



(12) 发明专利

(10) 授权公告号 CN 107256377 B

(45) 授权公告日 2021.03.16

(21) 申请号 201710262130.6

(22) 申请日 2013.09.12

(65) 同一申请的已公布的文献号
申请公布号 CN 107256377 A

(43) 申请公布日 2017.10.17

(30) 优先权数据
61/700,033 2012.09.12 US
13/838,511 2013.03.15 US

(62) 分案原申请数据
201380047668.5 2013.09.12

(73) 专利权人 威智伦富智堡公司
地址 加拿大不列颠哥伦比亚省

(72) 发明人 张忠 尹卫红
彼得·为奈蒂阿奈尔

(74) 专利代理机构 北京鸿元知识产权代理有限公司 11327

代理人 李琳 许向彤

(51) Int.Cl.
G06K 9/00 (2006.01)
G06T 7/73 (2017.01)

审查员 杜学惠

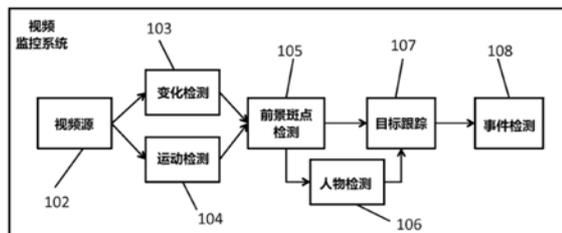
权利要求书3页 说明书18页 附图26页

(54) 发明名称

用于检测视频中的对象的方法、设备和系统

(57) 摘要

公开了用于进行视频内容分析以检测视频图像中感兴趣的人物或其他对象的方法、设备和系统。检测人物可以用于对人数进行计数、确定每个人的位置并且/或者对监测区域进行人群分析。



101

1. 一种检测视频中的人物对象的方法,包括:

确定作为前景像素的视频图像的像素,所述前景像素的组构成一个或多个前景斑点的前景斑点集;

对于所述视频图像中N个预定位置的对应的预定位置处的N个预定形状,其中N是大于1的正整数,通过将对应的预定形状的重叠与所述前景斑点集进行比较来计算第一值,所述第一值用于获得所述对应的预定位置处的人物的对应的概率,从而获得与N个预定位置对应的N个概率;

使用所述N个概率,确定由所述前景斑点集代表的X个人物,其中X是整数;并且

使用所述X个人物的代表的确定来提供报告、警报和事件检测的至少一种,

其中,响应于视频系统的校正来确定所述N个预定位置的每个位置的所述对应的预定形状的大小,并且

其中,对于所述N个预定形状,所述对应的预定形状与所述前景斑点集的比较包括分析所述对应的预定形状与所述前景斑点集的重叠面积的量。

2. 根据权利要求1所述的方法,进一步包括使用所述N个概率以确定所述X个人物的每个人物的位置。

3. 根据权利要求2所述的方法,其中,所述X个人物的每个人物的确定的所述位置是与所述视频图像对应的图像平面中的位置。

4. 根据权利要求2所述的方法,其中,所述X个人物的每个人物的确定的所述位置是相对于与真实世界对应的物理地平面的位置。

5. 根据权利要求1所述的方法,

其中,确定所述视频图像的前景像素包括将没有前景对象的视频图像的第一帧与包括所述前景对象的所述视频图像的第二帧进行比较。

6. 根据权利要求1所述的方法,其中,所述N个预定位置的每个位置的所述预定形状相同。

7. 根据权利要求1所述的方法,其中,所述N个预定位置的至少一些位置的所述预定形状具有不同大小。

8. 根据权利要求1所述的方法,

其中,所述视频系统的校正包括确定所述视频图像的与所述N个预定位置的每个位置的平均人物大小对应的一部分的图像大小,并且

其中,响应于对应的所述图像大小来确定所述N个预定位置的每个位置的所述预定形状的大小。

9. 根据权利要求1所述的方法,

进一步包括:在确定作为所述前景像素的所述视频图像的所述像素之前,对于所述N个预定位置的每个位置,通过估计在人物存在于所述对应的位置时在所述视频图像中要被占据的前景图像部分来确定所述对应的预定形状。

10. 根据权利要求9所述的方法,其中,基于将真实世界中的人物模型投影到所述视频图像的图像平面上来估计所述N个预定位置的每个位置的所述前景图像部分。

11. 根据权利要求1所述的方法,

其中,所述视频图像包括多个图像帧,每个图像帧包括具有所述N个预定位置的二维图

像,所述N个预定位置的每个位置通过所述二维图像中的对应x、y坐标对来进行识别。

12. 根据权利要求11所述的方法,

其中,所述N个预定位置的每个位置与相对于与所述视频图像对应的图像平面的所述N个预定形状的对应的预定形状相关。

13. 根据权利要求1所述的方法,进一步包括,对于所述N个预定位置的每个位置,计算所述对应的预定形状和所述前景斑点集的调用率以确定相关的概率。

14. 根据权利要求13所述的方法,其中,对于所述N个预定位置的每个位置,计算所述调用率包括确定 (a) 包括所述预定形状占据的面积与所述前景斑点集占据的面积的重叠面积与 (b) 所述预定形状占据的面积之比。

15. 根据权利要求1所述的方法,进一步包括:

使用所述N个概率创建概率映射;以及

确定所述概率映射的概率的局部最大值。

16. 根据权利要求15所述的方法,进一步包括:

选择所述N个预定位置中的与所述概率映射的局部最大值对应的第一位置;

获得与所述第一位置对应的第一预定形状;以及

分析所述第一预定形状和所述前景斑点集占据的重叠面积的量。

17. 根据权利要求16所述的方法,进一步包括:

计算 (a) 包括所述第一预定形状和所述前景斑点集占据的重叠面积的面积与 (b) 所述前景斑点集的面积的第一比值,

其中,所述第一比值用于确定由所述前景斑点集代表的X个人物。

18. 根据权利要求15所述的方法,进一步包括计算所述N个预定位置中的m个位置的每个位置的精度值和调用值,m为整数,所述m个位置的每个位置与所述概率映射的局部最大值对应。

19. 根据权利要求1所述的方法,进一步包括:

基于所述N个概率选择所述N个预定形状的子集;以及

分析所选择的所述N个预定形状的子集占据的区域与所述前景斑点集占据的区域的重叠。

20. 根据权利要求1所述的方法,进一步包括:顺序地选择所述N个预定位置的1至m个位置,第(m-1)个位置的选择排除了落入所述第(m-1)个位置的第一预定距离内的所述N个预定位置的随后的一个位置的选择。

21. 根据权利要求1所述的方法,进一步包括:顺序地选择所述N个预定位置的1至m个位置,其中,所述N个预定位置的下一个位置的选择包括基于其与所述视频图像的底部边缘的接近度来选择位置。

22. 一种检测视频中的人物对象的方法,包括:

确定作为前景像素的真实世界场景的视频图像的像素,所述前景像素的组构成一个或多个前景斑点的前景斑点集;

对于所述视频图像中N个预定位置的对应的预定位置处的N个预定形状的每个预定形状,其中,N是大于1的正整数,通过将对应的预定形状的重叠与所述前景斑点集进行比较来计算第一值,所述第一值用于确定由所述前景斑点集代表的X个人物,其中,X是整数,并且

所述X个人物的每个人物的位置被确定为真实世界的水平面内的位置;以及

当人群密度超过阈值时,使用所述X个人物的代表的确定来提供报告、警报和事件检测的至少一种,

其中,响应于视频系统的校正来确定所述N个预定位置的每个位置的所述对应的预定形状的大小,

其中,所述视频系统用于获得所述视频图像,

其中,对于所述N个预定形状每个预定形状,所述对应的预定形状与所述前景斑点的比较包括分析所述对应的预定形状与所述前景斑点集的重叠面积的量。

23. 根据权利要求22所述的方法,进一步包括通过回顾所述X个人物的位置中的至少一些位置来检测人群的存在。

24. 根据权利要求22所述的方法,进一步包括当确定所述X个人物中的Y个人物位于所述真实世界的水平面的区域内时确定人群的存在。

25. 根据权利要求24所述的方法,其中,所述真实世界的水平面的区域包括在所述真实世界内具有预定面积大小的预定几何形状。

26. 根据权利要求24所述的方法,其中,所述真实世界的水平面的区域包括由圆限定的区域。

27. 根据权利要求22所述的方法,进一步包括:

确定与所述视频图像的第一帧对应的第一区域内的第一人群密度;

确定与所述视频图像的第二帧对应的所述第一区域内的第二人群密度;

响应于所述第一人群密度和所述第二人群密度确定人群聚集事件。

28. 根据权利要求22所述的方法,进一步包括:

确定与所述视频图像的第一帧对应的第一区域内的第一人群密度;

确定与所述视频图像的第二帧对应的所述第一区域内的第二人群密度;

响应于所述第一人群密度和所述第二人群密度确定人群分散事件。

29. 一种视频监控系统,包括:

视频源,被配置为提供真实世界场景的视频图像;以及

计算机,被配置为检测所述视频图像的前景像素,所述前景像素的组构成一个或多个前景斑点的前景斑点集,

其中,对于所述视频图像中N个预定位置的对应的预定位置处的N个预定形状的每个预定形状,其中,N是大于1的正整数,所述计算机被配置为通过将所述对应的预定形状的重叠与所述前景斑点集进行比较来计算第一值,被配置成使用所述第一值来确定由所述前景斑点集代表的X个人物,其中,X是整数,被配置为对于所述N个预定形状的每个预定形状,通过分析所述对应的预定形状与所述前景斑点集的重叠面积的量来将所述对应的预定形状与所述前景斑点集进行比较,以及被配置为当人群密度超过阈值时,使用确定的所述X个人物的代表来提供报告、警报和事件检测的至少一种。

用于检测视频中的对象的方法、设备和系统

[0001] 本案是分案申请,其母案为于2013年9月12日提交的题为“用于检测视频中的对象的方法、设备和系统”的申请号为201380047668.5的申请。

[0002] 相关专利申请的交叉引用

[0003] 本申请要求于2013年3月15日提交的美国申请第13/838,511号以及于2012年9月12日提交的美国临时专利申请第61/700,033号的优先权,以上两个申请的每一个的内容通过引用方式全文并入本文中。

技术领域

[0004] 本公开涉及视频监控,例如,视频监控方法和系统以及视频查证方法和系统。公开了可以检测人物的视频监控系统、设备和方法。视频监控系统、设备和方法可以统计人数并且/或者监测视频流中的人群场景。

背景技术

[0005] 智能视频监控 (IVS) 系统可以用于(例如,通过回顾此前记录并存储的视频)实时或离线检测视频中感兴趣的事件。通常,该任务通过检测并跟踪感兴趣的目标来完成。当场景不拥挤时,这通常很好地发挥作用。然而,这种系统的性能在拥挤场景中会显著降低。实际上,这种拥挤的场景频繁发生,因此,对能检测人群中的人物具有极大兴趣。这种检测人物可以用于统计以及其他人群分析,例如,人群密度、人群形成和人群分散。

[0006] 此前的人群分析工作处理一些具体的极端拥挤的场景,例如某些体育运动或宗教活动。然而,还需要关注偶尔会形成大量人群的更普通的监控场景。这些场景包括公共场所,例如,街道、购物中心、机场、公共汽车站和火车站等。

[0007] 最近,人群密度评估或统计人群中的人物的问题越来越得到研究社群以及行业的高度关注。现有的方法主要包括基于映射(间接)的方法和/或基于检测(直接)的方法。

[0008] 基于映射的方法可以尝试将人物目标的数量映射到提取的图像特征,例如,运动像素的数量、前景斑点 (foreground blob) 的大小、前景边缘、前景角落的群以及其他图像特征。基于映射的方法通常要求针对不同类型的视频场景进行训练。研究主要集中在查找与人数良好对应的可靠特征以及如何处理一些特殊问题,例如,阴影和摄像机视图透视。在许多场景中,在提供足够训练视频的情况下,基于映射的方法可以提供相当精确的人数估计。然而,效果通常与场景有关系,并且每个个体的实际位置可能难以获得。

[0009] 基于检测的方法可以通过识别每个单独的人物目标来对场景中的人数进行计数。研究已经集中在人物检测、人体部位检测以及检测与跟踪的结合考量。这些方法可以在稍微拥挤的场景中提供更加精确的检测和计数。如果使得每个个体的位置可以获得,那么就有可能计算局部人群密度。这些方法的关键挑战是更高的计算成本、依赖视角的知识以及较大的人物图像大小要求。

[0010] 本发明所述的实施例解决了现有系统的这样一些问题。

发明内容

[0011] 所公开的实施例提供了用于对视频图像进行智能分析以检测对象(例如人物对象)的方法、设备和系统。

[0012] 在某些实施例中,一种检测视频中的人物对象的方法包括:确定作为前景像素的视频图像的某些像素,前景像素的组构成一个或多个前景斑点的前景斑点集;对于视频图像中N个位置的每个位置,其中N是整数(integer),将预定形状与前景斑点集进行比较以获得人物在该位置的相应概率,从而获得与N个位置对应的N个概率;并且使用N个概率,确定由前景斑点集代表的X个人物,其中X是整数(whole number)。

[0013] 一种检测视频中的人物对象的方法可以包括:将真实世界场景的视频图像的像素确定为前景像素,前景像素的组构成一个或多个前景斑点的前景斑点集;对于视频图像中N个位置的每个位置,其中N是整数,将预定形状与前景斑点集进行比较以确定由所述前景斑点集代表的X个人物,其中X是整数。

[0014] 方法可以包括确定X个人物的每个人物的位置。X个人物的每个人物的位置可以确定为真实世界的水平面上的位置,例如,真实世界的物理地平面上的位置。

[0015] 检测人物对象可以用于对人物进行计数,用于人群分析和其他事件检测。

[0016] 公开了可以被配置成执行这些方法的系统和设备。

[0017] 包括可用于配置计算机以执行本发明所述操作的软件的计算机可读介质进一步包括本发明的实施例。

附图说明

[0018] 结合附图,从以下详细描述可以更加清晰地理解示例性实施例。附图代表本发明描述的非限制性示例性实施例。

[0019] 图1图示了根据本发明的示例性实施例的示例性视频监控系统。

[0020] 图2图示了根据本发明的示例性实施例的示例性视频监控系统的视频流的示例性帧。

[0021] 图3A图示了根据本发明的示例性实施例的用于目标检测和计数的示例性流程图。

[0022] 图3B图示了几个人物模型占用二维视频图像的实例,每个人物模型相对于二维视频图像与不同的位置对应。

[0023] 图3C图示了一行(x,y)识别坐标321,每个识别坐标与对应的人物模型320相关联。

[0024] 图3D图示了用于计算人物概率映射(human probability map)的示例性方法。

[0025] 图3E图示了执行单次扫过(pass)概率映射的示例性方法,该方法是找到视频图像内人物模型的最佳数量的一部分。

[0026] 图3F图示了执行多次扫过概率映射的方法,该方法用于找到视频图像内人物模型的最佳数量。

[0027] 图4图示了一般人物模型,包括三维圆柱体模型及其对应的二维凸壳(convex hull)模型。

[0028] 图5图示了可以使用几个人物图像样本进行校正的一般平地摄像机模型。

[0029] 图6A、图6B和图6C示出了示例性检测结果。

[0030] 图7A、图7B和图7C图示了关于基于人物检测结果的人群密度的实例。

- [0031] 图8图示了用于检测各种人群相关事件的示例性实施方式。
- [0032] 图9图示了如何定义并检测拥挤区域的示例性方法。
- [0033] 图10图示了对每个检测的人物目标的示例性处理。
- [0034] 图11图示了对每个拥挤区域的示例性处理。
- [0035] 图12图示了可用于定义并检测人群“聚集”和“分散”事件的方法。
- [0036] 图13图示了定义人群聚集点的一个实例。
- [0037] 图14A和图14B示出了人群聚集点的一个实例。
- [0038] 图15图示了检测人群聚集点的示例性方法。
- [0039] 图16图示了更新人群聚集点并检测人群“聚集”和“分散”事件的示例性方法。
- [0040] 图17图示了使用多个摄像机的示例性实施方式。

具体实施方式

[0041] 以下将参照附图更全面地描述各个示例性实施例,附图图示了一些示例性实施例。然而,本发明能够以许多不同的方式来实施并且不应当被理解为限于本发明所述的示例性实施例。这些示例性实施例仅仅是实例,并且不要求本发明提供的细节的许多实施方式和变型是可行的。还应当强调的是,本公开提供替代实例的细节,但是这种替代方式的列举不是详尽的。此外,多个实例之间的任何细节一致不应当理解成要求这种细节,因为无法针对本发明描述的每种特征列举每种可行的变型。在确定本发明的要求时应当参照权利要求书的语言。在附图中,为了清晰起见,可以夸大层和区域的大小和相对大小。在整篇文中相同的附图标记指代相同的元件。

[0042] 应当理解,虽然术语“第一”、“第二”、“第三”等可以在本文中用于描述多种元件,但是这些元件应当不受这些术语的限制。这些术语用于使一个元件与另一个元件区分开。因此,在不脱离本发明构思的教导的情况下,以下讨论的第一元件可以称为第二元件。本发明中使用的术语“和/或”包括一个或多个相关的所列项目的任何和所有组合。

[0043] 应当理解,当元件被称为与另一个元件“连接上”或“耦接上”时,它可以与另一个元件直接连接上或耦接上,或者可以存在中间元件。反之,当元件被称为与另一个元件上“直接连接上”、“直接耦接上”时,不存在中间元件。用于描述元件之间的关系的其他词语应当用相同的方式进行理解(例如,“...与...之间”与“.直接在..与...之间”,“相邻”与“直接相邻”等)。

[0044] 本发明中使用的术语仅仅用于描述特定示例性实施例的目的,并且并非旨在限制本发明的发明构思。本发明中单数形式“一个”、“一种”和“所述”旨在还包括复数形式,除非上下文另有清晰的表示。还将进一步理解,当在本说明书中使用术语“包含”和/或“包括”指的是存在所述的特征、整数、步骤、操作、要素和/或元件,但是不排除存在或增加一个或多个其他的特征、整数、步骤、操作、要素、元件和/或它们的组合。

[0045] 除非另有说明,包括本发明中使用的技术术语和科技术语的所有术语具有与本发明构思所属的技术领域的普通技术人员通常理解的意思相同意思。还应当理解的是,术语,例如常用的字典中定义的术语,应当被理解为具有与相关领域上下文中的意思相一致的意思,并且不应当以理想化或过于正式意义的方式进行理解,除非本发明中特别作出了如此定义。

[0046] 定义:在描述本发明时,以下定义通篇适用(包括以上定义)。

[0047] “视频”可以指以模拟和/或数字形式呈现的动画。视频的实例可以包括:电视;电影;来自摄像机或其他观测器的图像序列;来自实时输入的图像序列;计算机生成的图像序列;来自计算机图像引擎的图像序列;来自例如计算机可读介质、数字化视频光盘(DVD)或高清晰度光盘(HDD)的存储设备的图像序列;来自基于IEEE 1394接口的图像序列;来自视频数字化仪的图像序列;或者来自网络的图像序列。

[0048] “视频序列”可以指一些或所有视频。

[0049] “摄像机”可以指用于视觉记录的设备。摄像机的实例可以包括以下的一个或多个:视频成像仪和透镜装置;摄像机;数字摄像机;彩色照相机;黑白照相机;照相机;摄影机;计算机摄像头;网络摄像头;红外线(IR)摄像机;弱光摄像机;热摄像机;闭路电视(CCTV)摄像机;平板/倾斜/可变焦(PTZ)摄像机;以及视频感测装置。摄像机可以定位成监控感兴趣的区域。

[0050] “视频处理”可以指任何操纵和/或分析视频,包括例如压缩、编辑、监控和/或查证。

[0051] “帧”可以指视频内的特定图像或其他离散单元。

[0052] “计算机”可以指能接收结构化输入、根据指定规则处理结构化输入并且产生处理结果作为输出的一个或多个装置和/或一个或多个系统。计算机的实例可以包括:计算机;固定计算机和/或便携式计算机;具有单个处理器、多个处理器或多核处理器的计算机,可以并行和/或不并行运行;通用计算机;超级计算机;大型计算机;超级小型计算机;小型计算机;工作站;微型计算机;服务器;客户端;互动电视;网络设备;具有互联网接口的电信设备;计算机和互动电视的混合组合;便携式计算机;台式个人计算机(PC);个人数字助理(PDA);便携式电话;用于模拟计算机和/或软件的专用硬件,例如,数字信号处理器(DSP)、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用指令集处理器(ASIP)、一个芯片、多个芯片或芯片组;片上系统(SoC),或微处理器片上系统(MPSoC);光学计算机;量子计算机;生物计算机;以及可以接收数据、可以根据一个或多个存储的软件程序处理数据、可以生成结果并且通常可以包括输入单元、输出单元、存储单元、算术单元、逻辑单元和控制单元的装置。

[0053] “软件”可以指用于操作计算机的规定规则。软件的实例可以包括:软件;代码段;指令;小程序;预编译代码;编译代码;解释型代码;计算机程序;和编程逻辑。

[0054] “计算机可读介质”可以指用于存储计算机可访问的数据的任何存储设备。计算机可读介质的实例可以包括:磁硬盘;软盘;光盘,例如,CD-ROM和DVD;磁带;快闪可移动存储器;存储芯片;和/或可以在上面存储机器可读指令的其他类型的介质。

[0055] “计算机系统”可以指具有一个或多个计算机的系统,其中每个计算机可以包括实施软件来使计算机运行的计算机可读介质。计算机系统的实例可以包括:用于通过由网络链接的计算机系统处理信息的分布式计算机系统;通过用于在计算机系统之间发送并且/或者接收信息的网络连接在一起的两个或更多个计算机系统;以及可以接收数据、可以根据一个或多个存储的软件程序处理数据、可以生成结果并且通常可以包括输入单元、输出单元、存储单元、算术单元、逻辑单元和控制单元的一个或多个装置和/或一个或多个系统。

[0056] “网络”可以指可以通过通信设施连接的多个计算机和相关设备。网络可以涉及永

久连接,例如,电缆或临时连接,比如通过电话线或其他通信线路实现的连接。网络可以进一步包括硬连线连接(例如,同轴电缆、双绞线、光纤、波导等)和/或无线连接(例如,射频波形、自由空间光波波形、声学波形等)。网络的实例可以包括:互联网,例如,因特网;内联网;局域网(LAN);广域网(WAN);以及例如因特网和内联网的网络的组合。示例性网络可以用多种协议的任意一种运作,例如,互联网协议(IP),异步传输模式(ATM),和/或同步光学网络(SONET),用户数据报协议(UDP),IEEE 802.x等。

[0057] 在一些实施例中,人群密度估计方法、系统和设备可以以现有的视频内容分析方法、系统和设备为基础。除基本估计精度要求之外,该方法还可以包括以下各项的一个或多个:

[0058] • 摄像机视图的独立性可以允许实施例在宽范围的应用场景中工作,而不顾摄像机位置、视角、目标像素数量等的变化。

[0059] • 可以实时运行的相对较低的计算成本。实施例可以在嵌入式系统上实施。

[0060] • 可以减少并且/或者消除复杂的初始化设置和训练,从而允许更方便以及更低的购置成本。

[0061] 本发明公开的一些实例包括基于检测的方法并且可以不要求训练。实例可以用通用IVS系统实施,该系统已经执行基本检测和跟踪任务并且提供可靠的前景掩模。可以对每个图像像素计算凸形区域人物图像模型,该模型可以用于估计每个前景区域中的人物目标数量。摄像机校正数据可以提供从图像平面到物理世界的地平面的映射,这可以用于在摄像机视图内的区域中提供实际人群密度测量。通过使用实际人群密度测量,可以检测到其他感兴趣的事件,例如,“人群热点”、“人群聚集”、“人群分散”等。

[0062] 图1图示了根据本发明的示例性实施例的视频监控系统101。视频监控系统101可以被配置成检测并监测视频流中的人群活动。视频监控系统101可以在对人物检测感兴趣的各种各样的应用中使用,例如,用于人群密度分析。例如,实施例可以用于可疑人物聚集检测、行人交通统计收集、异常人群形成和/或分散等。视频监控系统101可以包括视频源102(例如,具有存储视频的摄像机或存储器,例如硬盘驱动)、变化检测模块103、运动检测模块104、前景斑点检测模块105、人物检测模块106、目标跟踪模块107和事件检测模块108。在此实例中,视频源(例如,摄像机)是固定的。然而,普通技术人员会认识到本发明也适用于移动视频源。在此实例中,视频源提供单个视频流。然而,本发明还设想出使用并处理多个视频流。

[0063] 视频监控系统可以用典型的固定平台IVS系统来实施。举例来说,参见授予Venetianer等人的美国专利第7,868,912号以及授予Lipton等人的美国专利第7,932,923号,这两份专利通过引用的方式全部并入本发明中,详见可用于实施本发明所述的实施例的IVS系统的示例性细节。美国专利第7,868,912号和美国专利第7,932,923号也通过引用方式并入用于使用生成的视频基元从视频获得信息的视频基元(或元数据)生成和下游处理(可以是实时处理或随后处理)的示例性细节,例如事件检测,这可以与本发明公开的实施例一起使用。每个模块103-108以及它们的单独元件单独地或者与其他模块/元件结合可以通过专用硬件(电路)、软件和/或固件来实施。例如,用软件编程的通用计算机可以实施所有的模块。这样,包括可用于配置计算机以执行本发明所述操作的软件的计算机可读介质进一步包括本发明的实施例。作为另一个实例,为了实施本发明所述的系统、设备和方

法,可以使用各种计算和光学元件,例如以下各项的一个或多个:通用计算机;超级计算机;大型计算机;超级小型计算机;小型计算机;工作站;微型计算机;服务器;互动电视;计算机和互动电视的混合组合;智能手机;平板电脑;以及仿真计算机和/或软件的专用硬件。这些可以包括一个或多个处理器;一个或多个现场可编程门阵列(FPGA);计算机存储器;计算机可读介质,例如,用于存储计算机可访问的数据的任何存储设备(例如,处理器可以对从摄像机接收的数据进行各种运算,并且计算机存储器随后可以存储有关各种像素的信息并且可以存储斑点检测、目标检测和事件检测的结果)。计算机可读介质的实例包括:磁硬盘;软盘;光盘,例如,CD-ROM和DVD;磁带;存储芯片;固态存储设备;以及用于携带计算机可读电子数据的载波,例如,在发送和接收电子邮件或者在访问网络时使用的载波。有形计算机可读介质包括在物理上有形的计算机可读介质,例如以上列举的实例。此外,软件可以结合计算元件和/或光学元件使用以实施本发明所述的方法。软件可以包括使计算机操作的规则和/或算法,并且可以包括例如代码段、指令、计算机程序和编程逻辑。视频源102和模块103-108可以在单个系统内或者可以分散。例如,视频源102可以包括在待监测区域的摄像机。视频源102给模块103至107所处的监测位置(例如,在待监测位置之外的单独的第三位置)提供视频流。事件检测模块108可以设置在与监测位置和第三位置分开的第三位置(例如,中心站)。本发明所述的各种模块、计算机、摄像机和其他图像设备可以通过网络连接,网络可以涉及永久连接,例如,电缆或临时连接,比如通过电话线或其他通信线路实现的连接,并且还可以包括无线通信线路。网络的实例包括:互联网,例如,因特网;内联网;局域网(LAN);广域网(WAN);以及例如因特网和内联网的网络的组合。上述各种硬件和软件的实例同样在通过引用方式并入本发明中的专利文献中有更详细地描述。

[0064] 变化像素可以通过变化检测模块103检测到,因为视频源102提供的视频图像的像素不同于此前获得的背景图像。背景图像可以是动态的。可以从输入的视频帧连续构建并更新动态背景图像模型。因此,在背景图像中可以考虑到改变视频图像的亮度、气候等的变化。在104中,帧差异可以用于检测运动像素。在105中,考虑来自模块103的变化像素和来自模块104的运动像素之一或两者以确定在空间上集成前景斑点的前景像素。视频图像可以通过现有的视频内容分析系统和方法来处理以提取前景、前景斑点和感兴趣的前景斑点(例如,人物前景斑点),例如详见于2010年11月2日公布的授予Zhang等人的美国专利第7,825,954号,该专利的内容通过引用的方式全部并入本发明中。深度传感器信息可以任选地用于估计作为潜在的人类检测的每个对象的真实世界的高度或大小,并且因此,可以更加准确地识别与潜在的人物目标对应的斑点(与不感兴趣的斑点形成对比)。深度传感器信息可以任选地用于估计阴影、高光、作为感兴趣的区域之外检测的对象、太远的对象(例如,可能靠得不够近,以至于不允许进行精确分析)或者会增加视频图像错误分析的风险的视频图像的其他要素。使用深度信息的示例性细节可以详见于授予Zhang等人的美国专利申请序列号13/744,254,该申请的内容通过引用的方式全部并入本发明中。随时跟踪斑点以在目标跟踪模块107中形成时空目标,并且最后,事件检测模块108使用目标检测和跟踪处理的输出来检测使用者定义的感兴趣的事件。并非或者除了将前景像素简单地在空间上集成斑点之外,人物检测模块106使用校正信息和凸形区域形状的人物模型以便甚至在拥挤的场景中检测人物。在一些实例中,不需要或者需要最低限度的提前训练来检测场景中的人物对象。并且在事件检测模块108中,可以实施一些新的事件检测方法,这些方法可以使

用人物检测模块106中的人物检测结果。

[0065] 图2图示了与IVS系统101的一些典型应用场景(包括户外广场、街道、旅游景点、火车站、大型购物中心、地铁站等)对应的视频图像。可以看出,根据相对于被拍摄的场景的摄像机的位置,占据视频图像的人物的相对大小和形状并不相同。

[0066] 图3A图示了提供视频监控系统101的更多示例性细节的框图。前景斑点检测模块105可以与图1的模块相同。模块301、302、303、304、305和306可以是图1的人物检测模块106的要素。人体像素检测模块301根据来自变化检测模块103的变化像素结果检测人体像素。这些像素或明显不同于背景图像模型(例如,亮度差和/或色差超过各自的阈值),或位于高度确信的前景边缘像素之间。它们最可能被认为是图像中合理的人体像素。例如参见图6A的301a,该图图示了检测的人体像素的实例。可以从进一步的人物检测处理排除其他变化像素,因为它们最可能代表阴影或反射。人体边界像素检测模块302检测前景斑点的边界与当前视频帧的图像边界对齐的人体边界像素。例如参见图6A的302a,该图图示了检测的人物边界像素的实例。当进行人物检测时,可以实施其他分析(除了上述分析之外或者用于代替上述分析)来帮助确定已经检测到人体。例如,可能要求各潜在人物斑点必须包含某一数量的边界前景边缘像素。作为另一个实例,其他处理可以识别斑点为很可能与除人物之外的对象(例如,车辆)相关,并且排除这些斑点进行进一步的人物检测处理。可以从前景斑点集排除不认为是潜在人物的其他前景斑点。可替代地,任何检测的斑点可以是前景斑点集的一部分。

[0067] 一般人物模型模块303提供一般的人物三维和二维模型。例如,一般人物模型模块303可以通过将真实世界的三维人物模型映射或投射到视频图像的二维图像平面,从而将三维人物模型转换成二维人物模型。图4图示了映射到图像平面330上对应的二维人物模型303b的示例性三维模型303a。三维人物模型303a可以是一组简单的三维形状,例如一组圆柱体(例如,一个圆柱体代表腿,一个圆柱体代表躯干,一个圆柱体代表头)。同一个三维人物模型303a(例如,圆柱体模型)可以与各种摄像机位置一起使用使得摄像机相对于地面(真实世界的地平面)的不同角度可以用于在摄像机的图像平面中获得不同形状的二维人体模型303b。例如,以三维圆柱体人物模型为例,提供特定位置的俯视图的摄像机视角可以在二维图像平面中映射成圆,然而具有同一位置的倾斜视图的摄像机视角可以将三维圆柱体人物模型映射成具有细长形式的不同形状。在图17所示的实例中,摄像机1702可以比摄像机1704具有更多俯视的三维人物模型303a,而后者可以比前者具有更多侧视的三维人物模型303a。如果摄像机1702和1704到三维人物模型303a的距离相同,那么映射到摄像机1702的图像平面上对应的二维人物模型可以具有比映射到摄像机1704的图像平面上的二维人物模型更紧凑(例如,更矮)。二维人物模型可以具有三维人物模型投射到二维图像平面的外边缘的内插点而获得的凸形形状。

[0068] 图4图示了一般人物模型,包括三维圆柱体模型303a以及映射到二维图像平面330的对应的二维凸壳模型303b。三维人物模型303a由腿圆柱体、躯干圆柱体和头部圆柱体组成。各圆柱体的长度和半径可以与代表典型的普通人的典型尺寸的物理统计数据对应。如图4所示,这三个圆柱体具有四个关键平面:头部平面、肩部平面、臀部平面和脚平面。为了获得在特定位置的对应的二维人物模型,我们可以沿着四个关键平面的周界统一取样并且通过使用摄像机校正参数将各三维样本点投射到二维图像平面,以确定相对于二维图像空

间内的特定位置的合适大小和取向。这些对应的图像样本点然后可以用于通过凸形形成方法在图像上形成凸壳,该凸壳可以用作二维图像人物模型。

[0069] 图5图示了可以使用几个人物图像样本校正的一般平地摄像机模型。摄像机模型可以仅包括三个参数:摄像机相对于地面的高度、摄像机的倾转角以及摄像机的焦距。通过使用来自视频帧的三个或更多个人物样本可以估计这些参数,详见于“A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration”(第八届国际视频监控研讨会,2008,Z.Zhang,P.L.Venetianer和A.J.Lipton)以及于2010年9月21日公布的授予Zhang等人的美国专利第7,801,330号,上述内容通过引用的方式全部并入本发明中。

[0070] 在替代形式中,或者此外,一般人物模型模块303可以具有能够响应于拍摄视频图像的摄像机的相机角度进行修改(例如,伸长、收缩、相对于二维图像平面的垂直轴倾斜等)的预定的二维模型。通过一般人物模型模块303可以提供几个一般人物模型。人物模型还可以包括用于典型配件的建模。例如,当在户外使用系统时,第一个人物模型可以在温暖的天气使用,第二个更大的人物模型可以在寒冷的天气使用(此时可以预料到穿上了外套并外套是人物模型的一部分),并且第三个人物模型可以在下雨的天气使用(此时可以预料到使用了雨伞并且雨伞是人物模型的一部分)。

[0071] 一般人物模型模块303还提供估计在图像空间内对应位置的各种大小的二维人物模型。图像空间可以与视频源102提供的视频帧中图像的二维空间对应。图像空间可以通过像素增量来测量,使得图像空间内的位置通过像素坐标来识别。摄像机可以拍摄视频图像,包括三维真实世界的二维图像。当人物出现在真实世界的某一位置时,可以预料到人物在二维视频图像内的某一位置处占用一定量的前景。如果人物远离摄像机,可以预料到人物的图像大小会比靠近摄像机的人物的图像大小更小。对于二维视频图像空间内的各个位置,一般人物模型模块303可以提供具有与二维图像空间内的位置对应的大小的人物模型。对于每个位置,二维人物模型可以具有响应于二维视频图像的图像空间内的各位置的尺寸和/或大小。这些人物模型的取向也可以响应于二维图像空间内的位置。例如,一些摄像机镜头(例如,广角镜头)可以代表真实世界中的垂直方向,视频图像帧的一侧为第一方向,视频图像帧的第二侧为不同的第二方向。二维人物模型响应于真实世界的垂直方向的不同表示可以在视频图像帧的不同侧(和其他位置)具有不同的取向。

[0072] 二维视频图像空间内各个人物模型的位置可以与二维视频图像空间内的识别坐标关联。识别坐标可以与具有二维视频图像空间的视频的像素位置对应。例如,与像素阵列第10行第22列对应的位置可以与识别坐标(10,22)对应。对于二维视频图像空间内的各个位置,一般人物模型模块303可以将人物模型的特定映射到相关的识别坐标。例如,人物模型的特定映射点可以是与人物头部对应的人物模型的顶部、与人物的脚对应的人物模型的底部、与人物的中心对应的人物模型形状的形心。人物模型的剩余部分可以基于人物模型的特定映射点与人物模型的剩余部分之间的固定关系相对于人物模型的相关识别坐标和大小映射到二维视频图像空间。举例来说,假设人物模型是一个圆。对于二维视频图像空间内的每个像素,对应圆的中心经过映射(例如,与二维视频图像空间的(x,y)坐标相关),此时考虑到圆形的对应大小(以及圆与其中心的已知关系),圆形形状的剩余部分映射到二维视频图像空间。三维真实世界中人物的特定部分(例如,人物头部的顶部、人物脚的底部、人物的中

心)的位置可以与它在二维视频图像中的位置具有唯一对应关系,并且因此,在二维视频图像内存在人物的这个特定点可以用于确定人物在三维真实世界的位置。

[0073] 一般人物模型模块303也可以针对二维图像空间内的每个识别位置确定人物模型的大小。也可以从视频监控系统101的校正获得人物模型的大小。例如,在视频监控系统101为了校正目的而拍摄视频的同时,已知大小的校正模型可以在待检测区域周围运动。校正模型可以是检测区域周围走动的已知身高的人。在校正期间,系统可以识别视频中的校正模型为前景斑点并且(例如,通过访问提供给视频监控系统101的有关校正模型的大小的校正信息)辨认出前景斑点与预定大小对应(例如,预定高度)。这里,随着校正模型在视频校正期间移动经过待监控的区域,对于视频图像内的各个位置,系统可以将校正模型的已知高度与二维视频图像中的大小关联上。例如,当校正模型的中心在位置 (x_1, y_1) 时,校正模型的高度可以是15个像素(或者可以通过一些其他的测量方式进行测量)。当校正模型的中心在位置 (x_2, y_2) 时,校正模型的高度可以是27个像素。因此,视频监控系统101可以通过将二维视频图像大小与校正模型的已知大小(例如,高度)关联上而将二维视频图像中特定位置(例如, (x, y) 坐标)的二维视频图像的尺寸与真实世界中的大小(例如,高度)关联上。基于真实世界的大小与二维图像中特定位置(例如, (x, y) 坐标)的二维视频图像中的尺寸之间的(通过这种校正获得的)已知相互关系,可以针对二维视频图像内的各个位置((x, y) 坐标)计算二维视频图像空间内的人物模型的二维大小以对应于真实的三维世界中的平均人物大小。

[0074] 对于校正过程的实例,参见授予Lipton等人的美国专利第7,932,923号以及授予Zhang等人的美国专利第7,801,330号,这两份专利通过引用的方式全部并入本发明中。一般而言,使用通过校正过程输入或获得的参数,例如摄像机高度(H)、垂直和水平摄像机视野角度(θ_H, θ_V)以及摄像机倾角(α)和其他信息,例如对象的检测到的外边界(例如,人物的顶部和底部),摄像机系统为了识别目的一般可以确定目标在真实世界中的大小和形状。

[0075] 基于人物的摄像机校正模型304可以连同视频图像空间内合适的对应位置一起从一般人物模型模块303接收并存储具有合适大小的人物模型。这些人物模型和对应的位置可以存储在查找表中。例如,视频图像空间内外的各个坐标 (x, y) 可以用于识别对应的人物模型。例如,当 (x, y) 识别坐标与人物模型的形心对应时,在估计以位置 (x_1, y_1) 为中心的二维视频图像内存在人物对象时,基于人物的摄像机校正模型304的查找表可以接收位置 (x_1, y_1) 为输入,并且提供对应的人物模型(包括它在二维图像空间内的大小和位置)。例如,输出可以包括二维图像空间内的边界,或者可以包括图像空间内的整个像素集(例如,所有像素的 (x, y) 坐标)以描述对应的人物模型。

[0076] 图3B图示了几个人物模型占用二维视频图像的实例,每个人相对于二维视频图像与不同的位置对应。如图所示,四个人物模型320a、320b、320c和320d与相对于二维视频图像的不同的 (x, y) 识别坐标相关。人物模型320a是最小的,与三维真实世界中距离视频源最远的位置对应。人物模型320b、320c和320d与依次越来越靠近视频源的三维真实世界中的位置对应。人物模型320a、320b、320c和320d可以全部是从相同的完整人物形状模型得到的。然而,可以估计到,仅一部分完整人物形状模型可以在某些位置占据二维视频图像。这里,估计到与人物形状320c和320d对应的完整人物形状模型仅部分占据二维视频图像空间330;估计到人物模型320c作为完整人物形状模型的躯干和头部的组合,此时人物模型320d

仅与完整人物形状模型的头部对应。

[0077] 各人物模型320a、320b、320c和320d与相对于二维视频图像的 (x, y) 识别坐标相关。在此实例中,人物模型320a、320b和320c的识别坐标与人物模型的形心对应。与估计的形状320a、320b和320c相关的 (x, y) 识别坐标分别是321a、321b和321c,并且落入视频图像的 (x, y) 坐标内。与估计的形状320d相关的 (x, y) 识别坐标落入视频图像的 (x, y) 坐标外。也就是说,与320d相关的人物形状模型的形心位于视频图像下方,并且因此其识别 (x, y) 坐标具有负 y 轴值,这在此实例中位于视频图像的坐标外(并且图3B未示出)。为了容易计算, (x, y) 识别坐标可以以像素为单位增大,使得识别坐标321a、321b和321c也识别视频图像的像素。

[0078] 为了便于说明,图3B仅图示了与四个对应的识别坐标相关的四个人物模型。然而,基于人物的摄像机校正模型304可以存储人物模型的大量 (x, y) 识别坐标,几个人物模型可以彼此重叠。图3C图示了均与对应的人物模型320相关的单行 (x, y) 识别坐标321。为了便于说明,仅图示了单行,但是可以给多行 (x, y) 识别坐标提供人物模型,这些人物模型可以规则地分布在图像空间330中的 x 和 y 方向。按照所讨论的,形状的大小可以因位置的不同而不同(尽管它们在图3C中被图示为具有相同大小)例如,基于人物的摄像机校正模型304可以将二维图像空间330中每个像素的人物形状存储为二维图像空间330的 (x, y) 识别坐标以及与至少部分地在二维图像空间330内的人物模型相关的二维图像空间330外的 (x, y) 坐标。例如,对于视频图像空间330内的所有 (x, y) 像素坐标,当人物模型的形心在视频图像的视频图像空间330内位于该 (x, y) 识别坐标处时,基于人物的摄像机校正模型304可以存储预料到人物将会占据的视频图像空间330内的子空间的 (x, y) 识别坐标和相关人物模型(可以包括边界或像素集)。 (x, y) 识别坐标也可以包括与视频图像空间330内的子空间中的人物模型相关的视频图像空间330外的所有 (x, y) 识别坐标(也就是说,完整人物模型的一部分可以位于视频图像空间330的子空间内)。对于一些情形,上述子空间可以包括整个视频图像空间330(与估计何时人物的位置完全占据视频图像对应)。基于人物的摄像机校正模型304可以存储 (x, y) 识别坐标和相关人物模型为查找表。在完整人物形状模型的形心与本实例中人物模型的 (x, y) 识别坐标对应时,可以使用人物形状模型的其他识别点(例如,眼睛、鼻子、头部中心、头部顶部、脚趾、脚底等)。

[0079] 人物概率映射计算模块305使用前景斑点检测模块105输出的视频图像的特定帧的前景斑点集以及人物模型(其对应的识别坐标由基于人物的摄像机校正模型304输出),以针对二维视频图像内的各个位置(例如各图像像素位置)计算人物目标概率。多个计算的概率可以与多个位置相关以构建概率映射。多个位置可以与人物模型的 (x, y) 识别坐标相同。

[0080] 对于各 (x, y) 识别坐标,进行计算以确定视频图像中存在人物模型的相应概率。当 (x, y) 识别坐标与视频图像的像素具有一一对应关系时,针对视频图像的各像素进行概率计算。例如,对于每个图像像素,对应的人物概率可以计算为存在人物目标(其图像中心处于所考虑的像素中)的可能性。可以创建概率映射,从而将每个概率计算值映射到每个 (x, y) 识别坐标。查找表中可以存储概率映射,从而使各 (x, y) 坐标(作为输入)与相关的计算概率关联上。查找表可以与基于人物的摄像机校正模型模块304的查找表(存储人物模型作为条目)相同,或者可以是单独的第二查找表。

[0081] 如上所述,识别坐标可以落入视频图像空间外,并且因此可以进行计算以确定在视频图像内存在人物目标的相应概率(关于落入与这些识别坐标相关的图像空间(人物模型)内的部分对应完整人物二维模型)。例如,如果二维完整人物模型的形心与识别坐标对应,那么该形心可能在视频图像空间外,但是可能与作为完整人物模型的一部分的视频图像空间内的二维人物模型对应。例如,完整人物模型的肩部和头部可以构成二维人物模型(肩部和头部落入图像空间内),即使此完整人物模型的形心(例如,完整人物模型的肚脐附近)落入图像空间外(与用于识别对应的肩部/头部二维人物模型的识别坐标对应的形心)。在一些实例中,某一百分比的完整人物二维模型必须落入将要(或者考虑)进行概率计算的图像空间内。例如,当低于10%或低于20%的完整人物二维模型在图像空间内时(或者,当人物模型低于完整人物二维模型的10%或20%时),与识别坐标相关的概率值可以设置为零或被忽略。在一些实例中,当不到40%的完整人物二维模型在图像空间内时,与识别坐标相关的概率值可以设置为零。

[0082] 各 (x,y) 识别坐标的概率计算可以调用与对应的 (x,y) 识别坐标和前景斑点集相关的人物模型。例如,各 (x,y) 识别坐标的概率计算可以调用与对应的 (x,y) 识别坐标相关的人物模型内的人体像素和人物边界像素。与对应的 (x,y) 识别坐标相关的人物模型可以从(例如,存储在模块304的查找表中)的基于人物的摄像机校正模型模块304输出。前景斑点集可以从前景斑点检测模块105输出。用前景斑点集调用的估计形状可以计算为与前景斑点集重叠的人物模型区域与人物模型区域的比值。可以忽略不超过某一阈值的概率计算值。例如,计算的概率小于0.4(在0至1的量级)可以表示没有人物目标以该位置为中心。可以进行除调用计算之外的计算以确定在与各个估计的形状对应的视频图像中存在人物对象的概率。应当理解,计算的概率是估计值。因此,计算的概率为1(在0至1的量级)并不表示绝对确定在相关的对应位置存在人物。

[0083] 图3D图示了用于计算人物概率映射的示例性方法,该方法可以由图3A的系统实施。在步骤S340中,在304中校正的摄像机模型可以用于将二维图像空间的图像平面映射到真实世界的地面上。在步骤S342中,对二维图像空间中的N个位置可以获得人物模型(N是等于或大于2的整数)。校正的摄像机模型304可以用于获得对应的凸壳形状的人物模型,作为二维图像空间中每个图像像素位置的人物模型。各人物模型可以与二维图像空间中的识别坐标相关。例如,当执行映射到识别坐标时,人物模型的人物形心点可以用作参考点。假设二维图像空间的识别坐标是图像空间中人物的形心,通过校正的摄像机模型可以计算真实世界的地面上人物的对应物理脚印位置(例如,如图5所示)。然后将一般的三维(例如,多个圆柱体)人物模型放置在该脚印位置。三维模型的大小可以与此前获得的校正数据对应。一般的三维人物模型可以投射或映射到二维图像平面上以获得二维图像空间中的人物模型。例如,投射三维多圆柱体人物模型可以用于形成对应的二维图像凸壳,因为图像人物模型的形心在相关的识别坐标(例如,所考虑的图像点)。这样,各有效的图像像素可以具有表示在该图像位置合适的人物大小和形状的对应的凸形区域形状的人物模型(作为人物模型)。为了降低计算成本,在初始化系统时可以提前计算凸形区域形状的人物模型,人物凸形模型的矩形边界框通过使用积分图像可以用于获得近似的人物调用率。在步骤S344中,可以从视频图像提取前景斑点集。前景斑点集可以包括使用由模块301提取的人物前景像素和/或由模块302提取的人物边界像素所检测到的一个或多个前景斑点。在步骤S346中,对于各

N个位置,计算在各个位置存在人物的概率以获得概率映射。假设在图像人物凸形模型中存在足够多的人物边界像素,人物概率测量可以限定为人物调用率。该实例中的人物调用率是图像人物凸形模型中在301计算的人物前景像素的数量与该人物凸形模型的总面积的比值。可以按照与图示顺序不同的顺序执行图3D的处理的步骤顺序。例如,可以在步骤340和342之一或两者之前执行步骤344。

[0084] 参见图3A,基于在305计算的人物概率映射,人物目标估计模块306可以找到视频图像中最佳数量的人物模型(例如,人物对象)及其位置。全局优化方法可以用于找到最佳数量的人物模型及其位置。如果 $m(m_1, \dots, m_M)$ 代表图像空间内所有潜在人物模型中的M组人物模型,那么目的是找到最佳组 n^* ,使得临界函数 $f(n^*)$ 达到全局最大值。也就是说,目的是找到:

$$[0085] \quad \underset{n \in m}{\operatorname{argmax}} f(n)$$

[0086] 其中, n 是图像空间中多个人物模型的特定组, $f(n)$ 是计算用于这个人物模型组的函数。

[0087] 按照以下进一步所讨论的,计算几个选择的人物模型组的每一组的函数 $f(n)$,每组从概率映射选择 m_i 个位置(每次扫过选择 m_i 个位置,其中这些扫过次数的每一次扫过的数量 m_i 可以不同)。扫过(或扫描)概率映射可以选择每组人物模型,而某些约束准则用于选择每次扫过均变化的位置。这里,函数 $f(n)$ 的定义为:

$$[0088] \quad f(n) = w_R * R(n) + w_P * P(n) - w_0 * O(n)$$

[0089] 其中, R 是人物调用率,其定义为人物前景面积与 n 个选择的人物模型的组的整个面积的百分比; P 是人物精度,是与 n 个选择的人物模型的组重叠的前景面积的百分比,并且 O 是人物重叠比,是 n 个选择的人物模型的任意人物模型彼此重叠的面积与全部 n 个选择的人物模型占据的面积之比,并且 w_R 、 w_P 和 w_0 是权重。有利的是,在没有太多人物重叠的情况下找到前景区域(前景斑点集)与人物模型的联合(m 个人物模型的组)之间的最佳匹配。在实施过程中,如何确定上述三个权重会显著影响检测结果,例如,如果增加更多权重来减小人物重叠率,这会得到更小的人物计数。

[0090] 通过参照人物概率映射计算模块305输出的概率映射可以选择 m_i 个选择的人物模型的每个人物模型。可以进行几次扫过以执行计算 $f(n)$,每次扫过从一般人物模型模块303提供的二维人物模型选择 m_i 个人物模型的子集并且与基于人物的摄像机校正模型304中(例如,查找表中的) (x, y) 识别坐标相关。如前所述,这些扫过的每次扫过的 m_i 值可以不同。每次扫过的人物模型选择准则可以不同,使得对不同扫过选择不同的人物模型(并且有可能的是,对不同扫过选择不同数量的 m_i 个人物模型)。选择准则可以包括要求选择的人物模型与概率映射阐述的概率阈值 P_{th} 相关。选择准则还可以包括下一个选择的二维人物模型与任何此前选择的二维人物模型的距离是最小距离 D_{min} 。最小距离 D_{min} 可以是在真实世界的地面上的距离。例如,二维人物模型的形心可以映射或转换到三维真实世界中的位置,并且可以计算它们之间的距离。可以在二维图像平面内计算最小距离 D_{min} ,但是二维图像平面内的距离可以反映对应的三维位置,使得对于视频图像源附近的人物模型,可以比更远的人物模型在二维图像平面中要求更大的间隔。

[0091] 在一些示例性实施例中,概率映射的一个或多个快速一次扫描用于确定人物计数

和对应位置。图3E图示了执行单次扫过概率映射的方法,该方法是找到视频图像内人物模型的最佳数量的一部分。图3E的方法可以由人物目标估计模块306实施。在步骤S350中,扫描概率映射以找到局部最大值(这可以通过某一选择准则来判断是否合格)。可以扫描输入概率映射以定位与真实世界上最靠近视频源的位置对应的可用的未选择的局部最大值。概率映射的底部可以与视频图像的底部对应。在许多实施方式中,执行监控功能的摄像机可以安装在比待监测的区域内的人的头部水平高的位置。因此,视频图像的底部可以与最靠近视频源的位置对应。在此实例中自底向上扫描概率映射允许选择不太可能与视频图像内的闭塞对象(occluded object)对应的人物模型。

[0092] 自底向上扫描概率映射可以找到局部最大值点,从而代表用于图像空间内的多个位置的每个位置的此前计算的概率(存储在概率映射中)的局部最大值。局部最大值可以是具有比每个紧邻的 (x, y) 识别坐标(例如,紧邻的像素)的概率值更高的概率值的 (x, y) 识别坐标(例如,像素)。在步骤S352中,一旦找到局部最大值点,就将与作为其识别坐标的该局部最大值相关的人物模型选择为 m_i 个人物模型的组中的一个。在步骤S354中,在这次扫过中排除该选择的模型感兴趣的区域内(例如,落入二维人物模型的边界)的所有像素以及与距离该选择的模型最小距离 D_{min} 对应的像素(例如,在真实世界的地面上代表最小距离的视频图像中的像素)不作进一步考虑(并且可以从用于这次扫过的概率映射临时去除)。要注意,在此实例中,与人物模型的识别坐标对应的像素及其描述同样适用于不是像素位置的识别坐标。在一些实例中,在这个阶段不需要进一步分析视频图像本身,并且仅仅通过从概率映射临时去除像素可以排除该像素而不作进一步考虑。再次扫描概率映射以选择比概率阈值 P_a 更大且未被排除的与像素相关的人物概率映射的概率的另一个局部最大值点。在步骤S356中,确定是否已经考虑任何有效的像素。也就是说,在这次扫过概率映射中,可以对概率值既未被选择准则排除也未被选择其他人物模型排除的概率进行回顾。持续扫描概率映射,直到考虑并从概率映射去除所有有效的像素。因此,通过这次扫过概率映射可以选择 m_i 个人物模型。对于此次扫过,对这组 m_i 个人物模型计算函数 $f(m_i)$ 。

[0093] 可以对概率映射执行额外的扫描,而每次通过扫描(on-pass scan)具有不同组的选择准则。图3F图示了执行多次扫过概率映射的方法,该方法用于找到视频图像内人物模型的最佳数量。图3F的方法可以由人物目标估计模块306来实施。这里,对于每次扫描, D_{min} (最小距离)和 P_{th} (概率阈值)的至少一个的值可以不同。在步骤S360中,对特定的通过扫描设置选择准则。在考虑到所需精度和计算开销的情况下,按照具体情况具体分析的原则可以确定多少次选择准则的变更(并且因此多少次扫描)。在步骤S362中,对概率映射进行扫描以根据选择准则选择 m 个人物模型的组。值 m 是等于或大于零的整数,并且对于每次选择(例如,对于执行步骤S362的图3F的每次循环)都可以不同。步骤S362可以与图3E的方法对应。在步骤S364中,对选择的 m_i 个人物模型计算准则函数,例如,对这次扫描中选择的 m_i 个人物模型计算对应的 $f(m_i)$ 。可以用新的选择准则执行额外的扫描(S366)。当完成概率映射的所有扫描时,确定扫描的组的 $f(n)$, $n \in \{m_1, \dots, m_m\}$ 的最大值。确定与该最大值对应的人物模型的组以便与视频图像内的人物对象对应(S368)。通过使用确定用于代表视频图像中的人物对象的人物模型的人物的 (x, y) 识别坐标(例如,像素位置),可以确定真实世界地面上的位置。

[0094] 在替代实施例中,如果 m 代表图像空间内所有潜在人物模型中的人物模型组,那么目的可以是找到最佳组 m^* ,使得临界函数 $g(m^*)$ 达到全局最大值。也就是说,目的是找到 g

(m) 的最大值:

$$[0095] \quad g(m) = \sum_{n=1}^m f(n)$$

[0096] 其中n是图像空间中的多个人物模型的特定一个人物模型,m是选择的人物模型的数量(对于不同的求和计算可以不同),并且f(n)是计算用于m个人物模型的每个人物模型的函数,而不是人物模型组的函数。

[0097] 这里,函数f(n)的定义为:

$$[0098] \quad f(n) = w_R * R(n) + w_P * P(n) - w_0 * O(n)$$

[0099] 其中,R是人物调用率,其定义为人物前景面积与选择的人物模型的整个面积的百分比;P是人物精度,是与选择的人物模型重叠的前景面积的百分比,并且O是人物重叠比,是选择的第n个人物模型与第1个至第n-1个人物模型占据的面积(计算 $\sum f(n)$ 时,当前扫过之前所选择的人物模型占据的面积)的重叠率,并且 w_R 、 w_P 和 w_0 是权重。每次通过扫描如上所述的概率映射可以与计算 $\sum f(n)$ 相关,在选择每次扫过的概率映射的局部最大值时对选择准则由不同的约束。可以使用除本文描述的这些函数之外的其他函数f(n)。

[0100] 图6A、图6B和图6C图示了根据一个实例的视频监控系统101的检测结果。对于一个输入帧,图6A是人体检测模块301和人物边界像素检测模块302的输出,其中像素301a表示检测的人体像素,并且像素302a表示人物边界像素。在图6A中前景斑点集表示为检测的人体像素301a与人物边界像素302a的组合。检测的人体像素和人物边界像素叠加在限定视频图像空间330的原始视频图像帧上。在此实例中,此视频图像帧中的视频图像的剩余部分(除了前景斑点集之外)是背景图像的一部分。

[0101] 图6B图示了从图6A计算的人物概率映射。在此实例中,人物概率映射代表用灰度等级表示的计算的概率,黑色与零(0)概率对应,白色与一(1)概率对应。在对应于与对应的人物模型的识别坐标对应的像素的图像空间330内的位置表示每个计算的概率。

[0102] 图6C示出了最终的人物检测结果,图示了与检测的人物对应的多个人物模型320(粉色凸形轮廓)。这些人物模型的每个人物模型可以通过能识别三维真实世界中检测的人物的位置的识别坐标(例如,形心)相关联,并且映射到真实世界的地面(未图示)。

[0103] 图7A、图7B和图7C图示了基于人物检测结果测量人群密度的实例。图7A图示了视频监控系统101检测结果的示例性结果,示出了多个二维人物模型320(粉色凸壳),每个人物模型与检测的人物对应,重叠在原始视频图像上。图7B图示了将检测的人物映射到真实世界的物理地平面,示出了自顶向下表示图7A的视频图像,图中每个圆圈代表如同映射到真实世界的物理地平面的物理地平面的人物模型320,因此识别真实世界中检测的人物的位置。检测的人物目标可以映射到物理地平面上,因为对校正模型的已知大小、二维图像中的位置与图像空间中的对应大小之间的对应关系已经进行校正。根据已知的位置,可以进行计算以对某一识别区域(例如,由使用者选择)或整个场景内的人数进行计数。也可以进行计算来确定单位面积的人数。也可以直接计算每个地面位置的真实人群密度测量。人群密度测量的实际定义可以根据实际应用,特别是,根据待监测的人群的大小。例如,对于图6A、图6B和图6C所示的场景,可以使用2米半径内的人数作为人群密度测量。而对于图7A、图7B和图7C的场景,位置的人群密度可以定义为6米半径内的人数。图7C图示了使用6米半径的人群密度映

射,更高亮度的粉色表示更高的人群密度。

[0104] 根据对于每个视频帧的人群密度测量,可以检测许多与人群相关的事件,如图8所示,包括人群检测、人群聚集和人群分散,这些可以分别由图1的事件检测模块108的模块801、802和803来检测。图9图示了如何定义并检测拥挤区域的示例性方法。方框901图示了如何定义拥挤区域事件。使用者可以首先在图像上(例如,图像空间内)选择感兴趣的区域。接着,一些人群密度阈值可以用于确定对多少人群感兴趣。阈值可以是某一半径的区域内的人数。滞后阈值可以用于更稳健的性能。例如,如果我们定义人群密度为3米半径区域内的人数,可以设置两个人群密度阈值: $T_{高}=10$ 并且 $T_{低}=8$ 。仅当对应的人群密度大于或等于 $T_{高}$ 时,才能认为区域是拥挤区域。仅当对应的人群密度变成小于或等于 $T_{低}$ 时,拥挤区域才变得不拥挤。拥挤区域可以由识别的人群定义,并且不同帧之间的拥挤区域可以改变位置和/或形状。拥挤区域的形心可以用于描述拥挤位置。最小持续时间阈值可以定义拥挤区域在触发事件检测之前必须保持拥挤的最小持续时间。对于新的视频帧输入,方框902浏览所有检测的人物目标以查看其是否属于拥挤区域,然后方框903检查所有的拥挤区域以更新它们的状态。一旦检测到人群,可以在视频图像中逐帧跟踪人群和他们的位置。例如,只要检测到人群并且人群继续满足最小阈值 $T_{低}$,与拥挤区域相关的人物模型可以定义视频图像的后续帧中的人群,只要人群保持在满足最小人群密度的区域中。当人群进入检测的拥挤区域中时,额外的人物模型可以添加到检测的人群中。

[0105] 图10图示了各检测的人物目标的示例性处理。方框1001检查当前目标是否在现有拥挤区域中或附附近。如果“是”,则方框1001更新该区域的人数。如果“否”,则方框1002计算当前目标位置的人群密度,然后方框1004检查人群密度测量是否大于或等于阈值 $T_{高}$ 。如果“是”,则以当前目标为中心创建新的拥挤区域。如果“否”,则继续处理下一个目标。

[0106] 图11图示了对每个拥挤区域的示例性处理。方框1101根据目标处理结果更新区域面积和人群计数;方框1102检查密度计数是否仍然大于使用者定义的阈值;如果“否”,则从监测列表中去掉拥挤区域。方框1104进一步检查处理中的拥挤区域的拥挤持续时间是否大于或等于使用者定义的阈值。如果“是”,方框1105进一步检查是否已经报告对应的拥挤事件,如果否,方框1106将采取措施,例如,报告拥挤事件并且将该拥挤区域标记为“已报告”。

[0107] 图12图示了可用于定义并检测人群“聚集”和“分散”事件的方法。这里“聚集”和“分散”指的是人群聚集点形成和结束的两个过程。在此实例中,人群聚集点指的是具有很高的局部固定人群密度的区域,并且不同于例如游行中的移动人群。然而,本发明不限于此并且该方法还可以应用于检测移动的人群聚集点。方框1201图示了可以如何定义人群聚集点。使用者可以首先在图像上选择感兴趣的区域。接着,一些人群密度阈值可以用于确定对多少人群感兴趣。最小持续时间阈值可以定义拥挤区域必须保持被认为有效聚集点的人群的最小持续时间。方框1202检测人群聚集点。方框1203更新并监测检测的人群聚集点,并且检测人群“聚集”和“分散”事件。

[0108] 图13图示了定义人群聚集点的一个实例。该实例包括由1301表示的内区以及由1302表示的外区。两个区可以由中心点 O 、短半径 r 和长半径 R 定义。在此实例中,人群聚集点可以满足以下两个准则:

[0109] • 内区的人群密度必须大于或等于预定阈值;

[0110] • 外区的人群计数必须小于(例如,2倍、4倍、10倍等更小)内区的人群计数。可替

代地,外区中的人群密度必须小于(例如,2倍、4倍、10倍等更小)内区的人群密度。

[0111] 上述两个准则可以表示内区是人群聚集点,而不只是具有一大群人的区域。

[0112] 图14A和图14B示出了人群聚集点的一个实例。图14A和图14B均图示了映射到真实世界的物理地平面上的视频帧和检测的人物目标。尽管图14A具有更多的人物目标,仅图14B包括按照上述方式定义的人群聚集点。

[0113] 图15图示了检测人群聚集点的示例性方法。对于每个检测的人物目标,方框1501检查人物目标是否属于现有的人群聚集点。如果“是”,则在方框1502中该人物目标用于更新对应的人群聚集点的当前状态。如果“否”,方框1503进一步检查当前目标是否是新的人群聚集点的中心。如果“是”,则方框1504开始用于进一步监测的新的人群聚集点。如果“否”,则模块继续检测下一个检测。

[0114] 图16图示了更新人群聚集点并检测人群“聚集”和“分散”事件的示例性方法。方框1601通过使用考虑中的视频帧的新的人物检测结果更新人群聚集点的位置和区域。方框1602检查从当前人群聚集点是否已经检测到人群“聚集”事件。如果“否”,则方框1603通过检查是否已经成功更新某一持续时间的人群聚集点来继续检测“聚集”事件。该持续时间阈值可以由使用者在规则定义时设置。一旦人群聚集点已经产生“聚集”事件,方框1604进一步监测聚集点以检测“分散”事件。这里,人群“分散”事件定义为人群聚集点在短时间段内变成空点或具有低密度的点(例如,低于最小人群密度阈值 $T_{低}$)。方框1604检测人群聚集点的两个特殊时刻:人群聚集点变得不拥挤的时间以及人群聚集点变空或密度变低的时间。如果这两个时刻之间的时间比使用者定义的阈值短,则检测到人群“分散”事件。

[0115] 图17图示了可以应用本发明的多摄像机系统的实例。在此实例中,两个摄像机1702和1704从不同的视角单独拍摄感兴趣的场景的视频图像。本发明描述的视频监控系统101和方法可以与本发明中针对每个摄像机1702和1704的变化检测模块103、运动检测模块104、前景斑点检测模块105、一般人物模型模块303、基于人物的摄像机校正模型模块304和人物概率映射计算模块305所描述的相同,也就是说,每个摄像机可以具有它们自己的模块或用于这些模块的模块功能(如果电路共享)。

[0116] 由各摄像机1702、1704的基于人物的摄像机校正模型模块304提供的相应图像空间的二维人物模型也可以与真实世界的物理地平面的坐标关联上。例如,对于各摄像机的基于人物的摄像机校正模型模块304,可以制作额外的条目用于对应的物理地平面坐标,从而使N个人物模型的每个人物模型与物理地平面坐标关联上。在计算每个摄像机1702、1704的人物概率映射中,各概率映射的概率可以映射到物理地平面而非二维图像空间。

[0117] 在一个实例中,检测人物的最佳数量的人物目标估计模块306可以按照上述方式对一个摄像机的第一概率映射进行扫描,也就是说,在搜索准则的约束范围内,搜索第一概率映射的局部最大值。在计算准则函数以确定M组人物模型 $m(m_1, \dots, m_M)$ 的最大值时,目的是找到:

$$[0118] \quad \underset{n \in m}{\operatorname{argmax}} f_1(n) + f_2(n)$$

[0119] 其中,n是多个三维人物模型的特定组,这些三维人物模型可以在两个人物概率映射的每个概率映射的概率所映射到的物理地平面中具有识别坐标。也就是说,在选择真实世界中的点与模型组的人物模型关联时,识别每个摄像机系统的与该点相关的二维图像空

间中的人物模型,一个人物模型用于计算 $f_1(n)$ 并且另一个人物模型用于计算 $f_2(n)$ 。 $f_1(n)$ 和 $f_2(n)$ 可以与本发明所述的函数相同(与人物前景斑点集对应或者与从合适的视频图像提取的人物前景区域对应):

$$[0120] \quad f(n) = w_R * R(n) + w_P * P(n) - w_0 * O(n)$$

[0121] 其中(对于与视频图像和该视频图像的人物前景区域相关的相应的n个选择的二维人物模型),R是人物调用率,其定义为人物前景面积与n个选择的人物模型的组的整个面积的百分比;P是人物精度,是与n个选择的人物模型的组重叠的前景面积的百分比,并且O是人物重叠比,是n个选择的人物模型的任意人物模型彼此重叠的面积与全部n个选择的人物模型占据的面积比值(选择的第n个人物模型具有第1至第n-1个人物模型占据的面积)(计算 $f(n)$ 时,当前扫过之前所选择的人物模型占据的面积)的比值,并且 w_R 、 w_P 和 w_0 是权重。要注意,函数 $f_1(n)$ 和 $f_2(n)$ 的权重可以不同。在选择下一个局部最大值中排除像素不作进一步考虑可以将与此前选择的人物模型的地平面坐标相关的三维人物模型返回投射到各个图像平面中的两个概率映射的每个概率映射上。

[0122] 在另外的替代方式中,单个概率映射可以供多个摄像机使用。在图17的实例中,可以对每个二维摄像机图像进行本发明所述的概率计算,并且创建两个图像平面概率映射,每个概率映射与相应的二维图像平面对应。图像平面的概率映射的概率可以设置成零,如果它们没有超过某一阈值(各图像平面的概率映射的阈值可以相同或不同)。对于各图像平面的概率映射,各图像平面的概率映射中的识别坐标可以转换成真实世界中的地平面坐标,从而为各视频图像构建地平面概率映射。两个地平面概率映射可以通过乘以共享相同地平面坐标的概率而合并,从而创建合并的概率映射。合并的地平面概率映射可以经过扫描以找到局部最大值。每个找到的局部最大值可以识别它们各自的图像空间内每个视频图像的单个人物模型,然后该局部最大值可以酌情用于计算 $f_1(n)$ 或 $f_2(n)$ (如上所述)。可以对多个局部最大值的合并的地平面概率映射进行多次扫描以找到后续的人物模型(每个视频图像找到一个人物模型)并且计算。

$$[0123] \quad f_1(n) + f_2(n)。$$

[0124] 可以改变选择约束(例如,三维真实世界中的最小概率阈值和最小距离),并且实施新的扫描以找到m个三维人物模型(在此实例中与 $2m$ 个二维人物模型对应)的最佳组。

[0125] 在另一个实例中,检测人物的最佳数量的人物目标估计模块306可以按照上述方式对一个摄像机的第一概率映射进行扫描,也就是说,在搜索准则的约束范围内,搜索第一概率映射的局部最大值。在计算准则函数以确定m个人物模型的组的最大值时,目的是找到:

$$[0126] \quad \sum_{n=1}^m f_1(n) + f_2(n) \text{ 的最大值。}$$

[0127] 其中,n是在两个人物概率映射的每个概率映射的概率所映射到的物理地平面中的识别坐标。也就是说,在选择真实世界中的点时,识别每个摄像机系统的与该点相关的二维图像空间的人物模型,一个人物模型用于计算 $f_1(n)$ 并且另一个人物模型用于计算 $f_2(n)$ 。 $f_1(n)$ 和 $f_2(n)$ 可以与上述函数相同(与人物前景斑点集对应或者与从合适的视频图像提取的人物前景区对应):

$$[0128] \quad f(n) = w_R * R(n) + w_P * P(n) - w_0 * O(n)$$

[0129] 其中,R是人物调用率,其定义为人物前景面积与选择的人物模型的整个面积的百

分比;P是人物精度,是与选择的人物模型重叠的前景面积的百分比,并且O是人物重叠比,是选择的第n个人物模型与第1个至第n-1个人物模型占据的面积(计算 $\Sigma f(n)$ 时,当前扫过之前所选择的人物模型占据的面积)的重叠率,并且 w_R 、 w_P 和 w_0 是权重。要注意,函数 $f_1(n)$ 和 $f_2(n)$ 的权重可以不同。在选择下一个局部最大值中排除像素不作进一步考虑可以将与此前选择的人物模型的地平面坐标相关的三维人物模型返回投射到各个图像平面中的两个概率映射的每个概率映射上。

[0130] 在另外的替代方式中,单个概率映射可以供多个摄像机使用。在图17的实例中,可以对每个二维摄像机图像进行本发明所述的概率计算,并且创建两个图像平面概率映射,每个概率映射与相应的二维图像平面对应。图像平面的概率映射的概率可以设置成零,如果它们没有超过某一阈值(各图像平面的概率映射的阈值可以相同或不同)。对于各图像平面的概率映射,各图像平面的概率映射中的识别坐标可以转换成真实世界中的地平面坐标,从而为各视频图像创建地平面概率映射。两个地平面概率映射可以通过乘以共享相同地平面坐标的概率而合并,从而创建合并的概率映射。合并的地平面概率映射可以经过扫描以找到局部最大值。每个找到的局部最大值可以识别它们各自的图像空间每个视频图像的单独人物模型,然后该局部最大值可以酌情用于计算 $f_1(n)$ 或 $f_2(n)$ (如上所述)。可以对多个局部最大值的合并的地平面概率映射进行多次扫描以找到后续的人物模型(每个视频图像找到一个人物模型)并且计算。

[0131] $\sum_{n=1}^m f_1(n) + f_2(n)$ 。

[0132] 可以改变选择约束(例如,三维真实世界中的最小概率阈值和最小距离),并且实施新的扫描扫过以找到m个三维人物模型(在此实例中与 $2m$ 个二维人物模型对应)的最佳组。

[0133] 上述内容是对示例性实施例的说明并不构成对示例性实施例的限制。虽然已经描述了几个示例性实施例,但是本领域技术人员容易理解的是,在不实际脱离本发明的新教导和优点的情况下可以在示例性实施例中进行许多修改。例如,尽管本公开已经描述了检测视频图像中的人物对象,但是本发明不应当被理解为局限于此并也可以检测感兴趣的其它对象。

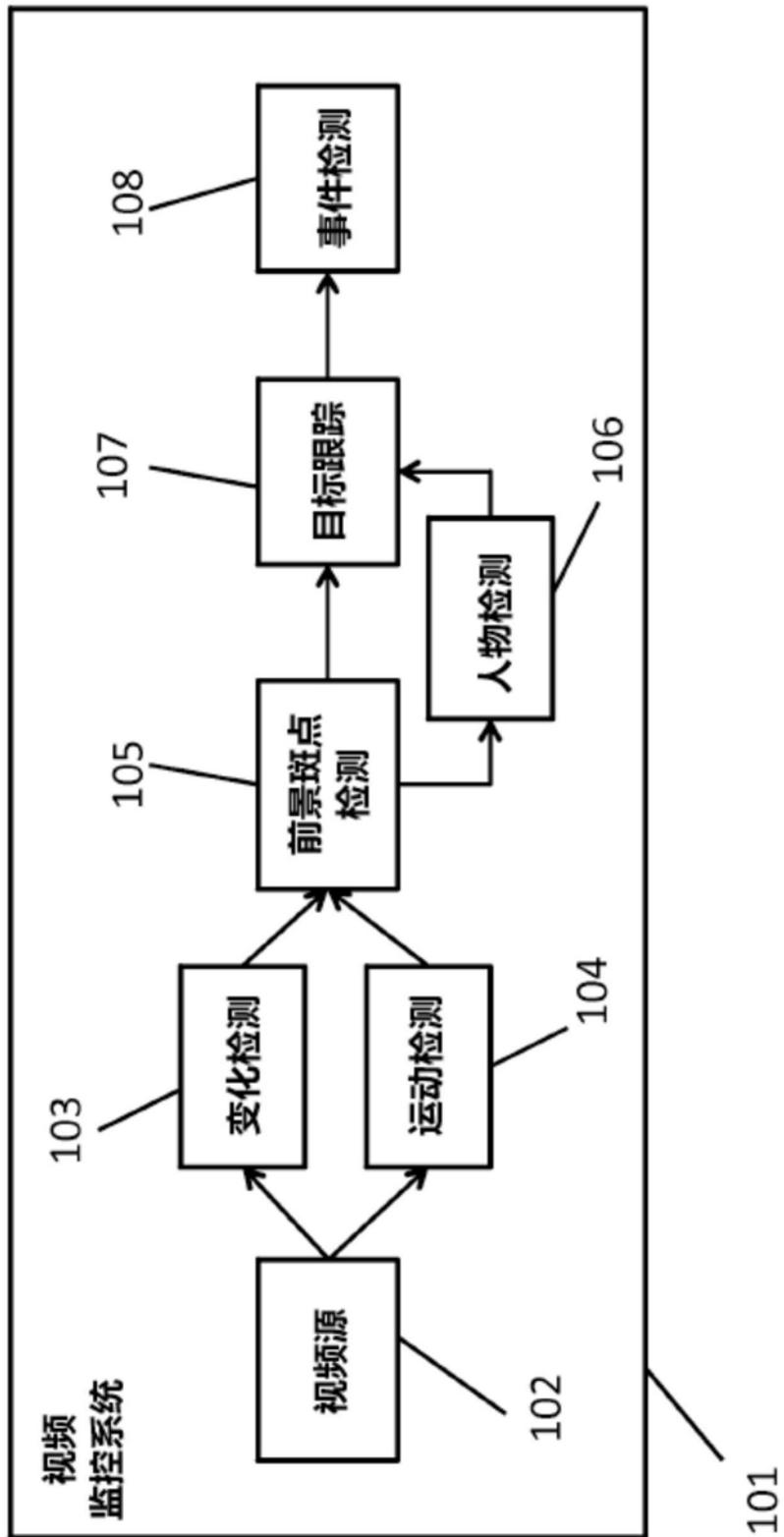


图1

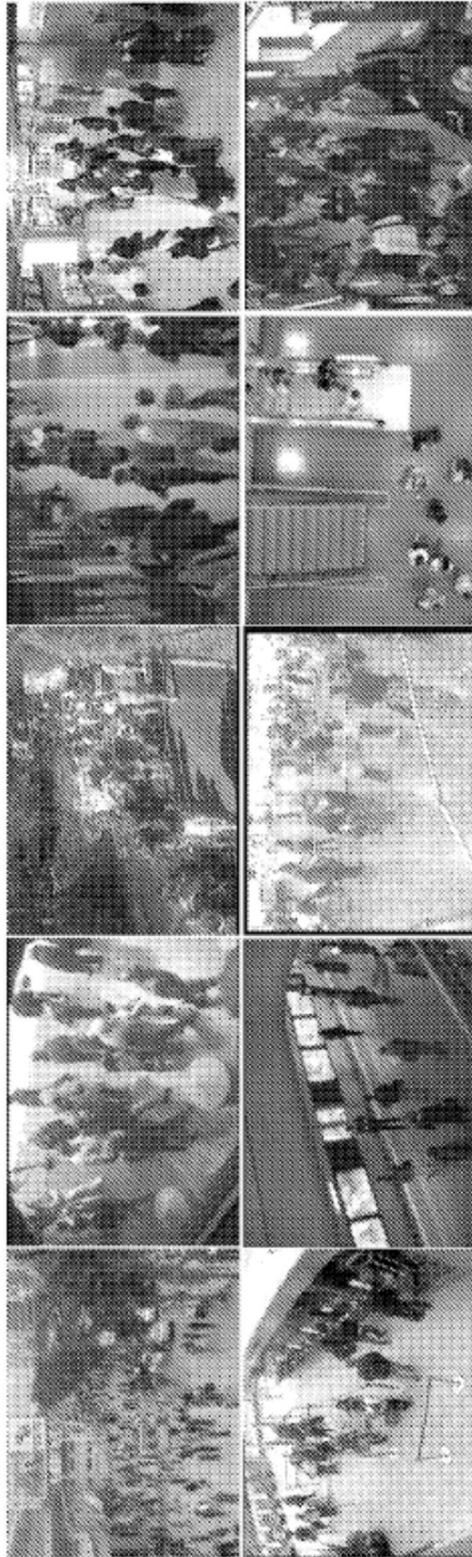


图2

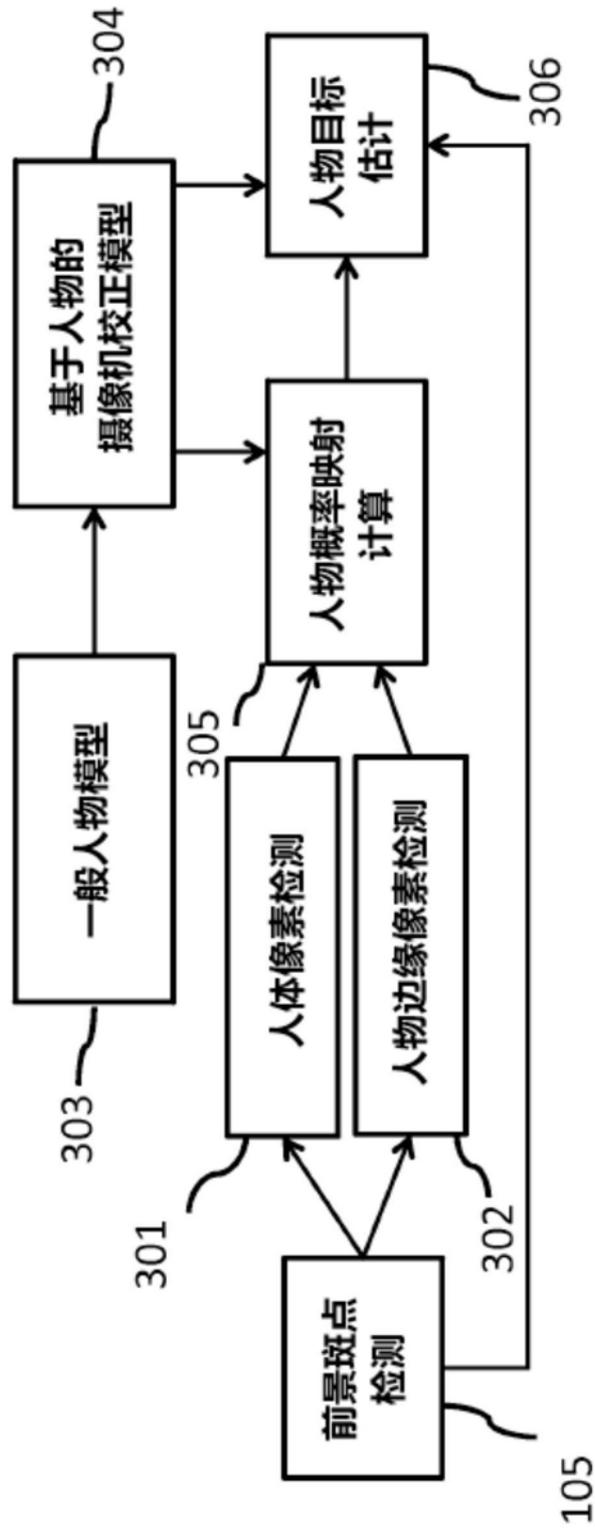


图3A

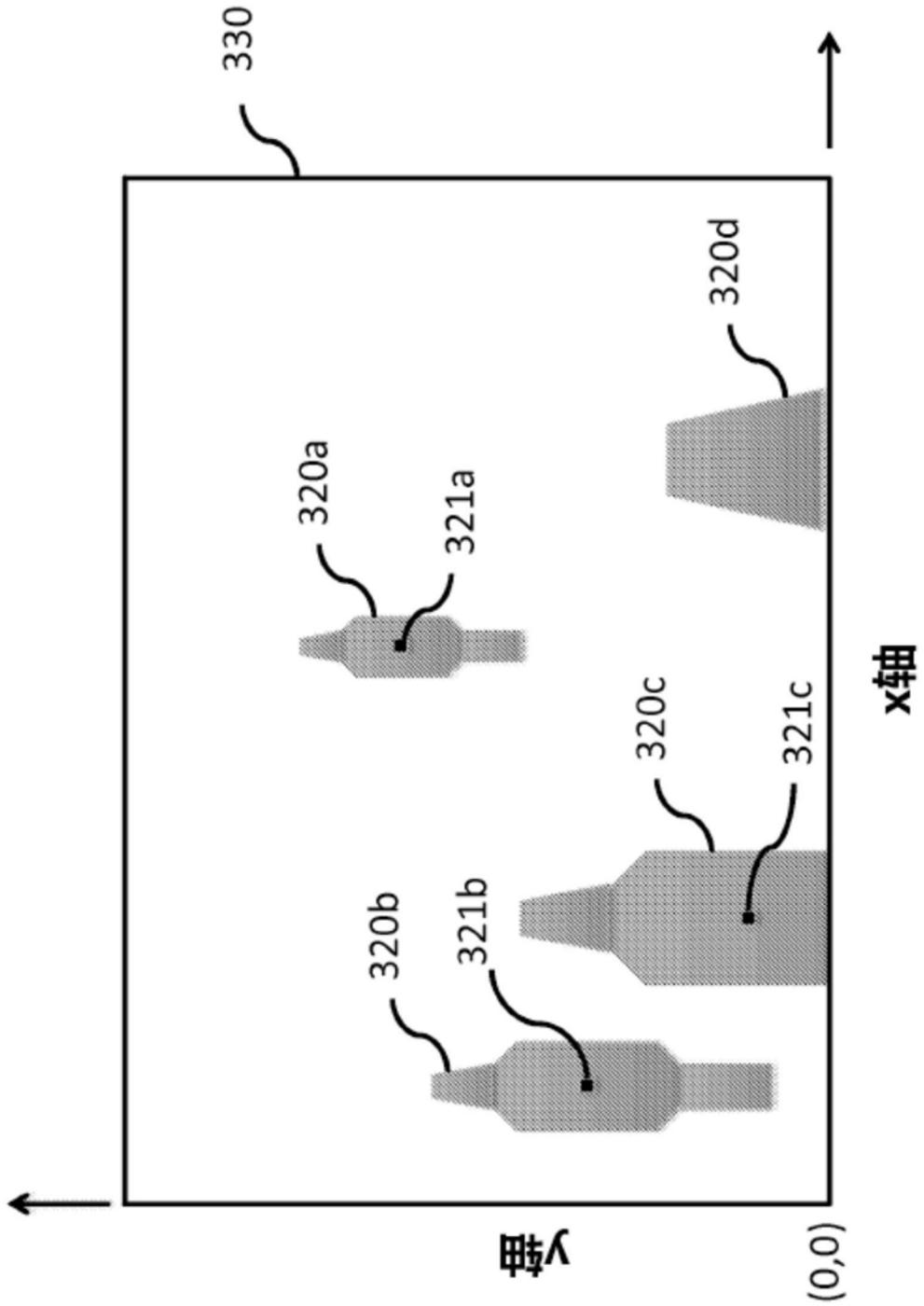


图3B

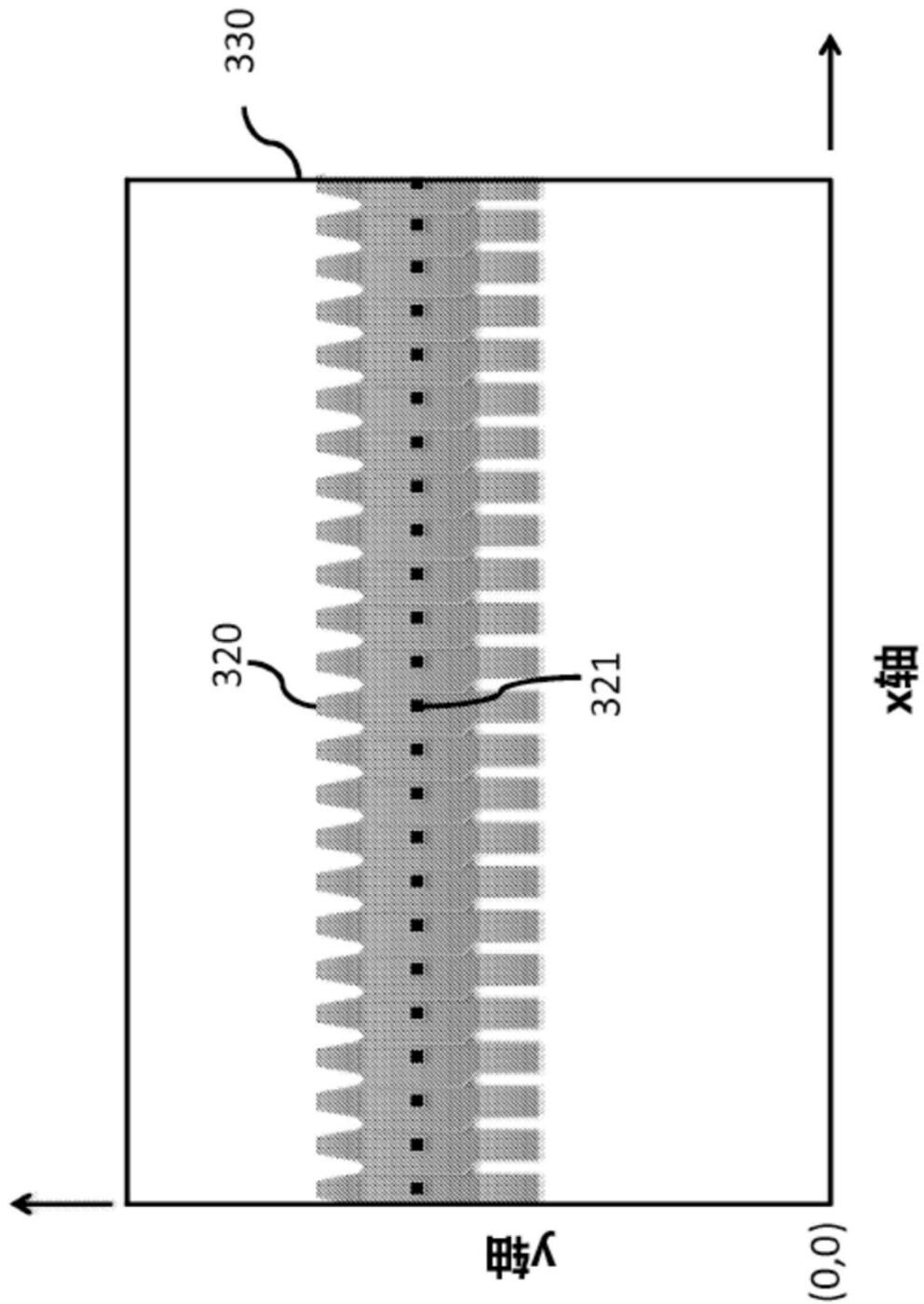


图3C

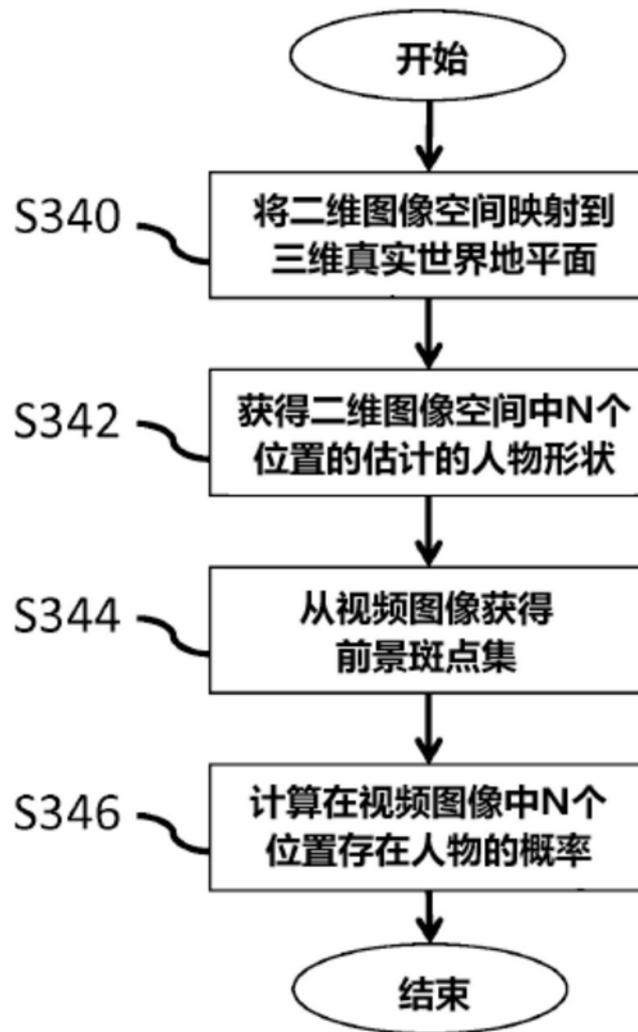


图3D

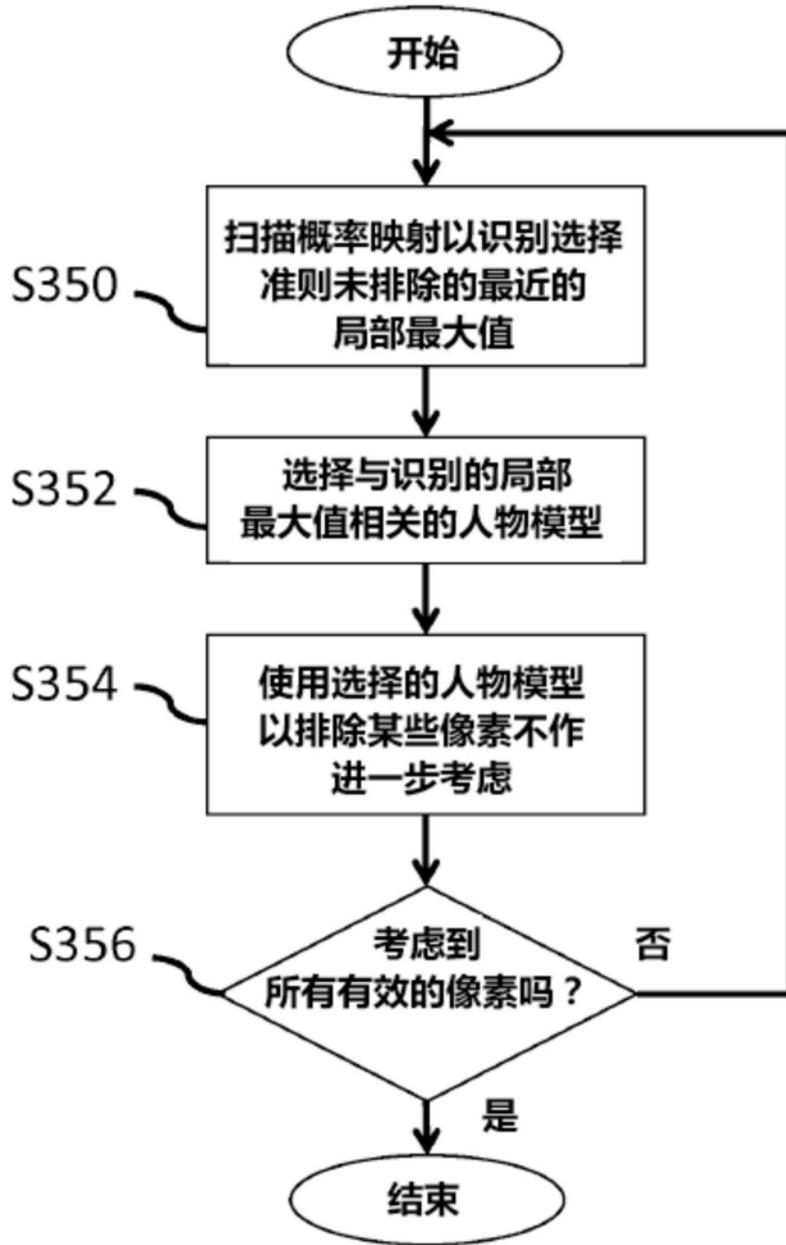


图3E

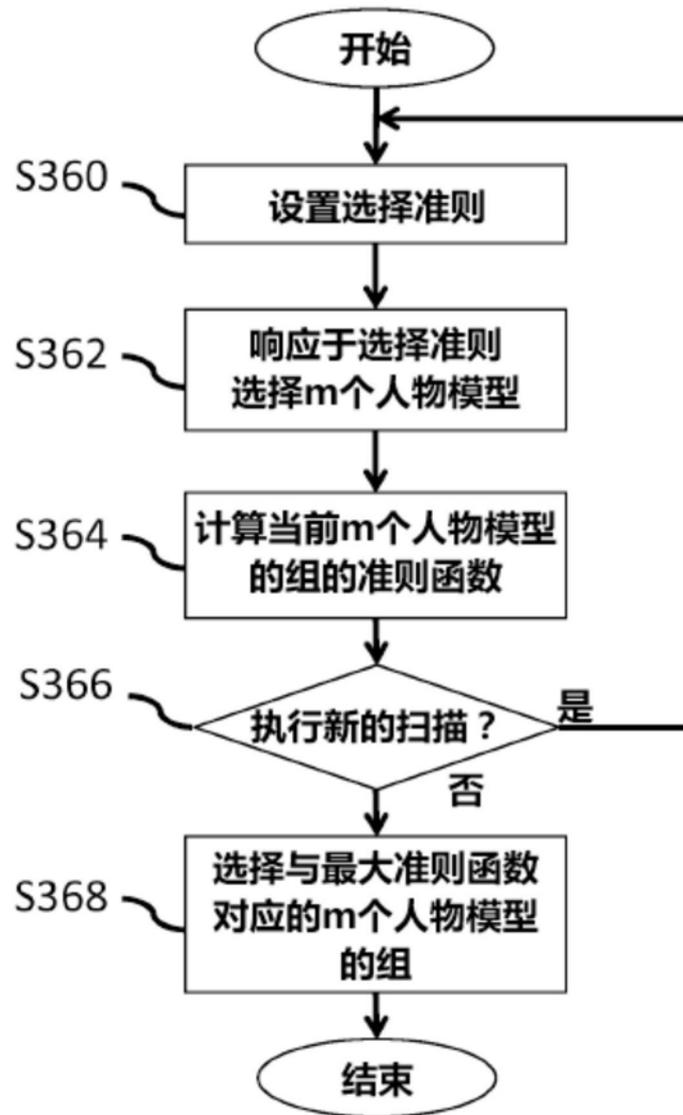


图3F

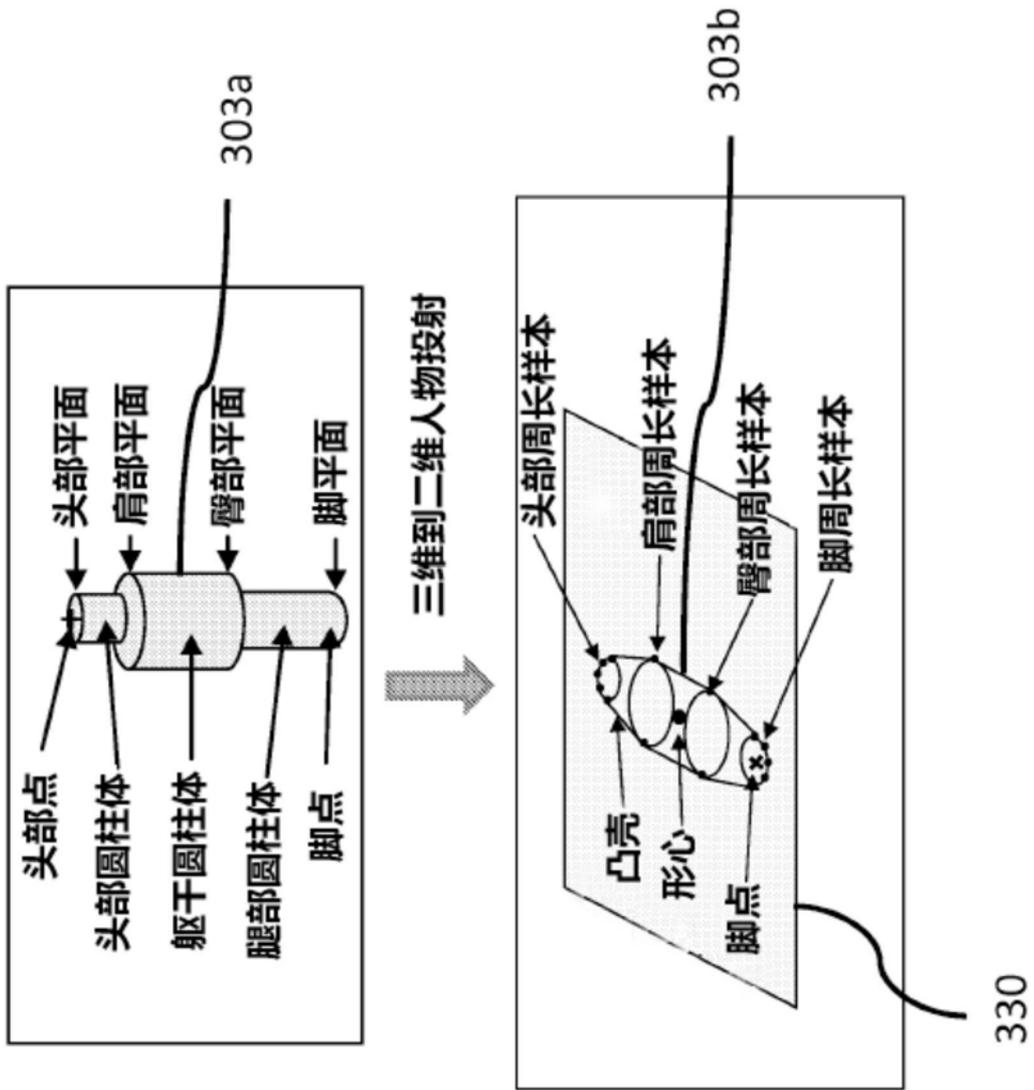


图4

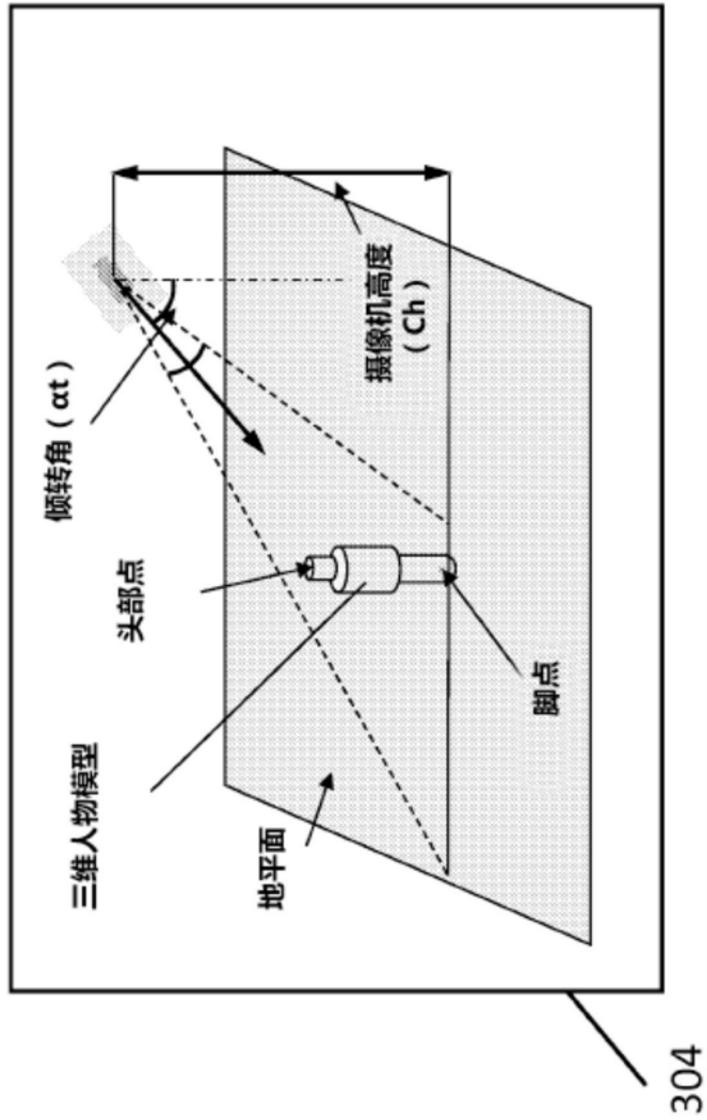


图5

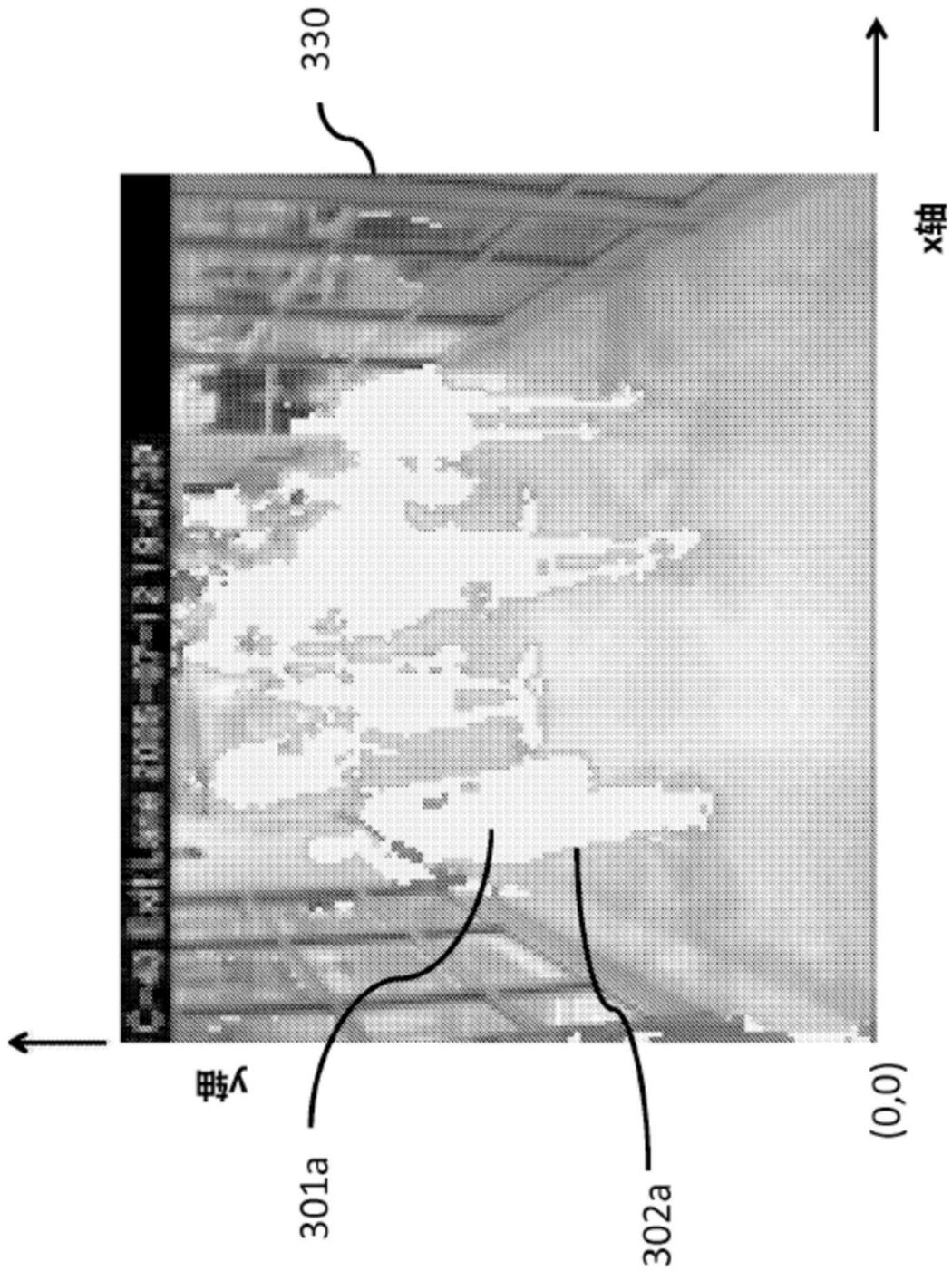


图6A

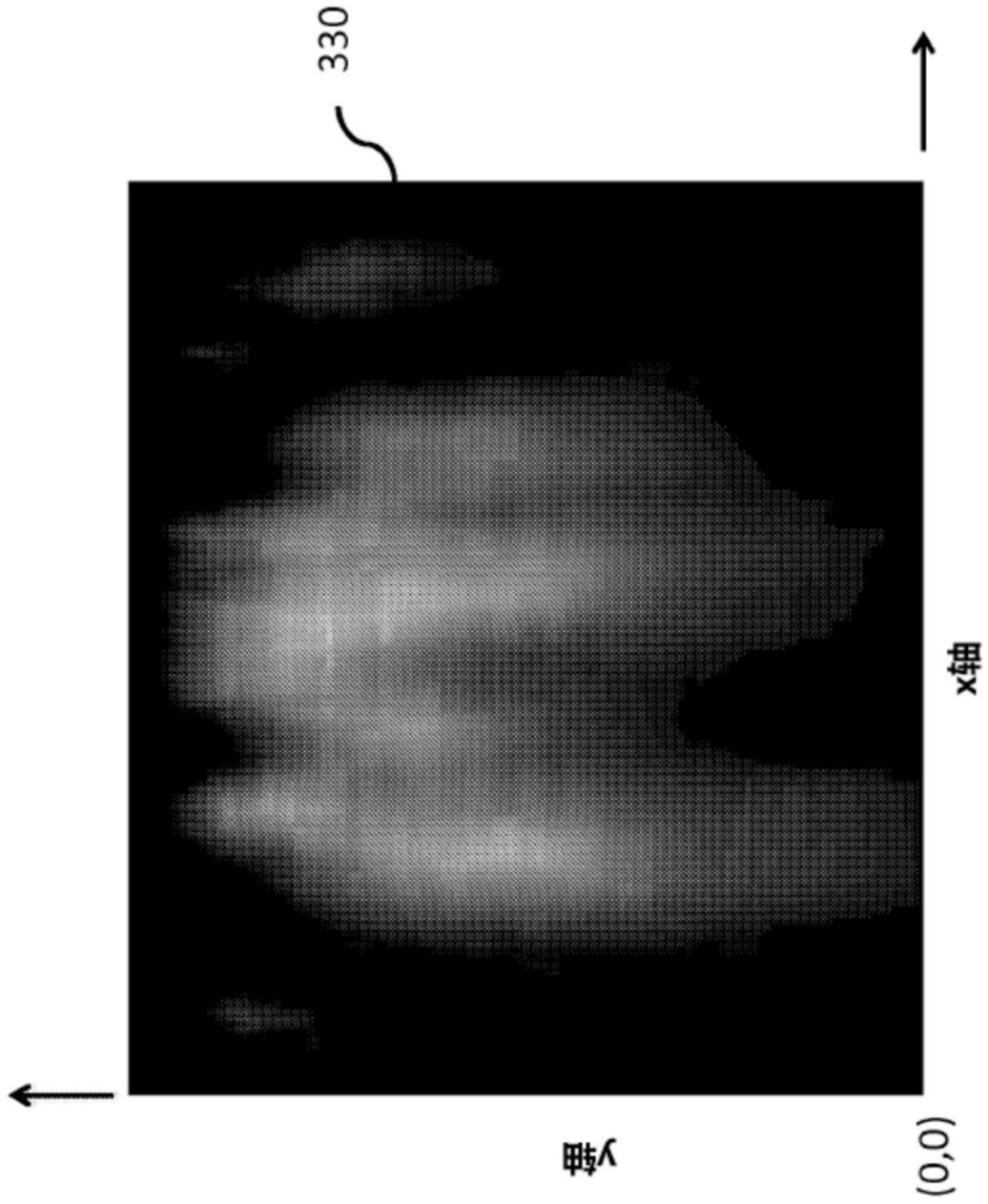


图6B

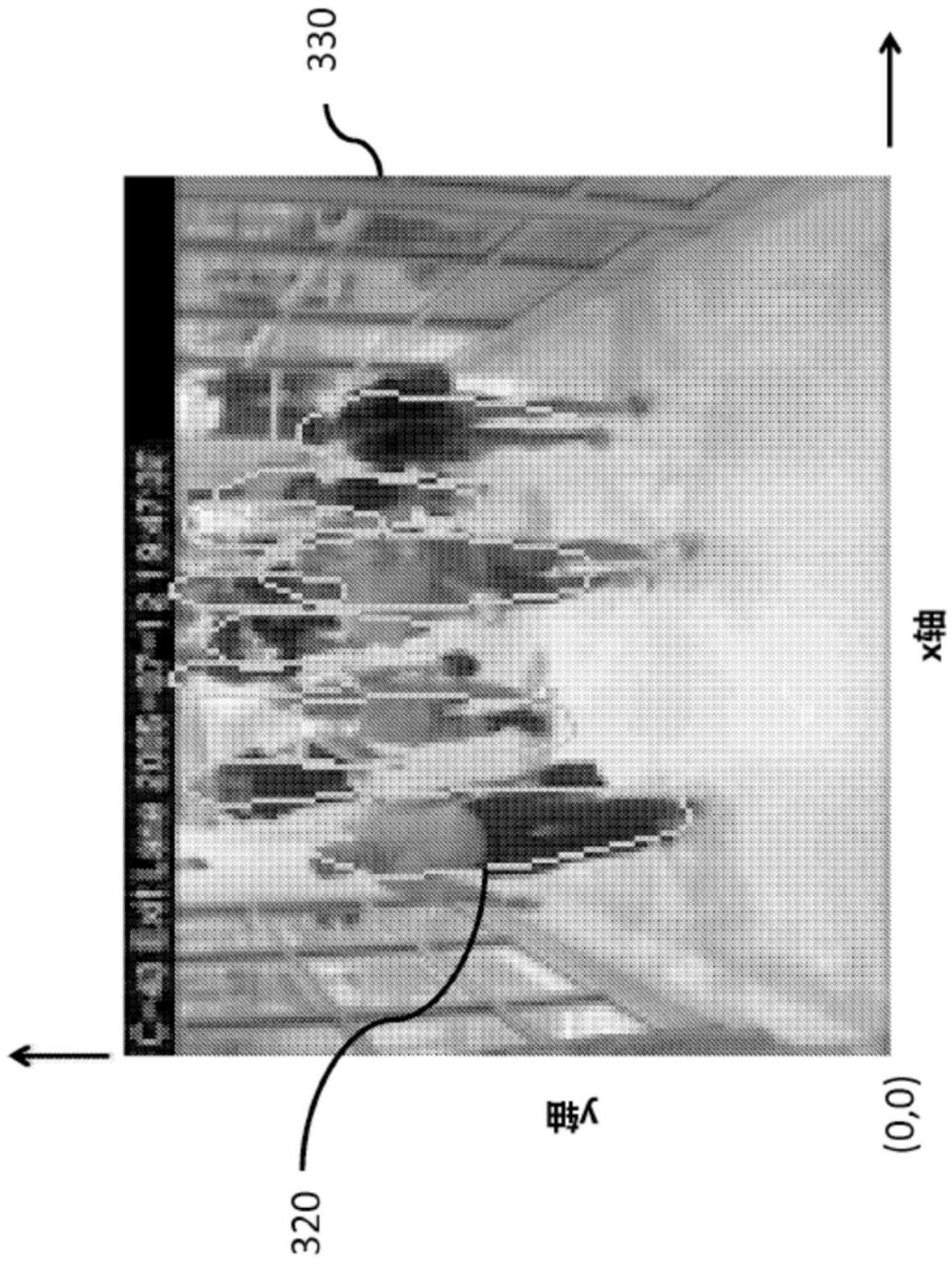


图6C



320

图7A

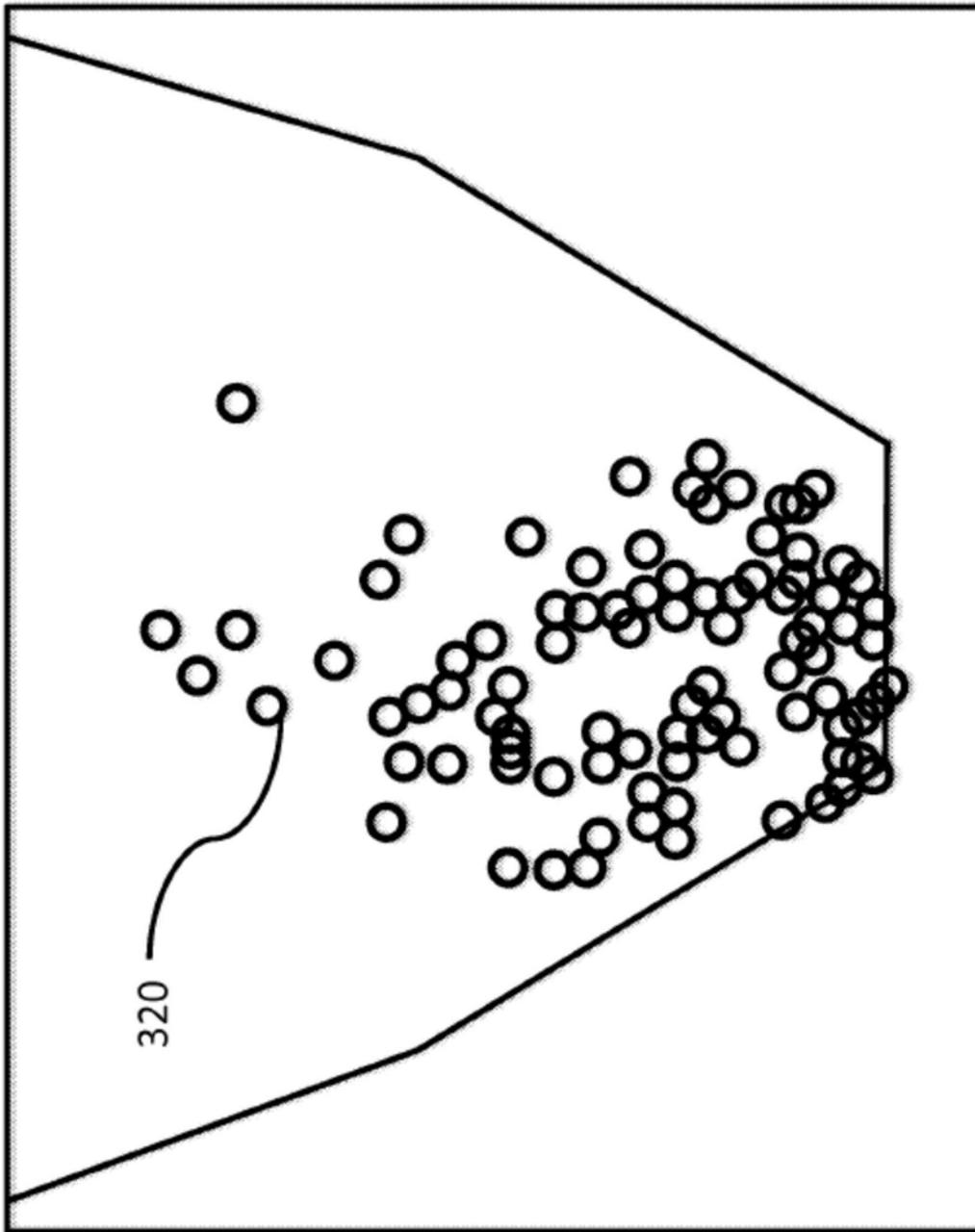


图7B



图7C

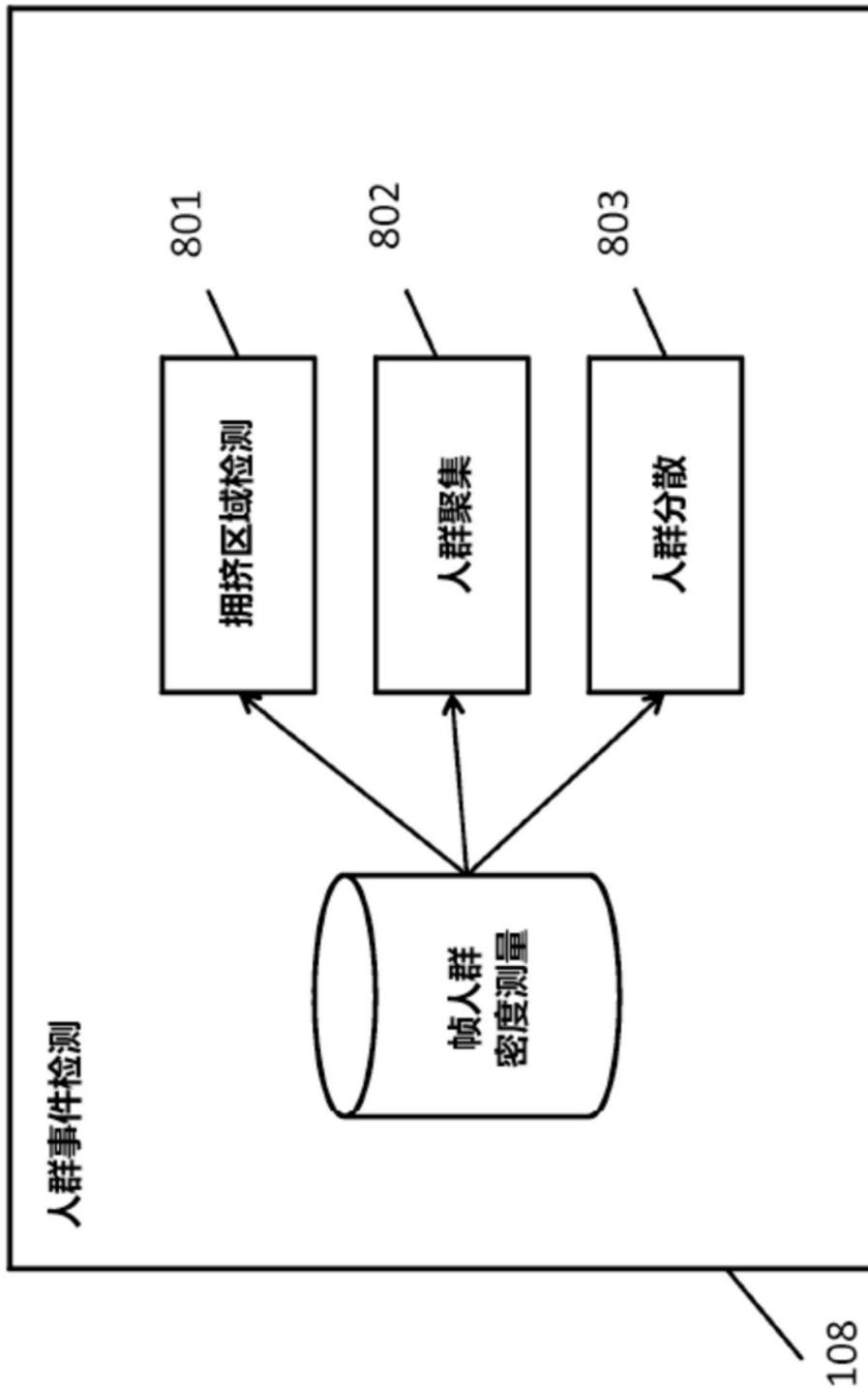


图8

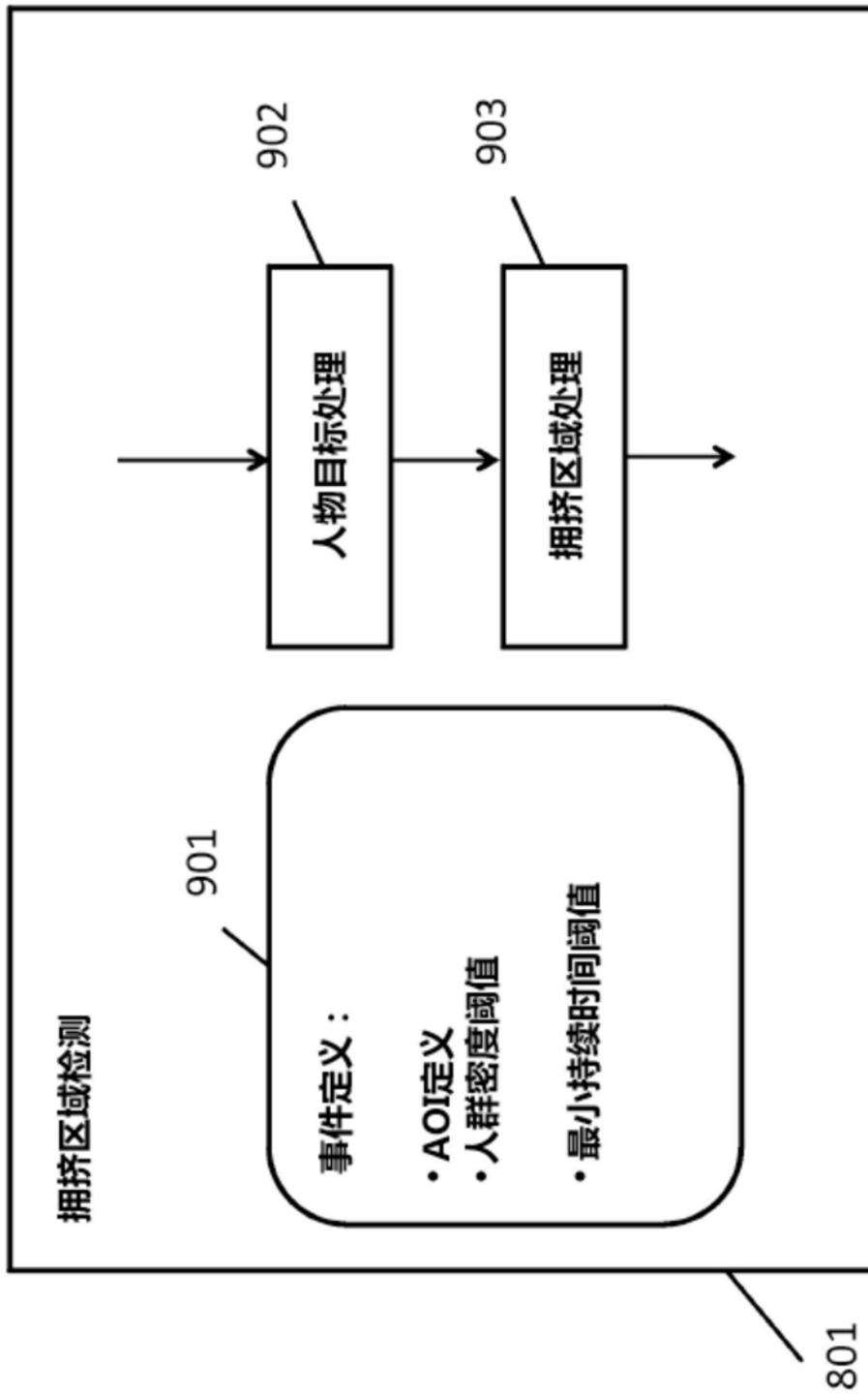


图9

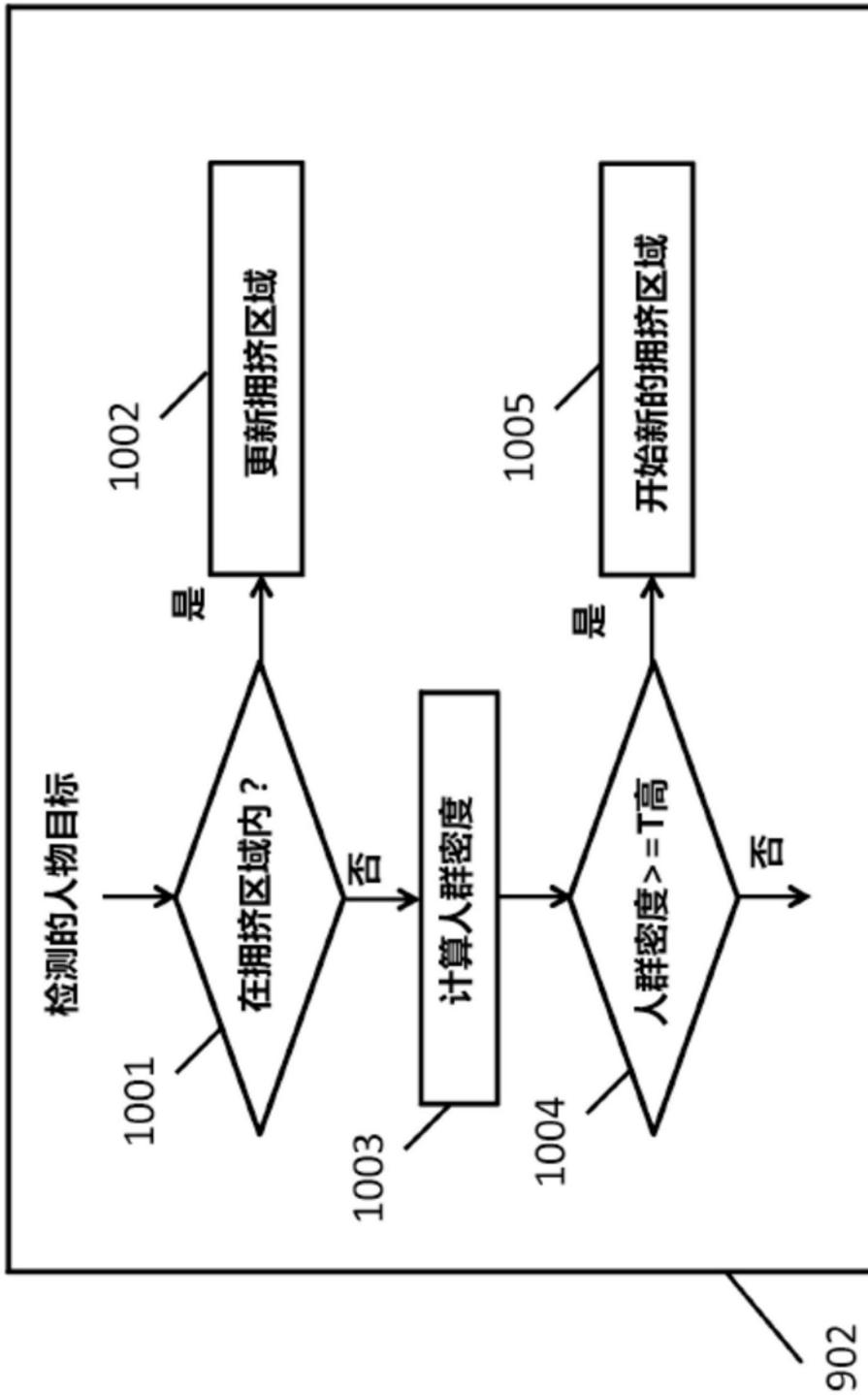


图10

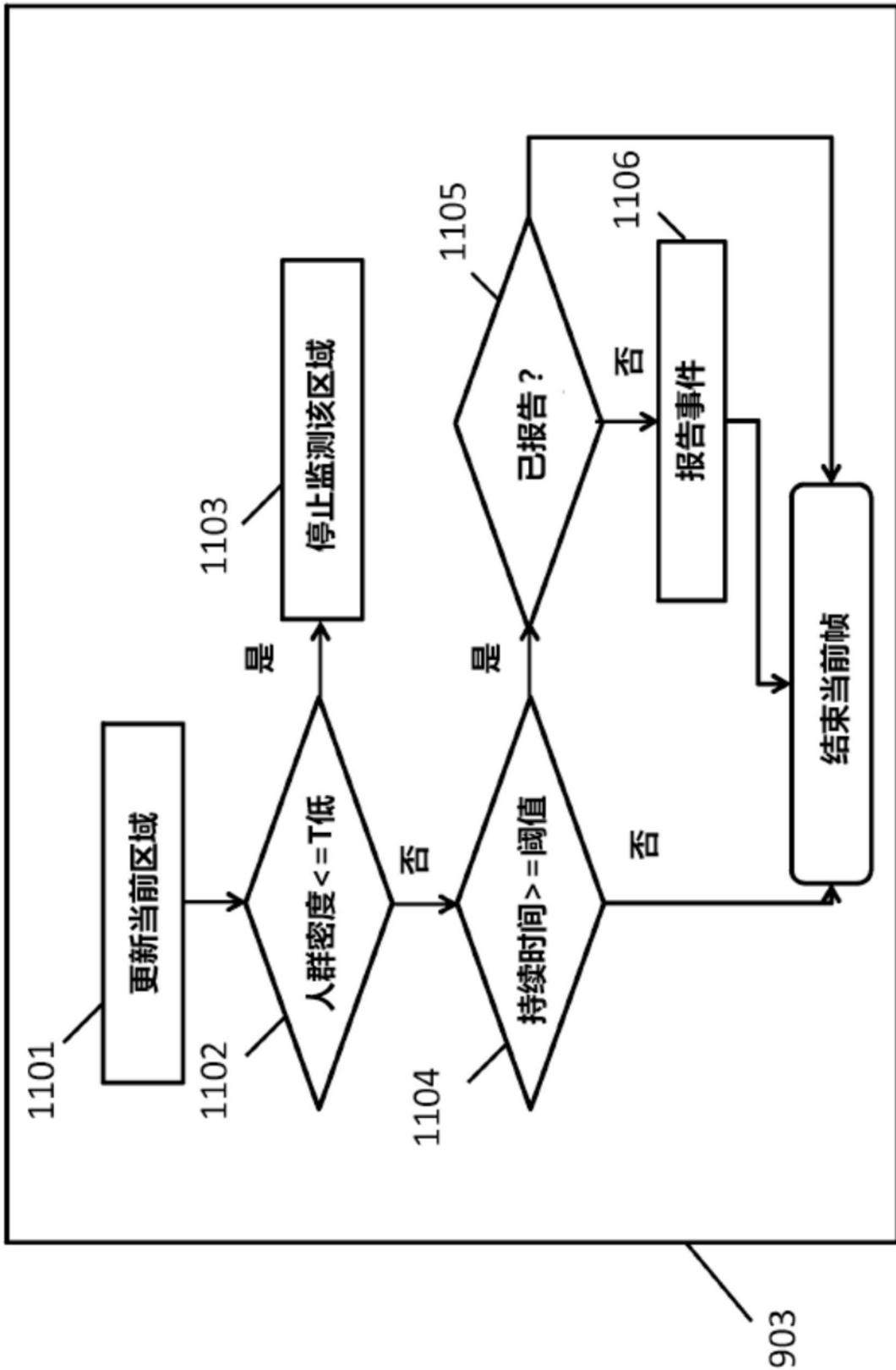


图11

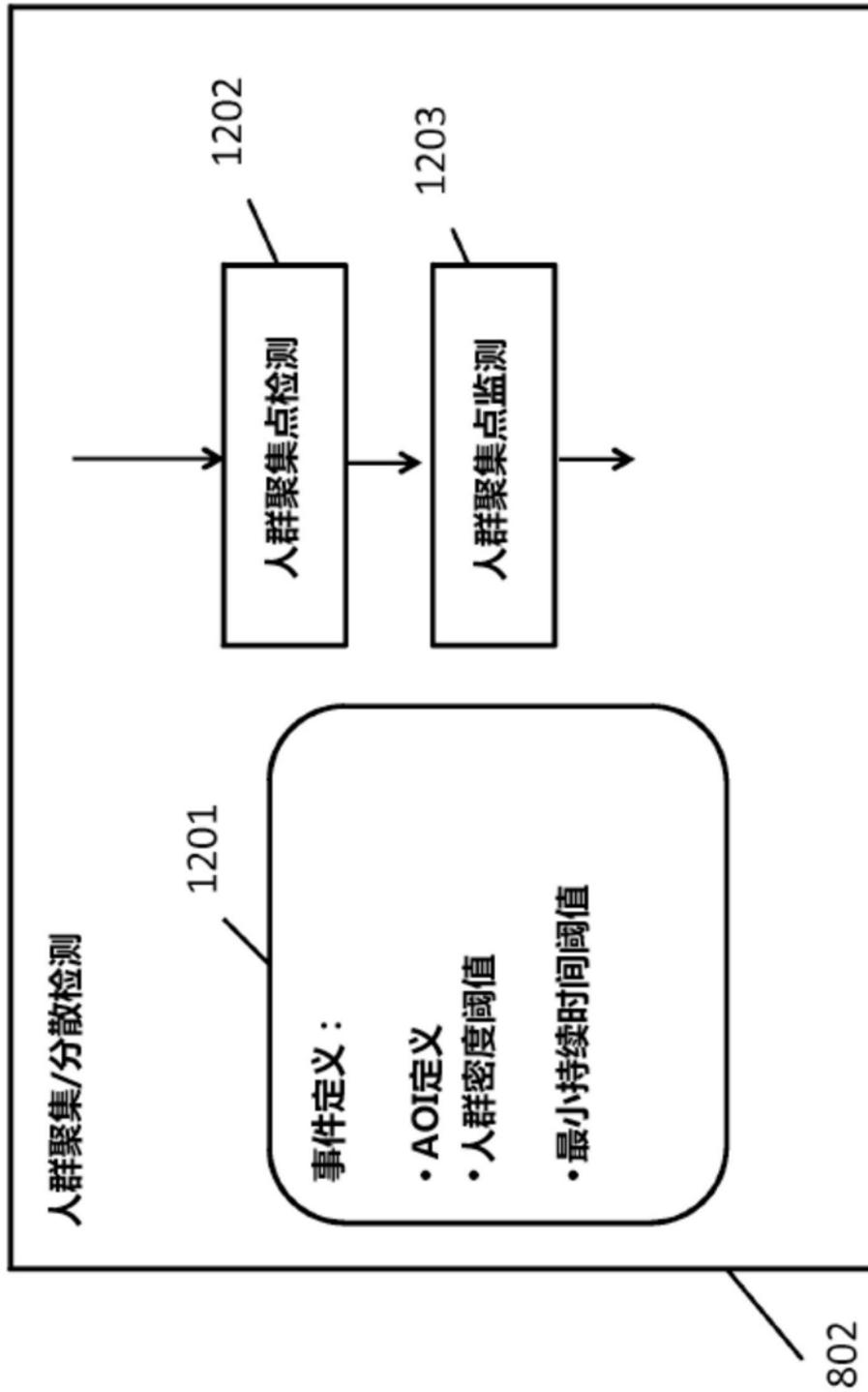


图12

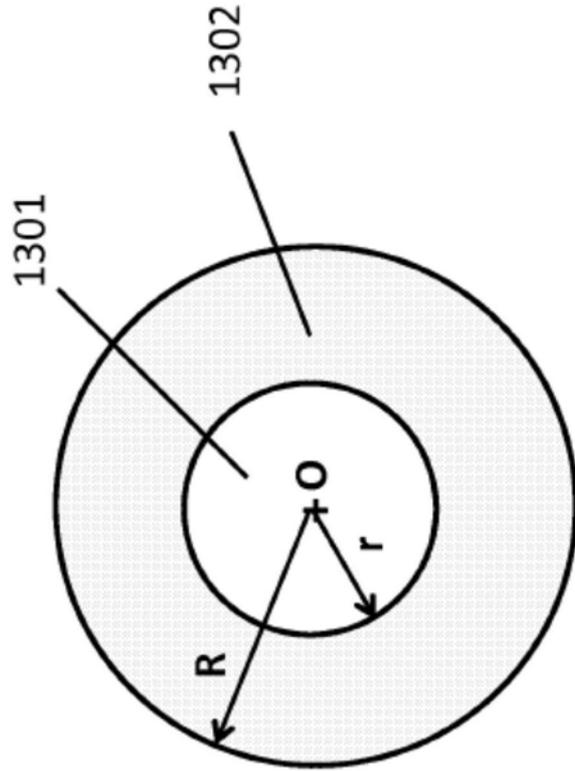


图13

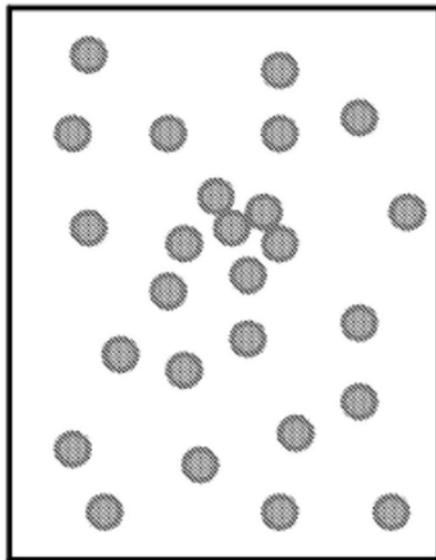


图14A

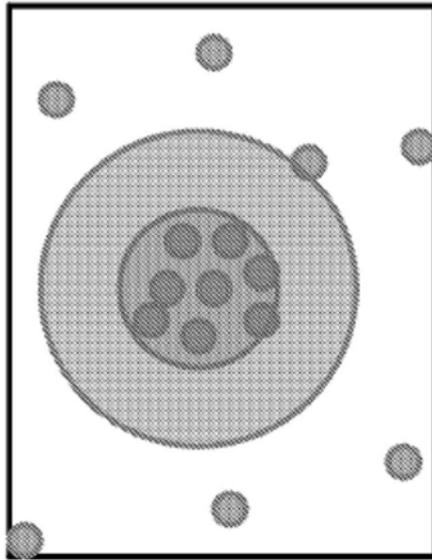


图14B

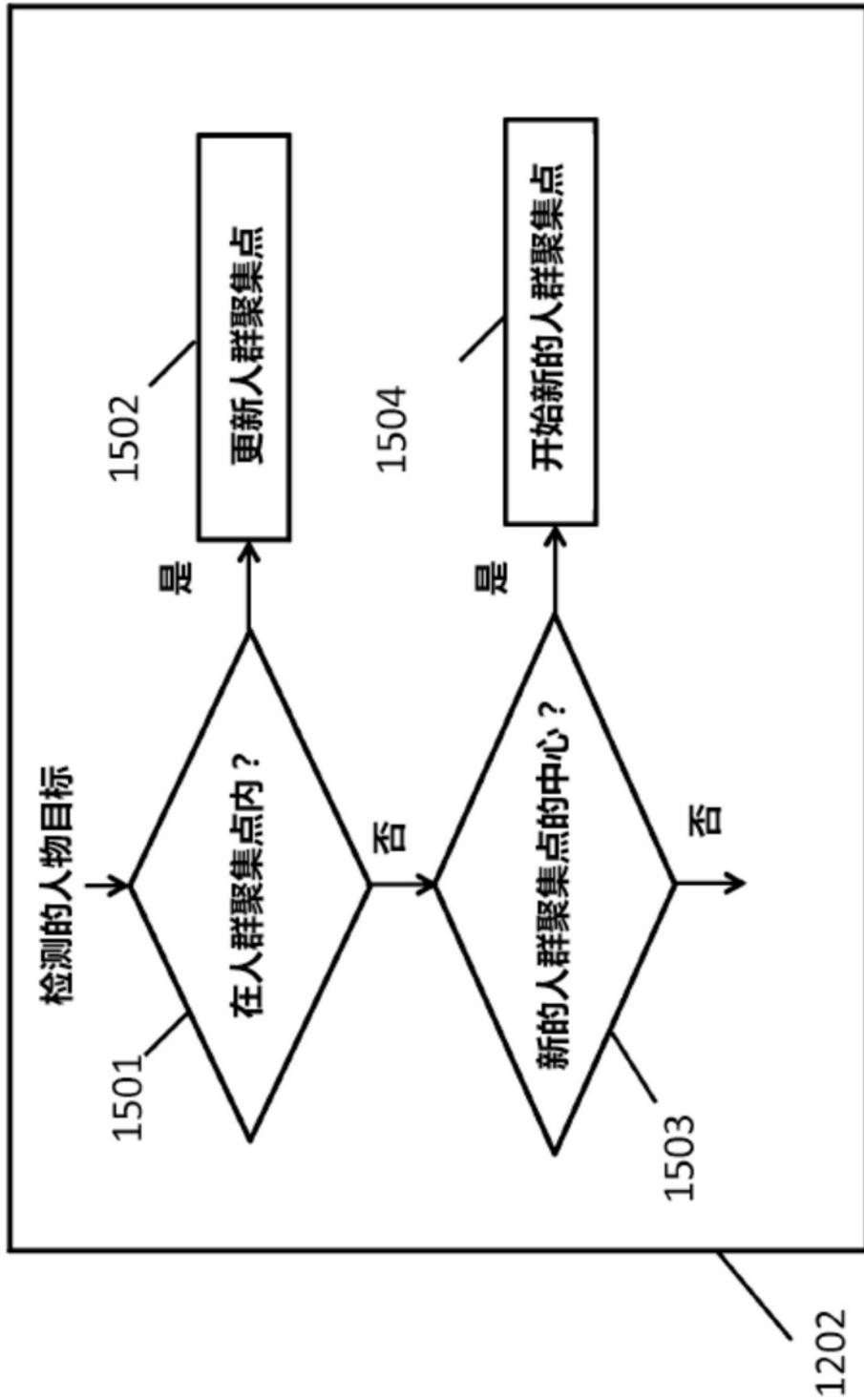


图15

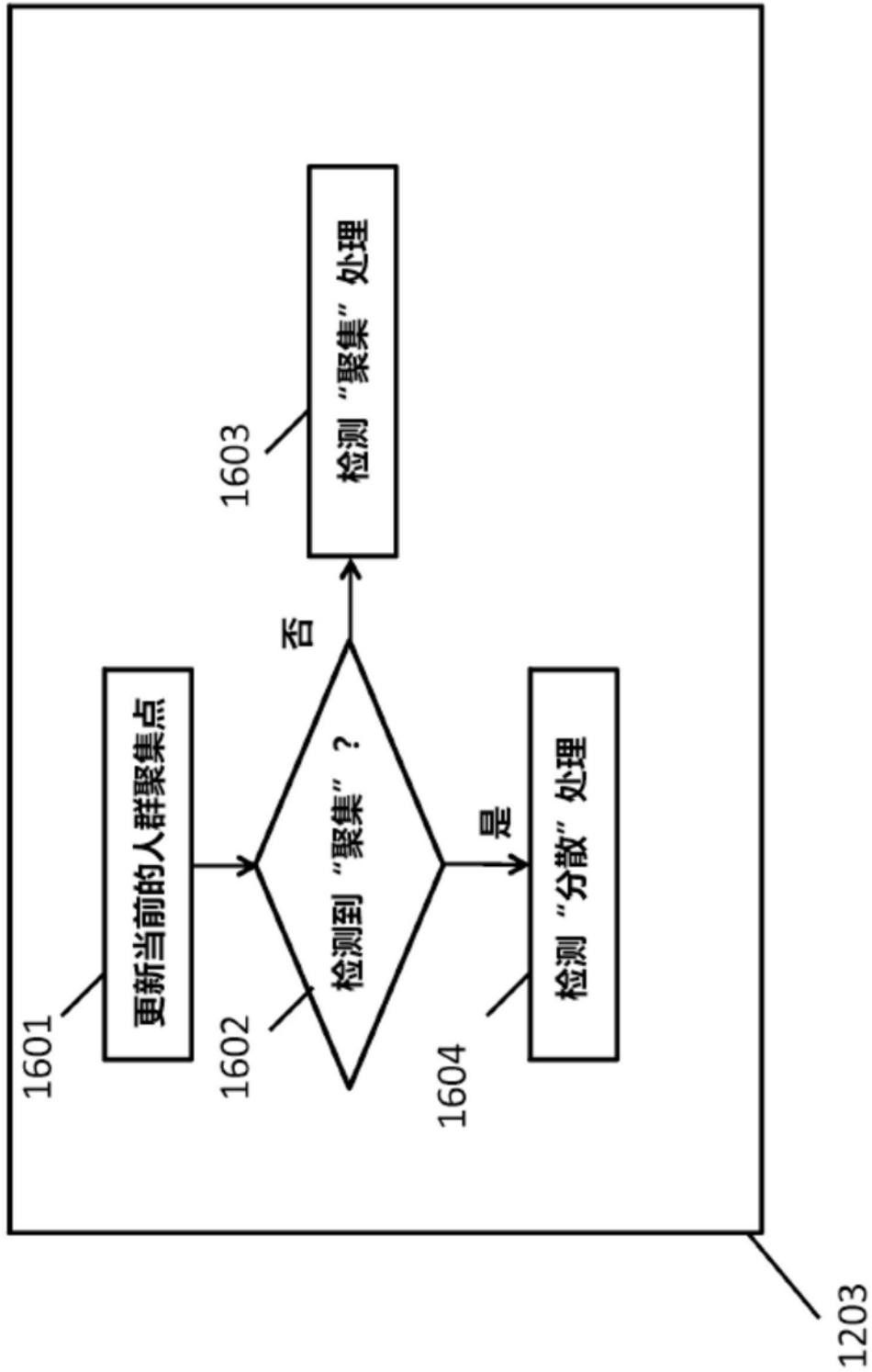


图16

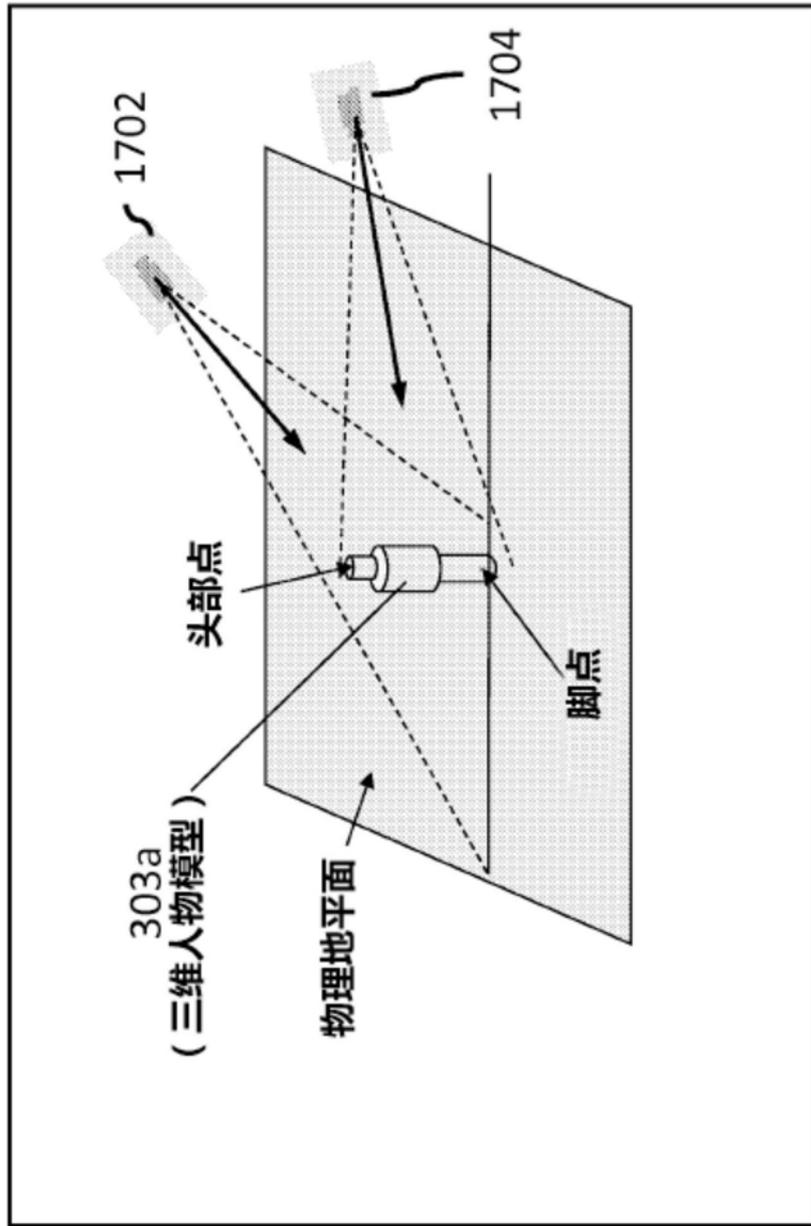


图17