内定位与行为识别的技术，对人们室内独居时可能存在的意外情况予以监护。

### 发明内容

[0004] 本发明针对现有技术中存在的技术问题，提供一种室内监护环境行为识别及定位方法，通过多角度多源数据对检测对象进行行为识别以及人头定位，从而判断每个检测对象的准确的行为类别，以实现室内监护环境当中对于受照护者行为的有效智能化监护。

[0005] 本发明解决上述技术问题的技术方案如下：

[0006] 本发明提供一种室内监护环境行为识别及定位方法，包括：

[0007] 获取多源视频，对每一源的视频进行预处理以获取样本集，

[0008] 将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息，将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息；

[0009] 通过极线约束的方式将多源视频相关联，通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号；

[0010] 根据行为识别信息、人头定位信息与ID号得到对应检测对象的行为分类信息。

[0011] 在上述技术方案的基础上，本发明还可以做如下改进。

[0012] 进一步，所述的对每一源的视频进行预处理以获取样本集，包括：

[0013] 将每一源的视频按照相同时间间隔分割成若干段视频片段，在每段视频片段中随机采样一帧图像，随机采样时间机制采用离散时间傅里叶变换，每次采样得到的全部帧图像组成一个样本集。

[0014] 进一步，所述的将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息，包括：

[0015] 将样本集输入训练好的第一2D卷积神经网络，以分别获取样本集中每帧图像的特

征图,将全部特征图进行堆叠,得到特征块;

[0016] 将所述特征块同时输入训练好的第二2D卷积神经网络以及3D卷积神经网络,在第二2D卷积神经网络中提取各帧的静态空间特征并进行堆叠,沿着帧维度对各帧的静态空间特征求平均值;在3D卷积神经网络中按照预设时间间隔跨帧提取特征块中的动态特征;

[0017] 将所述静态空间特征的平均值与所述动态特征进行拼接以得到行为识别信息,根据行为识别信息输出对应动作类别数目的一维向量。

[0018] 进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,包括:

[0019] 将样本集输入训练好的目标检测模型,依次提取样本集中的每帧图像的人头特征信息并记录每个人头对应的像素位置,根据人头特征信息为每个人头分配唯一的ID号,输出连续帧中对应ID号的人头的像素位置变化记录,所述像素位置变化记录即人头定位信息。

[0020] 进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

[0021] 为每个ID号分别构建跟踪模块,所述跟踪模块用于追踪连续帧中对应ID号的人头定位信息,通过人头定位信息计算当前人头在视频中的消失时间;

[0022] 所述跟踪模块中设有消失时间阈值,当判定对应ID号人头的消失时间达到消失时间阈值,则从内存中清除该ID号对应人头的实例;当判定对应ID号人头的消失时间未达到消失时间阈值,则历史记录中保留该ID号对应人头的实例。

[0023] 进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

[0024] 将相邻帧中各跟踪模块的人头定位信息进行交并比运算,将交并比值最大的人头定位信息进行同一ID迁移。

[0025] 进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

[0026] 基于跟踪模块中同一ID号的人头定位信息,得到对应人头的高程信息,根据所述高程信息计算连续帧中对应人头的高程变化信息,所述高程变化信息用于行为识别辅助判别。

[0027] 进一步,所述的通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号,包括:

[0028] 基于获取的多源视频的任意两个相机共视区域的位置关系构造对极线,求解两个相机图像的坐标归一化关系,通过坐标归一化关系获得相机之间点到极线一致性约束的基础矩阵;

[0029] 根据所述基础矩阵以及检测对象的人头定位信息,将一个相机图像中任一检测对象映射到另一个相机图像中同一检测对象上,并将该检测对象的ID号赋值给另一个相机图像中对应的同一检测对象。

[0030] 进一步,该方法还包括:

[0031] 获取毫米波雷达成像图,对毫米波雷达成像图提取特征以获取检测对象的行为识别辅助信息;

[0032] 当通过视频检测的行为识别结果置信度较低时,通过所述行为识别辅助信息判断对应检测对象的行为分类信息,并采用得到的行为分类信息对视频检测的行为识别结果进行修正,修正后得到的数据用于优化行为识别模型。

[0033] 进一步,该方法还包括:

[0034] 当通过视频检测以及毫米波雷达检测得到的行为识别结果置信度均较低时,对行为识别结果进行人工校准标注,修正后得到的数据用于优化行为识别模型。

[0035] 本发明的有益效果是:本发明提出了一种室内监护环境行为识别及定位方法,其融合了多源数据(例如多源视频、毫米雷达波)的行为识别及定位技术,构建了轻量化的行为检测网络,便于室内监护特别是养老场景的大范围部署实施。通过离散采样融合二三维卷积神经网络实现视频数据帧精简及信息提取融合的方式,可以提升行为识别算法的全局准确性,实现关联多跨度连续帧进行行为识别并提升识别效率;行为识别算法直接在特征图层面进行2D及3D的信息融合,具备算法高效性、可移植性,算法支持嵌入式端轻量级部署,以实现基于嵌入式端(Nvidia-nano)部署检测,节省数据传输及中心算力的成本;联合交并比(IoU)判定及目标关联机制,通过特定目标(例如人头)的检测与跟踪来产生行为识别事件,产生另一个维度的行为识别判据,进一步提高行为识别的准确度;通过极线约束的方式实现跨摄像头的目标重识别,可实现视觉全覆盖跟踪;基于毫米波雷达辅助行为检测,进一步提高系统的行为检测及定位精度;基于实际场景的多源信息对海量室内行为数据进行修正,实现行为检测数据集的生成及优化,同时基于最新数据以优化模型,实现行业数据、模型及算法的循环优化体系,从而提高识别精度,实现整个系统的高效性及稳定性。

## 附图说明

[0036] 图1为本发明视频数据行为识别方法流程图;

[0037] 图2为本发明的行为识别模型网络结构图;

[0038] 图3为本发明的目标检测模型IoU运算判别图解;

[0039] 图4为本发明的对极线构造原理图解;

[0040] 图5为本发明通过极线约束进行图像关联的推理图例;

[0041] 图6为本发明视觉检测结合毫米雷达波检测的流程图。

## 具体实施方式

[0042] 以下结合附图对本发明的原理和特征进行描述,所举实例只用于解释本发明,并非用于限定本发明的范围。

[0043] 为了进一步对本发明进行举例说明,首先在室内监护场景布置至少两个相机(摄像头),以获取监护场景中的多源视频数据。其中,任意两个相机具有一定的共视区域。本实施例以采用两个相机获取两源视频进行举例。

[0044] 本实施例提供一种室内监护环境行为识别及定位方法,如图1所示,包括:

[0045] 获取多源视频,对每一源的视频进行预处理以获取样本集,

[0046] 将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息,将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息;

[0047] 通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息

为每个检测对象分配唯一的ID号；

[0048] 根据行为识别信息、人头定位信息与ID号得到对应检测对象的行为分类信息。

[0049] 可以理解的是,本实施例中融合了多源视频数据的行为识别及定位技术,构建了轻量化的行为检测网络,便于室内监护,特别是养老场景的大范围部署实施。在视频数据处理层面,将每一源的视频数据分别提取样本集以进行特征提取、识别每一源视频中检测对象的行为识别信息以及人头定位信息;然后将任意一源的视频作为参考视频,根据人头定位信息为此视频中的每一个检测对象分配唯一的ID号;通过极线约束的方式将另一源视频与参考视频进行关联,并将不同源视频中的检测对象进行一一对应,然后将参考视频中检测对象的ID号赋值给关联视频中同一检测对象。最后将检测到的多源视频中同一ID检测对象的行为识别信息及其人头定位信息进行融合,综合判断检测对象的行为,并输出其行为分类信息。该方法通过实现关联视频进行行为识别,提升了行为识别算法的全局准确性。通过多源视频的相关联进行行为检测,适用于室内场景的大范围部署实施。

[0050] 在上述技术方案的基础上,本实施例还可以做如下改进。

[0051] 目前已有的基于视觉视频实现行为识别的算法,普遍存在以下问题:

[0052] (1) 模型在做任务推理时只关注视频局部信息,忽略了横跨一定时间间隔的动作联系;

[0053] (2) 已有模型只在视频局部处理进行效率改进,但是视频整体的处理效率较低,无法用于快速的视频检索或者在线长期动作分类;

[0054] 针对以上问题,本发明提出了一种更加高效的视觉行为识别算法,思路如下文所述。

[0055] 在一种可能的实施例中,所述的对每一源的视频进行预处理以获取样本集,包括:

[0056] 将每一源的视频按照相同时间间隔分割成若干段视频片段,在每段视频片段中随机采样一帧图像,随机采样时间机制采用离散时间傅里叶变换,每次采样得到的全部帧图像组成一个样本集。

[0057] 可以理解的是,对于视觉视频信号,由于相邻图像帧之间存在大量的信息冗余,基于冗余信息的数据进行训练,需要耗费大量的计算资源,且极大增加了标注的难度。因此本方案提出了一种基于离散采样的推理方式,进行数据处理。即在对某一源的视频进行检测前,需对其进行离散采样预处理以获取用于行为识别的图像样本集。

[0058] 在具体训练模型时,视频被分成n小段,每小段时间长度相等。每次用某段视频进行训练时,在每个时间段内随机采样一帧视频图像,采样得到的全部图像组成一个样本集,用于后续步骤中输入卷积网络。随机采样机制选择离散时间傅里叶变换,如以下公式:

$$\begin{aligned}
 F(\omega) &= \int_{-\infty}^{\infty} \left[ \sum_{n=-\infty}^{\infty} f(t)\delta(t - nT) \right] e^{-i\omega t} dt \\
 [0059] \quad &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} f(t)\delta(t - nT)e^{-i\omega t} dt \quad 。 \\
 &= \sum_{n=-\infty}^{\infty} f(nT)e^{-i\omega nT}
 \end{aligned}$$

[0060] 对于长段落视频训练而言,视频中的每一帧都得到了充分的利用。这样训练数据更加多样化,让模型能够适应视频中动作在各个瞬间的微妙变化。

[0061] 就单次前向和反向传播而言,模型对一段视频只需要处理n帧图像,因此模型的运行速度很快。

[0062] 在一种可能的实施例中,所述的将样本集输入训练好的行为识别模型以获得检测对象的行为识别信息,包括:

[0063] 将样本集输入训练好的第一2D卷积神经网络,以分别获取样本集中每帧图像的特征图,将全部特征图进行堆叠,得到特征块;

[0064] 将所述特征块同时输入训练好的第二2D卷积神经网络以及3D卷积神经网络,在第二2D卷积神经网络中提取各帧的静态空间特征并进行堆叠,沿着帧维度对各帧的静态空间特征求平均值;在3D卷积神经网络中按照预设时间间隔跨帧提取特征块中的动态特征;

[0065] 将所述静态空间特征的平均值与所述动态特征进行拼接以得到行为识别信息,根据行为识别信息输出对应动作类别数目的一维向量。

[0066] 可以理解的是,本实施例中,模型在前半段利用现有的预训练的图像分类模型(第一2D卷积神经网络)对视频各帧图像进行各自独立的特征提取,然后将各帧得到的特征进行堆叠,在后半段利用3D卷积神经网络进行跨帧的特征提取,以提取动态特征。同时,本方案在后半段加也上了第二2D卷积神经网络,与3D卷积神经网络平行,专门提取视频的静态空间特征,如图2所示。对于2D卷积网络提取的各帧表示,将它们堆叠在一起,沿着帧维度求平均值,得到的向量与3D卷积神经网络的输出拼接在一起。最后利用3D卷积网络提取出的视频表示和第二2D卷积网络提取视频表示输出视频的行为分类结果。

[0067] 如图2所示,S1~SN是从某一源视频中采样得到的N个RGB格式的视频片段。

[0068] 对于每个片段,采用共享的第一2D卷积子网络来得到96个28\*28大小的特征图(feature map),将特征图堆叠后得到一个N\*28\*28\*96 大小的特征块(volume)。

[0069] 对于得到的特征块(volume),采用一个3D子网络进行处理,直接输出对应动作类别数目的一维向量。

[0070] 这种后半段2D卷积神经网络和3D卷积神经网络并行的结构做到了分工合作,实现了“1+1>2”的效果,一方面2D卷积神经网络专注于提取图像的静态特征,保证在最后分类的中视频的静态空间信息得到充分的重视,另一方面3D卷积神经网络专注于提取视频复杂的时序特征(动态特征),最终可使行为识别模型能够处理与分辨多段时间内发生的动作,实现基于视觉的行为识别。

[0071] 进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,包括:

[0072] 将样本集输入训练好的目标检测模型,依次提取样本集中的每帧图像的人头特征信息并记录每个人头对应的像素位置,根据人头特征信息为每个人头分配唯一的ID号,输出连续帧中对应ID号的人头的像素位置变化记录,所述像素位置变化记录即人头定位信息。

[0073] 可以理解的是,本方案实现了对特定目标进行捕捉及获取实时动态信息的一种检测方法。以跌倒检测为例,本方案采用的Yolo v5l模型训练的人头检测模型去检测室内环境的人头,实现基于人头的捕捉、ID号分配及定位,同时基于实时定位信息,判断相关事件

是否发生,辅助行为算法进行判断。

[0074] 检测人头的初始化ID分配策略:由于该工程事先不清楚当前场景下的人员信息,因此需要在检测后对人头进行ID分配,确保每个相机下的人头的ID号唯一,确保了跟踪模块ID的唯一性和解算模块人头坐标的唯一性。使用全局变量global\_id为0,相机摄像头的影像依次传入,对每个摄像头依次进行检测操作,每检测出一个人头,为其分配全局的ID号,同时global\_id进行加1操作。这样,一个摄像头检测到的每个人头就具有唯一的ID号,即,同源视频中检测到的人头具有唯一的ID号。

[0075] 在一种可能的实施例中,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,更具体的,还包括:

[0076] 为每个ID号分别构建跟踪模块,所述跟踪模块用于追踪连续帧中对应ID号的人头定位信息,通过人头定位信息计算当前人头在视频中的消失时间;

[0077] 所述跟踪模块中设有消失时间阈值,当判定对应ID号人头的消失时间达到消失时间阈值,则从内存中清除该ID号对应人头的实例;当判定对应ID号人头的消失时间未达到消失时间阈值,则历史记录中保留该ID号对应人头的实例。

[0078] 具体的,定义一个Tracker类(即跟踪模块),记录目标检测(例如人头检测)的相关信息,其中包括目标像素位置、ID号及对应定位记录(定位算法详见后续方案)等信息。

[0079] Tracker类中的消失时间记录当前追踪ID消失的时间,如果当前ID在有效时间内存在(即判定对应ID号人头的消失时间未达到消失时间阈值),则历史记录会保留当前的Tracker实例;若检测到当前ID消失的时间超出了消失时间阈值,则会从内存中清除当前的ID实例。可以理解的是,这种时间策略确保了检测时由于检测模型的缘故,时不时出现漏检、但在该位置又存在人头的情况,同时清除了长时间未检测到的ID,确保了计算机内存不会由于积累的Tracker而导致内存溢出。

[0080] 更进一步,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

[0081] 将相邻帧中各跟踪模块的人头定位信息进行交并比(IOU)运算,将交并比值最大的人头定位信息进行同一ID迁移。

[0082] 可以理解的是,如图3所示,Tracker类中的定位记录记录了当前目标框的位置,通过上一帧记录的Tracker类中的定位记录与当前检测到的定位记录进行IOU(交并比)的计算,若该计算值超过了设定的阈值,就对当前帧检测到的目标框进行同一ID迁移的操作(即将当前生成的Tracker类中的ID属性设置为与上一帧相比IOU值最大的Tracker类的ID的数值)。公式为:

$$[0083] \quad \frac{A \cap B}{A \cup B} > \text{阈值}。$$

[0084] 在一种可能的实施例中,所述的将样本集输入训练好的目标检测模型以获得检测对象的人头定位信息,还包括:

[0085] 基于跟踪模块中同一ID号的人头定位信息,得到对应人头的高程信息,根据所述高程信息计算连续帧中对应人头的高程变化信息,所述高程变化信息用于行为识别辅助判别。

[0086] 可以理解的是,基于Tracker类中的同一ID目标的定位记录,可实时获取场景中目标人头的高程信息,可头连续帧内高程变化的速度来辅助判断检测对象的行为。例如,当连续帧内,检测对象的人头高程发生急剧变化时(例如从预设高程降低为0等),产生跌倒警报。

[0087] 基于以上思路,可针对检测对象的多种目标,基于连续帧目标检测的思路,实现相关的行为判断,从而达到行为检测的目的,为一种高效且易实现的行为识别辅助判别手段。

[0088] 本方案中优化的行为识别算法直接在特征图层面进行3D及2D的信息融合,具备算法高效性、可移植性;而对应的检测辅助策略选取基于yolo v5-small模型进行训练及推理,可实现基于嵌入式端(Nvidia-nano)部署检测,达到每秒超过10帧的处理速度,整体方案满足轻量化模型的设计思维,可实现边缘端部署及推理,节省数据传输及中心算力的成本。

[0089] 人在室内进行活动时,需要对其进行全区域的摄像头跟踪及行为识别,在包括养老在内的室内监护场景中,为了达到全场景覆盖,会涉及到跨摄像头的人员重识别问题。本实施例提出了一种基于共视区域的人员重识别方法,基于两个摄像头进行基础矩阵的解算进行同一个人的关联,并进行同一个人ID号的迁移,使得整个室内监护场景中同一检测对象具有唯一的ID号,便于行为识别的准确性。

[0090] 在一种可能的实施例中,所述的通过极线约束的方式将多源视频相关联,通过每一源视频中得到的人头定位信息为每个检测对象分配唯一的ID号,包括:

[0091] 基于获取的多源视频的任意两个相机共视区域的位置关系构造对极线,求解两个相机图像的坐标归一化关系,通过坐标归一化关系获得相机之间点到极线一致性约束的基础矩阵;

[0092] 根据所述基础矩阵以及检测对象的人头定位信息,将一个相机图像中任一检测对象映射到另一个相机图像中同一检测对象上,并将该检测对象的ID号赋值给另一个相机图像中对应的同一检测对象。

[0093] 理论上两个相机具有大部分的共视区域,共视区域点到同一个人(检测对象)的关联需要一个判断依据,因此基于极线约束,一张图像上的点在它匹配的图像上对应的极线一定穿过该同名点,此时匹配对图像上就存在极线,那便可以通过匹配对上的点到直线的距离作为衡量的标准来进行计算。具体的,可对这两个相机进行基础矩阵的计算,解算出基础矩阵,即可通过基础矩阵将一个相机图像中的检测对象映射到另一个相机图像中同一检测对象上。

[0094] 首先是进行对极线的构造,即计算出基础矩阵。

[0095] 如图4所示,对对极线的原理进行说明。

[0096] 相关基础概念如下:

[0097] 基线:两个相机中心的连线 $CC'$ 称为基线。

[0098] 对极点: $ee'$ 为对极点,是基线与两个成像平面的交点,即两个相机在另一个成像平面上的像点。

[0099] 对极平面:过基线的平面都被称为对极平面 $\pi$ ,两个相机的中心 $C$ 和 $C'$ 、三维点 $X$ 、以及其在两个相机的像点 $(x, x')$ ,这5点必定在同一个对极平面 $\pi$ 上。当三维点 $X$ 变化时,对极平面 $\pi$ 绕着基线旋转,形成对极平面束。

[0100] 对极线:对极平面 $\pi$ 和成像平面的交线,所有的对极线相交于极点。

[0101] 如图5所示,对极平面 $\pi$ 不通过两相机中心,过第一个相机的中心C和像点x的射线与对极平面 $\pi$ 相交于点X。该点X再投影到第二幅图像上得到像点 $x'$ ,这个过程称为点通过平面的转移。因此,构造出对极线 $l'$ ,即可通过对极线 $l'$ 将一个相机图像上的点和线映射到另一个相机图像上,从而实现不同源视频之间的关联。

[0102] 点X位于像点x和相机中心C确定的射线上,其在另一幅图像上的像点 $x'$ 必然位于该射线在另一幅图像的投影也就是对极线 $l'$ 上。点x和点 $x'$ 都是三维点X的像点,这样第一幅图像上的像点集合 $x_i$ 和第二幅图像上的像点集合 $x'_i$ 是射影等价的,它们都射影等价于共面的三维点集合 $X_i$ 。因此,存在一个2D单应矩阵 $H_\pi$ ,把每一个点x映射到对应的点 $x'$ 上。

[0103] 对对极线进行构造的过程如下:

[0104] 基于点 $x'$ ,通过 $x'$ 和对极点 $e'$ 的对极线 $l'$ 可表示为:

$$[0105] \quad l' = e' \times x' = [e']_{\times} \cdot x',$$

[0106] 由于 $x' = H_\pi x$  ( $H_\pi$ 是将x变换为 $x'$ 的单应性矩阵),带入上式可得:

$$[0107] \quad l' = [e']_{\times} H_\pi x,$$

[0108] 定义基础矩阵 $F = [e']_{\times} H_\pi$ ,则有

$$[0109] \quad l' = Fx.$$

[0110] 图像中由匹配的对点坐标组成的矩阵记为系数矩阵A, $Af=0$ ,其中f为八点法(通过匹配点估算基础矩阵的方法)的过程量。因为矩阵各列的数据尺度差异太大,最小二乘得到的结果精度一般很低,所以要对各个列向量进行归一化操作。

[0111] 图像坐标归一化操作步骤如下:

[0112] (1) 对点进行平移使其形心位于原点;

[0113] (2) 对点进行缩放,使它们到原点的平均距离为 $2\sqrt{2}$ ;

[0114] (3) 对两幅图像独立进行上述变换。

[0115] 具体的,设H是归一化的变换矩阵,可记为如下形式:

$$[0116] \quad H = S \begin{bmatrix} 1 & 0 & -\bar{\mu} \\ 0 & 1 & -\bar{\nu} \\ 0 & 0 & \frac{1}{S} \end{bmatrix},$$

[0117] 其中, $\bar{\mu}$ 与 $\bar{\nu}$ 是图像像点坐标两个分量的平均值,S表示尺度,其表达式为:

$$[0118] \quad S = \frac{\sqrt{2 \cdot N}}{\sqrt{\sum_{i=1}^N (\mu_i - \bar{\mu})^2 + (\nu_i - \bar{\nu})^2}}.$$

[0119] 这样,首先对原始的图像坐标进行归一化处理,再利用八点法求解基础矩阵,最后将求得的结果解除归一化,得到基础矩阵F。八点法为现有技术,本专利不再赘述。

[0120] 得到基础矩阵F后,即可根据所述基础矩阵F以及检测对象的人头定位信息,将一个相机图像中任一检测对象映射到另一个相机图像中同一检测对象上,并将该检测对象的

ID号赋值给另一个相机图像中对应的同一检测对象,使得两个相机视野中的同一检测对象具有唯一的ID号,便于对同一检测对象的行为进行判断,也对检测对象的行为识别结果进行分类输出。

[0121] 为了提高不同时段、光照条件下行为识别结果的准确性,本方案在搭建的室内监护场景中架设了具备相关行为检测的毫米波雷达,用以辅助行为识别检测。毫米波雷达本身具备相关检测与定位功能,但是存在数据的不稳定性,人数及空间局限性等缺点。在本方案中毫米波雷达安装在特定敏感区域,并将其数据与相关视觉监测数据均由基于zigbee协议接入到智能网关,以视觉数据为主要判别依据,而以毫米波雷达识别结果为辅助,经过融合判别机制,最终实现全场景覆盖的行为识别功能。

[0122] 具体的,如图6所示,毫米波雷达识别结合视觉监测的方法包括:

[0123] 获取毫米波雷达成像图,对毫米波雷达成像图提取特征以获取检测对象的行为识别辅助信息;

[0124] 当通过视频检测的行为识别结果置信度较低时,通过所述行为识别辅助信息判断对应检测对象的行为分类信息,并采用得到的行为分类信息对视频检测的行为识别结果进行修正,修正后得到的数据用于优化行为识别模型。同样的,当通过视频检测的行为识别结果置信度较高时,以视频检测的行为识别结果为准。

[0125] 本实施例的方法基于环境中摄像头的外参标定,可将毫米波雷达检测覆盖区域和摄像头覆盖区域进行强关联,从而可以获得两种特殊行为状态的判定依据;毫米波雷达与视觉摄像头均基于zigbee协议接入到智能网关,经相关配置可实现两个设备的时间同步;行为识别存在自有的置信度机制,当存在置信度较低的行为检测结果时,可关联毫米波检测结果进行联合判定,提高行为检测结果的准确性;视觉数据通过与毫米波雷达检测结果的配准,可对置信度较低的行为检测结果进行置信度修正,从而形成可用于优化模型的新数据集,基于该机制可实现数据+模型优化的内循环。

[0126] 该方法具有高效性、高精度,且可根据视觉、毫米波数据,相互配准,优化算法模型训练集。

[0127] 本实施例中,该方法还包括:

[0128] 当通过视频检测以及毫米波雷达检测得到的行为识别结果置信度均较低时,可对行为识别结果进行人工校准标注,修正后得到的数据用于优化行为识别模型。

[0129] 本方案通过多样室内监护视频数据进行算法训练及优化,因此在数据集及模型层面,本方案具备先进性及普适性。同时本方案设计了弱监督数据校准平台,基于视觉视频检测及毫米波雷达检测数据相互配准的优化算法模型训练集,可选择地进行人工校准标注(弱监督),实现对于整个居家照护行业相关数据集的集成、标注、校准及制作,同时基于最新数据优化模型,实现行业数据、模型及算法的循环优化体系。

[0130] 工作原理:

[0131] 本发明实施例提出了一种室内监护环境行为识别及定位方法,其融合了多源数据(例如多源视频、毫米波雷达等)的行为识别及定位技术,构建了轻量化的行为检测网络,便于室内监护场景的大范围部署实施。通过离散采样融合二三维卷积神经网络实现视频数据帧精简及信息提取融合的方式,可以提升行为识别算法的全局准确性,实现关联多跨度连续帧进行行为识别并提升识别效率;行为识别算法直接在特征图层面进行2D及3D的信息融

合,具备算法高效性、可移植性,算法支持嵌入式端轻量级部署,以实现基于嵌入式端(Nvida-nano)部署检测,节省数据传输及中心算力的成本;联合交并比(IOU)判定及目标关联机制,通过特定目标(例如人头)的检测与跟踪来产生行为识别事件,产生另一个维度的行为识别判据,进一步提高行为识别的准确度;通过极线约束的方式实现跨摄像头的目标重识别,可实现视觉全覆盖跟踪;基于毫米波雷达辅助行为检测,进一步提高系统的行为检测及定位精度;基于实际场景的多源信息对海量监护人行为数据进行修正,实现行为检测数据集的生成及优化,同时基于最新数据以优化模型,实现行业数据、模型及算法的循环优化体系,从而提高识别精度,实现整个系统的高效性及稳定性。

[0132] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

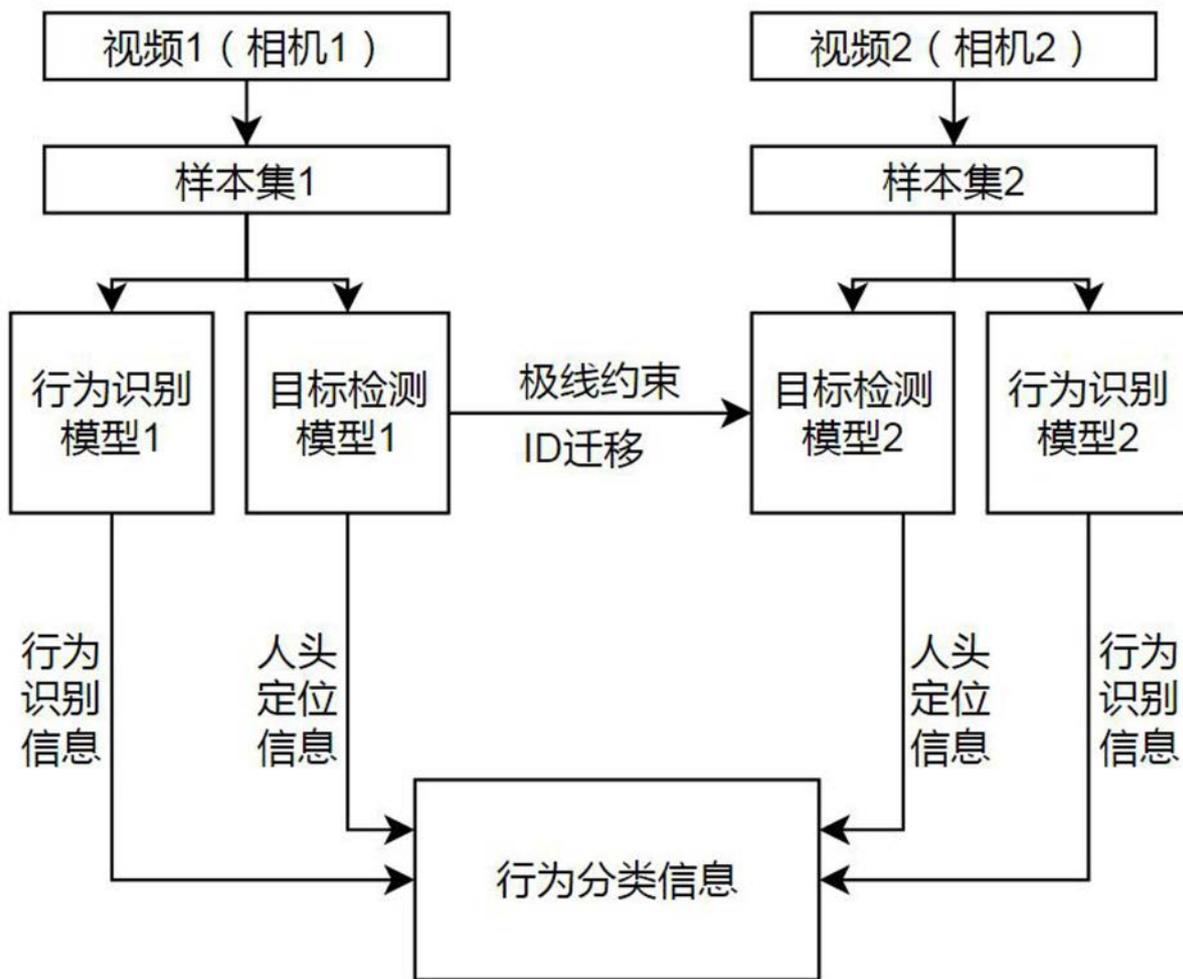


图1

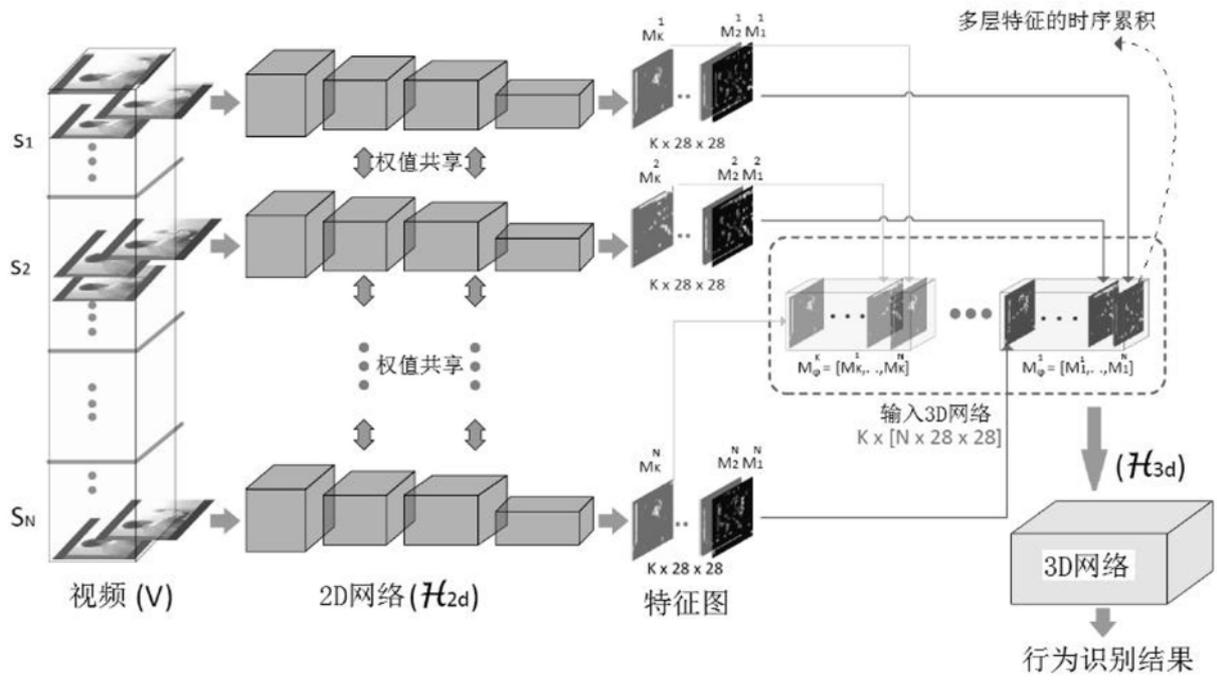


图2

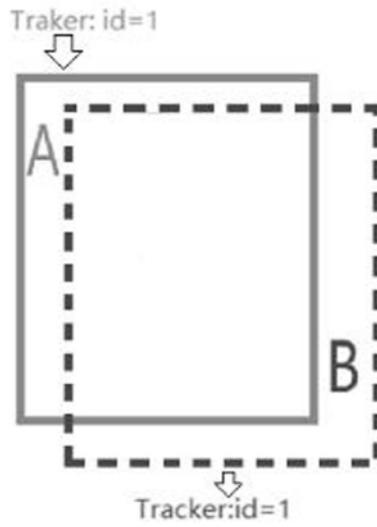


图3

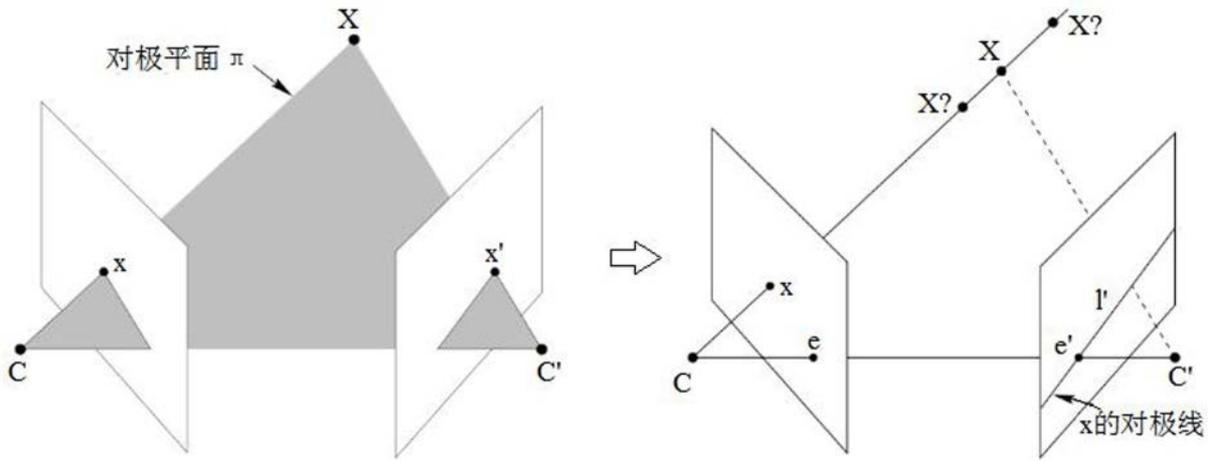


图4

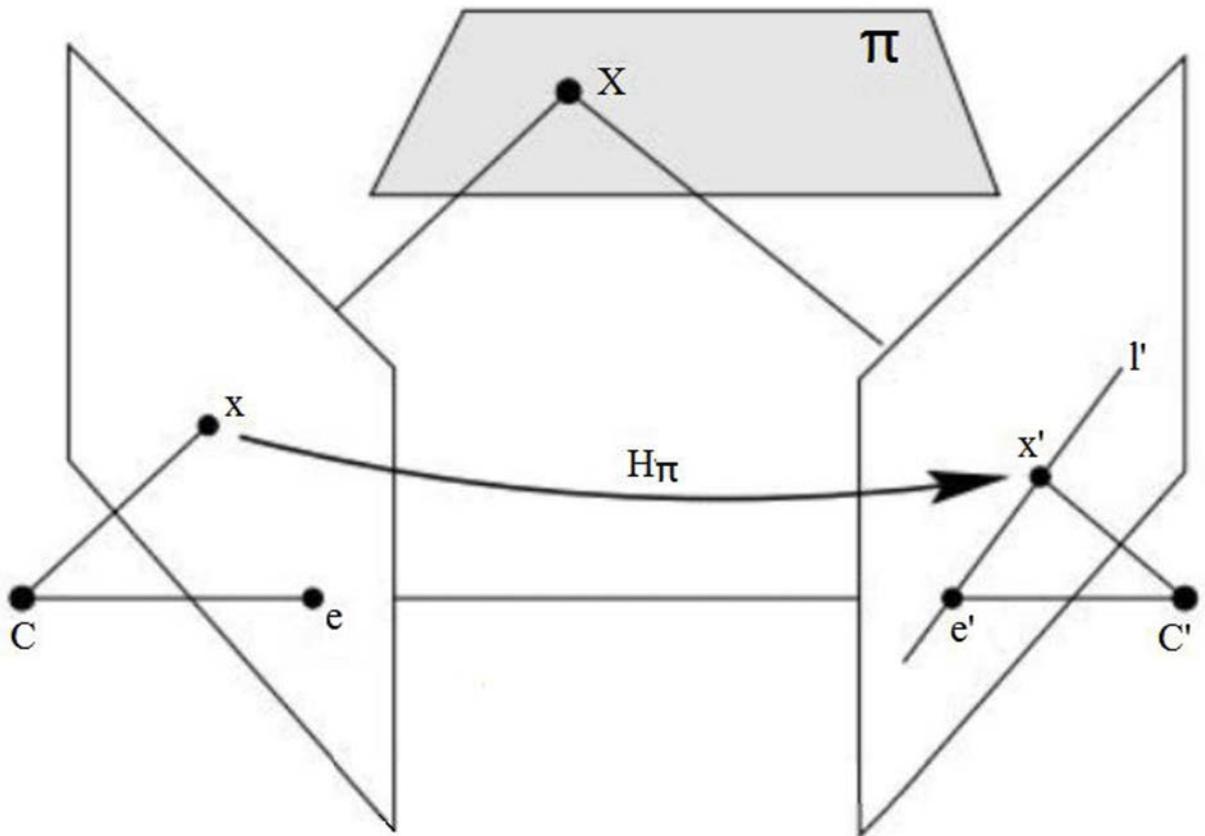


图5

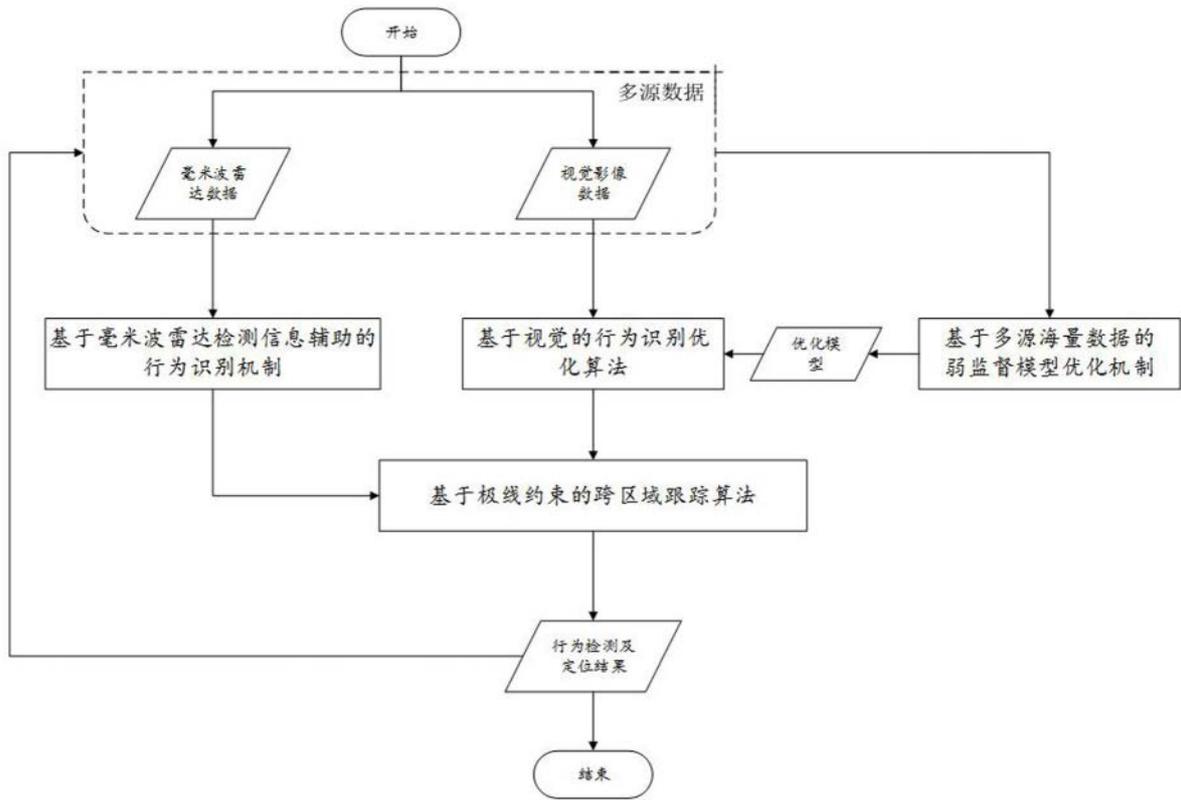


图6