



(12) 发明专利申请

(10) 申请公布号 CN 105531995 A

(43) 申请公布日 2016. 04. 27

(21) 申请号 201480039208. 2

(51) Int. Cl.

(22) 申请日 2014. 05. 09

H04N 7/18(2006. 01)

(30) 优先权数据

61/822, 051 2013. 05. 10 US

(85) PCT国际申请进入国家阶段日

2016. 01. 08

(86) PCT国际申请的申请数据

PCT/US2014/037449 2014. 05. 09

(87) PCT国际申请的公布数据

W02014/183004 EN 2014. 11. 13

(71) 申请人 罗伯特·博世有限公司

地址 德国斯图加特

申请人 N·奈卡尔 P·拉赫瓦尔迪

(72) 发明人 N·奈卡尔 P·拉赫瓦尔迪

(74) 专利代理机构 永新专利商标代理有限公司

72002

代理人 周家新

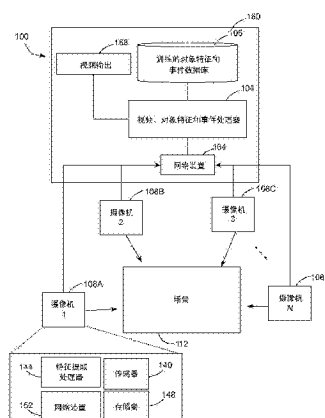
权利要求书4页 说明书13页 附图6页

(54) 发明名称

用于使用多个摄像机进行对象和事件识别的系统和方法

(57) 摘要

一种用于识别所关注的对象和事件的系统使用具有图像处理功能的一个或多个摄像机。该系统包括配置成从多个角度实施场景的图像处理的多个摄像机，以便提取和传输对应于场景中的对象或人的元数据。摄像机传输元数据至处理站，所述处理站配置成在时间上处理数据流，以便检测对象和受关注事件，从而警告场景中的对象或事件的监控人员。



1. 一种监视系统,包括:
 - 多个摄像机,其位于多个位置以便记录场景,每个摄像机还包括:
 - 传感器,其配置成生成场景的视频数据,该视频数据包括一系列帧;
 - 第一网络装置,其配置成将视频数据和与视频数据相关的特征向量传输至处理站;和
 - 特征提取处理器,其操作性地连接至所述传感器和所述网络装置,所述特征提取处理器配置成:
 - 识别由所述传感器生成的视频数据中的多个特征向量;
 - 在第一网络装置处于第一操作模式的情况下仅将所述多个特征向量传输至处理站;和
 - 在第一网络装置处于第二操作模式的情况下仅响应于处理站对视频数据的请求而将视频数据传输至处理站;和
 - 处理站,其还包括:
 - 第二网络装置;
 - 视频输出装置;和
 - 处理器,其操作性地连接至第二网络装置和所述视频输出装置,该处理器配置成:
 - 通过第二网络装置接收由所述多个摄像机中的每个摄像机生成的所述多个特征向量;
 - 参考从所述多个摄像机中的至少两个接收的所述多个特征向量而识别场景中的对象和对象动作;
 - 参考事件的预定数据库而识别与场景中的对象动作对应的事件;
 - 生成传送来自所述多个摄像机中的至少一个摄像机的视频数据的请求;和
 - 通过视频输出装置生成来自所述至少一个摄像机的视频数据的图形显示,以显示与所述事件相关的对象。
2. 根据权利要求1所述的监视系统,其特征在于,所述多个摄像机中的每个中的特征提取处理器还配置成:
 - 识别视频数据中的一帧的与场景的背景对应的一部分;和
 - 仅在所述一帧中的与场景的背景不对应的部分中识别视频数据中的所述多个特征向量。
3. 根据权利要求2所述的监视系统,其特征在于,每个摄像机中的特征提取处理器还配置成:
 - 生成边界框,该边界框对应于所述一帧的与场景的背景不对应的部分;
 - 对所述一帧的在边界框中的所述部分调整大小,以生成所述一帧的被包括在边界框中的所述部分的固定分辨率图像;和
 - 参考所述固定分辨率图像而在所述多个特征向量中识别至少一个特征向量。
4. 根据权利要求3所述的监视系统,其特征在于,特征提取处理器还配置成:
 - 在固定分辨率图像中识别多个图像梯度;和
 - 生成特征向量,所述特征向量包括与所述多个图像梯度对应的梯度直方图。
5. 根据权利要求1所述的监视系统,其特征在于,所述多个摄像机中的第一摄像机中的特征提取处理器还配置成:
 - 针对视频数据中的第一帧识别第一特征向量;
 - 通过第一网络装置将第一特征向量传输至处理站;

针对视频数据中的第二帧识别第二特征向量,第二帧在第一帧之后出现;和
通过第一网络装置将稀疏特征向量传输至处理站,该稀疏特征向量对应于第二特征向量的与第一特征向量不同的部分。

6. 根据权利要求4所述的监视系统,其特征在于,处理站中的处理器还配置成:

接收来自第一摄像机的第一特征向量;

接收来自所述多个摄像机中的第二摄像机的稀疏特征向量,该稀疏特征向量仅包括第二摄像机中所生成的、在生成较早特征向量后发生变化的特征向量部分,所述较早特征向量与来自第一摄像机的第一特征向量近似同时地生成;和

参考来自第一摄像机的第一特征向量而生成对应于来自第二摄像机的稀疏特征向量的完全特征向量。

7. 根据权利要求1所述的监视系统,其特征在于,处理站中的处理器还配置成:

参考所述多个特征向量,在来自所述多个摄像机中的每个的每个所述多个特征向量中独立地识别出事件的出现或未出现,所述多个特征向量对应于来自在单个时间段期间对应于对象动作的所述多个摄像机中的每个的多个关键帧;

响应于从所述多个摄像机中的大多数识别出事件的出现而识别出事件。

8. 根据权利要求1所述的监视系统,其特征在于,处理站中的处理器还配置成:

接收来自第一摄像机的对应于视频数据的第一帧的第一多个特征向量;

识别出来自第一摄像机的第一多个特征向量对应于对于事件的预定数据库中的事件而言的第一关键帧;

接收来自第二摄像机的、对应于视频数据的第二帧的第二多个特征向量;和

识别出来自第二摄像机的第二多个特征向量对应于来自事件的第二关键帧;和

生成分数,该分数对应于第一帧中的第一多个特征向量和第二帧中的第二多个特征向量之间的变形;和

参考所述分数来识别事件的出现或未出现。

9. 根据权利要求8所述的监视系统,其特征在于,处理站中的处理器还配置成:

识别对象的第一位置,该第一位置对应于从视频数据的第一帧提取的第一多个特征向量;

识别对象的第二位置,该第二位置对应于从视频数据的第二帧提取的第二多个特征向量;

实施单应变换,以便参考存在于视频数据的第一帧和视频数据的第二帧中的地平面来识别对象的第一位置与对象的第二位置之间的距离;和

参考第一位置和第二位置之间的距离而生成对应于第一帧中的第一多个特征向量和第二帧中的第二多个特征向量之间的变形的分数。

10. 一种用于监视场景的方法,包括:

通过第一摄像机中的传感器生成场景的第一视频数据,第一视频数据包括第一系列帧;

通过第一摄像机中的特征提取处理器识别第一视频数据中的第一多个特征向量;

在第一操作模式中通过第一摄像机中的网络装置仅将第一多个特征向量传输至处理站;

在第二操作模式中仅响应于处理站对第一视频数据的请求通过第一摄像机中的网络装置将第一视频数据传输至处理站；

通过第二摄像机中的另一传感器生成场景的第二视频数据，第二视频数据包括第二系列帧，第二摄像机与第一摄像机相比从不同的位置生成场景的第二视频数据；

通过第二摄像机中的另一特征提取处理器识别第二视频数据中的第二多个特征向量；

在第一操作模式中通过第二摄像机中的另一网络装置仅将第二多个特征向量传输至处理站；

在第二操作模式中仅响应于处理站对第二视频数据的请求通过第二摄像机中所述另一网络装置将第二视频数据传输至处理站；

通过处理站中的另一网络装置接收来自第一摄像机的第一多个特征向量和来自第二摄像机的第二多个特征向量；

通过处理站中的事件处理器参考第一和第二多个特征向量来识别场景中的对象和对象动作；

通过处理站中的事件处理器参考事件的预定数据库来识别对应于场景中的对象动作的事件；

通过处理站中的事件处理器生成用于传送来自第一摄像机和第二摄像机中的至少一个的视频数据的请求；和

通过视频显示装置生成从第一摄像机和第二摄像机中的至少一个接收的视频数据的图形显示，并通过视频输出装置显示与所述事件相关的对象。

11. 根据权利要求10所述的方法，其特征在于，该方法还包括：

通过第一摄像机中的特征提取处理器来识别第一视频数据中的一帧的对应于场景的背景的一部分；和

通过第一摄像机中的特征提取处理器仅在所述一帧中的与场景的背景不对应的部分中识别第一视频数据中的第一多个特征向量。

12. 根据权利要求11所述的方法，其特征在于，该方法还包括：

通过第一摄像机中的特征提取处理器而生成边界框，该边界框对应于所述一帧的与场景的背景不对应的部分；

通过第一摄像机中的特征提取处理器对所述一帧的在边界框中的所述部分调整大小，以生成所述一帧的被包括在边界框中的所述部分的固定分辨率图像；和

通过第一摄像机中的特征提取处理器参考所述固定分辨率图像而在第一多个特征向量中识别至少一个特征向量。

13. 根据权利要求12所述的方法，其特征在于，识别至少一个特征向量还包括：

通过第一摄像机中的特征提取处理器在固定分辨率图像中识别多个图像梯度；和

通过第一摄像机中的特征提取处理器生成所述至少一个特征向量，所述至少一个特征向量包括与所述多个图像梯度对应的梯度直方图。

14. 根据权利要求10所述的方法，其特征在于，该方法还包括：

通过第一摄像机中的特征提取处理器而针对视频数据中的第一帧识别第一特征向量；

通过第一摄像机中的网络装置将第一特征向量传输至处理站；

通过第一摄像机中的特征提取处理器而针对视频数据中的第二帧识别第二特征向量，

第二帧在第一帧之后出现;和

通过第一摄像机中的网络装置将稀疏特征向量传输至处理站,该稀疏特征向量对应于第二特征向量的与第一特征向量不同的部分。

15.根据权利要求14所述的方法,其特征在于,该方法还包括:

通过处理站中的事件处理器接收来自第一摄像机的第一特征向量;

通过处理站中的事件处理器接收来自第二摄像机的其他稀疏特征向量,该其他稀疏特征向量仅包括第二摄像机中所生成的、在生成第二摄像机的较早特征向量后发生变化的特征向量部分,所述较早特征向量与来自第一摄像机的第一特征向量近似同时地生成;和

通过处理站中的事件处理器参考来自第一摄像机的第一特征向量而生成对应于来自第二摄像机的其他稀疏特征向量的完全特征向量。

16.根据权利要求10所述的方法,其特征在于,事件的识别还包括:

通过处理站中的事件处理器,参考对应于多个关键帧的第一多个特征向量和对应于所述多个关键帧的第二多个特征向量而在第一多个特征向量和第二多个特征向量中独立地识别出事件的出现或未出现,所述多个关键帧对应于单个时间段期间的对象动作;

响应于从第一摄像机和第二摄像机中的一个或两者识别出事件的出现而识别出事件。

17.根据权利要求10所述的方法,其特征在于,事件的识别还包括:

通过处理站中的事件处理器接收来自第一摄像机的对应于视频数据的第一帧的第一多个特征向量;

通过事件处理器识别出第一多个特征向量对应于对于事件的预定数据库中的事件而言的第一关键帧;

通过事件处理器接收来自第二摄像机的对应于视频数据的第二帧的第二多个特征向量;和

通过事件处理器识别出第二多个特征向量对应于对于事件的第二关键帧;和

通过处理站中的事件处理器生成分数,该分数对应于第一帧中的第一多个特征向量和第二帧中的第二多个特征向量之间的变形;和

通过处理站中的事件处理器参考所述分数来识别事件的出现或未出现。

18.根据权利要求17所述的方法,其特征在于,分数的生成还包括:

通过处理站中的事件处理器识别对象的第一位置,该第一位置对应于从视频数据的第一帧提取的第一多个特征向量;

通过处理站中的事件处理器识别对象的第二位置,该第二位置对应于从视频数据的第二帧提取的第二多个特征向量;

通过处理站中的事件处理器实施单应变换,以便参考存在于视频数据的第一帧和视频数据的第二帧中的地平面来识别对象的第一位置与对象的第二位置之间的距离;和

通过处理站中的事件处理器,参考第一位置和第二位置之间的距离而生成对应于第一帧中的第一多个特征向量和第二帧中的第二多个特征向量之间的变形的分数。

用于使用多个摄像机进行对象和事件识别的系统和方法

[0001] 优先权声明

[0002] 本申请要求2013年5月10日提交的、题为“SYSTEM AND METHOD FOR OBJECT AND EVENT IDENTIFICATION USING MULTIPLE CAMERAS”的美国临时专利申请No.61/822051的优先权,其全部内容通过引用并入本文。

技术领域

[0003] 本公开总体涉及视频监控的领域,且更特别地涉及用于使用围绕场景以不同角度布置的多个摄像机监控对象和事件的系统和方法。

背景技术

[0004] 视频监控系统出于包括安保和公共安全等各种目的而被广泛地部署。在通常的视频监控系统中,在不同的位置上部署一个或多个摄像机来监控活动。例如,视频监控系统生成公共场所、运输设施、零售商店、工业设施和住宅以及其他私有物业的图像。监控系统通常包括:数据存储装置,其将记录的视频中的某些或全部存档,以便以后查阅;和一个或多个视频输出装置,其能够回放实时的和存档的视频数据。

[0005] 在某些监控系统中,摄像机生成视频数据,该视频数据由一个或多个操作人员监控,该操作人员可查看视频中的活动并在查看到事件时采取适当的行动。例如,在零售商店处的监控系统中,操作者查看商店中的个人的实时视频,并在个人试图盗取商品的情况下通知安保人员。在某些视频监控系统中,多个摄像机从不同位置和角度记录单个场景的视频。尽管从多个角度生成视频可有助于收集关于场景的附加细节,但是多个视频记录使操作人员难以以高效的方式观察。另外,在联网的视频监控系统中,多个视频流耗费大量带宽和网络资源,尤其是在无线视频监控系统中。因此,改进视频监控系统、从而以自动的方式在记录的视频数据识别所关注的事件并以高效的方式利用网络带宽是有益的。

发明内容

[0006] 一种视频监视系统包括与中心处理站通信的分布式摄像机。中心处理站与使用背景减除方法来提取前景对象的多个摄像机通信。我们的系统中的摄像机将元数据传输至中心处理站。将对应于人的元数据从对应于其他对象的元数据过滤出来。通过中心处理站来分析对应于人的前景元数据,以便识别由人实施的动作和事件。摄像机使用无线通信网络或其他合适的通信通道与中心处理站通信。

[0007] 在一个实施例中,视频监视系统包括:多个摄像机,其位于多个位置以便记录场景。每个摄像机包括:传感器,其配置成生成场景的视频数据,该视频数据包括一系列帧;第一网络装置,其配置成将视频数据和与视频数据相关的特征向量传输至处理站;和特征提取处理器,其操作性地连接至所述传感器和所述网络装置。所述特征提取处理器配置成:识别由所述传感器生成的视频数据中的多个特征向量;在第一网络装置处于第一操作模式的情况下仅将所述多个特征向量传输至处理站;和在第一网络装置处于第二操作模式的情况

下仅响应于处理站对视频数据的请求而将视频数据传输至处理站。视频监视系统还包括处理站,其具有:第二网络装置;视频输出装置;和处理器,其操作性地连接至第二网络装置和所述视频输出装置。该处理器配置成:通过第二网络装置接收由所述多个摄像机中的每个摄像机生成的所述多个特征向量;参考从所述多个摄像机中的至少两个接收的所述多个特征向量而识别场景中的对象和对象动作;参考事件的预定数据库而识别与场景中的对象动作对应的事件;生成传送来自所述多个摄像机中的至少一个摄像机的视频数据的请求;和通过视频输出装置生成来自所述至少一个摄像机的视频数据的图形显示,以显示与所述事件相关的对象。

[0008] 在另一实施例中,发展了一种用于实施场景监视的方法。该方法包括:通过第一摄像机中的传感器生成场景的第一视频数据,第一视频数据包括第一系列帧;通过第一摄像机中的特征提取处理器识别第一视频数据中的第一多个特征向量;在第一操作模式中通过第一摄像机中的网络装置仅将第一多个特征向量传输至处理站;在第二操作模式中仅响应于处理站对第一视频数据的请求通过第一摄像机中的网络装置将第一视频数据传输至处理站;通过第二摄像机中的另一传感器生成场景的第二视频数据,第二视频数据包括第二系列帧,第二摄像机与第一摄像机相比从不同的位置生成场景的第二视频数据;通过第二摄像机中的另一特征提取处理器识别第二视频数据中的第二多个特征向量;在第一操作模式中通过第二摄像机中的另一网络装置仅将第二多个特征向量传输至处理站;在第二操作模式中仅响应于处理站对第二视频数据的请求通过第二摄像机中所述另一网络装置将第二视频数据传输至处理站;通过处理站中的另一网络装置接收来自第一摄像机的第一多个特征向量和来自第二摄像机的第二多个特征向量;通过处理站中的事件处理器参考第一和第二多个特征向量来识别场景中的对象和对象动作;通过处理站中的事件处理器参考事件的预定数据库来识别对应于场景中的对象动作的事件;通过处理站中的事件处理器生成用于传送来自第一摄像机和第二摄像机中的至少一个的视频数据的请求;和通过视频显示装置生成从第一摄像机和第二摄像机中的至少一个接收的视频数据的图形显示,并通过视频输出装置显示与所述事件相关的对象。

附图说明

[0009] 图1是视频监控系统的示意图。

[0010] 图2是示出了用于使用从观察场景的多个摄像机传输的元数据来识别事件的流程图的图。

[0011] 图3是在多数票决配置中针对从多个摄像机生成的事件的特征向量节点图表的图。

[0012] 图4是在多链条配置中针对从多个摄像机生成的事件的特征向量节点图表的图。

[0013] 图5是在多视场配置中针对从多个摄像机生成的事件的特征向量节点图表的图。

[0014] 图6是由监视系统中的多个摄像机生成的场景的一组图像。

具体实施方式

[0015] 为了有助于理解本文所述实施例的原理的目的,现在将参照附图和随后的书面说明中的描述。并非由此意图对主题的范围进行限制。说明书还包括对所示出的实施例的任

意改变和修改,且包括本公开所属领域的普通技术人员通常会想到的所述实施例的原理其他应用。

[0016] 在本文中,术语“场景”描绘了由监视系统使用多个摄像机监控的单个区域,所述多个摄像机位于多个位置处以便从不同方向查看场景。场景的示例包括、但不限于房间、门厅、广场、出入路线、街道、十字路口、零售商店、停车场设施等。

[0017] 在本文中,术语“稀疏编码”涉及用于生成相应于大量输入的数据的方法,所述数据使用多个“基向量”和“稀疏权重向量”而编码为向量。基向量使用惩罚优化过程(penalized optimization process)生成,所述惩罚优化过程应用于在训练过程期间提供的多个预定输入向量。在一个实施例中,现有技术已知的 l_1 优化过程用于生成相应于多个输入训练向量的基向量和稀疏权重向量。所使用的涉及向量或者矩阵的术语“稀疏”描述具有多个元素的向量或矩阵,其中给大多元素分配零值。在本文中,当应用于向量时术语“维度”涉及向量中的元素的数量。例如,具有三个元素的行向量或者列向量称为具有维度3,而另一个具有四个元素的行向量或者列向量称为具有维度4。

[0018] 在本文中,术语“元数据”涉及在视频或者其它传感器数据中所辨识的对象的特性。例如,如果对象沿随路径穿过视频摄像机的视场,则相应于对象的元数据可选地包括对象在视频数据的帧中的二维位置、对象的速度、对象的动作方向、对象的大小以及对象存在于摄像机的视场中的时间持续。如以下所描述的那样,参照对象的所观察的元数据来辨识事件。元数据不要求以特定性来辨识对象。在一个实施例中,元数据不辨识对象是特定的人或者甚至不辨识人类。然而,如果事件与预期的人动作相似,则替代的实施例推断出元数据相应于人、如朝一个方向并且以相应于人走过摄像机的速度动作的对象的元数据。附加地,仅仅对于短的时间追踪单个对象并且元数据不在延长的时间期间上辨识相同的对象。因此,除为了以后的调取存储视频数据拍摄以外,所存储的元数据和高兴趣事件根据元数据的辨识不需要个人可辨识信息(PII)的收集和存储。

[0019] 在本文中,术语“特征向量”或更简单地说“特征”是指与对象中的、在对象的视频数据中识别的区别结构对应的元数据向量。元数据中的每个元素也被称为“特征描述符”,且特征向量包括多个特征描述符。例如,人体或人体部分(比如臂和腿)的近似形状在视频数据中被识别。人体不同于周围环境,且特征向量包括在视频数据中描述人体的方面的数据,视频数据例如包括场景中的对象的大小、位置和方位。如果视频数据包括多个人,且每个人可使用单个特征向量来描述,或每个人可使用用于不同身体部分(比如臂、腿、躯干等)的多个特征向量来描述。

[0020] 在本文中,术语“词典”是指使用稀疏编码过程生成的多个基础向量。当词典在训练过程期间生成后,词典中的基础向量用于识别任意输入向量与用于在训练过程期间生成词典中的基础向量的输入向量之间的相似度。优化技术用于使用稀疏权重向量选择基础向量的组合,以生成估计任意输入向量的重构向量。重构的估计向量和实际的输入向量之间的识别误差提供了输入向量与词典之间的相似度的度量。

[0021] 在本文中,术语“关键帧”是指在由场景中人或对象实施的动作的视频序列中被看做是代表整个动作的图像帧。动作的视频序列通常包括两个或多个关键帧,且下文更详细描述的训练过程包括识别视频序列中的有限的N个关键帧。特定事件的每个视频序列包括相同数量的N个关键帧,每个关键帧所出现的时间可根据视频序列的角度以及在用作

训练数据的不同视频序列之间变化。在训练过程期间从一个或多个角度记录的受关注事件包括视频数据的一系列帧。例如,描述人从座位站立起来的视频序列是事件。注释者在来自围绕人定位的多个摄像机的视频流中的人站立的视频序列中识别关键帧。然后,事件处理器或另一合适的处理装置从所识别的关键帧提取特征,以便识别对应于事件的一系列特征向量。从不同视角描述由一个或多个个人或对象实施的相同事件的一组训练的多个视频序列形成了用于从视频序列中的每个选择关键帧的基础。在训练数据中从视频序列中选择的关键帧提取的特征形成了用于词典的基础,该词典并入数据库中,以用于识别由视频监视系统所监控的不同场景中的其他人或对象实施的类似动作。

[0022] 在本文中,术语“同步帧”是指视频数据的一帧,该帧在摄像机中生成且包括由摄像机中的特征提取处理器提取的特征以形成完全特征向量。完全特征向量包括与视频数据的帧中的被识别特征对应的所有数据。当对象(比如人)移动通过场景时,随后的图像帧中的视频数据捕获该移动,且特征提取处理器生成稀疏特征向量,所述稀疏特征向量仅包括被识别特征中相对于包括该特征的先前帧、比如同步帧的变化。在一些实施例中,视频摄像机以定期间隔(例如视频数据的每60帧一次)生成同步帧。本领域已知的特征向量提取技术包括、但不限于维数约减技术,包括主分量分析、边缘检测和尺度不变特征变换。在一些实施例中,场景中的被识别对象被编码以方向梯度直方图(HOG)外观特征描述符。如上所述,视频数据的关键帧在受关注事件期间以特定时间出现,且在摄像机操作期间不必对齐于同步帧和中间帧的生成。因此,在受关注事件期间生成的视频数据关键帧可通过摄像机中的同步帧或中间帧捕获。

[0023] 图1示出了视频监控系统100,其配置成记录关于场景中对象的视频数据并显示选择的视频以便由操作人员进行额外分析。视频监控系统100包括处理站160和分别被定位以从不同位置和角度记录场景112的多个摄像机108A-108N。处理站160还包括视频、对象特征和事件处理器104,对象和特征数据库106、网络装置164和视频输出装置168。在图1的实施例中,网络装置164是有线或无线数据网络适配器,且视频输出装置168包括一个或多个显示屏,比如LCD面板或其他合适的视频显示装置。

[0024] 在视频监控系统100中,处理站160中的特征处理器104包括一个或多个数字处理器,比如中央处理单元(CPU)、图形处理单元(GPU)、数字信号处理器(DSP)、专用集成电路(ASIC)等,其配置成执行存储程序指令,以便处理从摄像机接收的特征和事件数据以及从摄像机接收的视频数据。处理器104还包括一个或多个存储器装置,其存储编程的指令数据,以用于通过处理器104执行一个或多个软件程序。处理器104操作性地连接至数据库106、网络装置164和视频输出装置168。在操作过程中,处理站160通过网络装置164从摄像机108A-108N接收特征向量数据和可选地接收视频数据。处理站160中的处理器104通过将来自摄像机108A-108N中的一个或多个的特征向量数据结合存储在训练的对象特征和事件数据库106中的预定特征向量和事件数据进行合成来识别受关注对象和受关注事件。

[0025] 训练的对象特征和事件数据库106存储训练数据的词典。训练数据在系统100的训练阶段期间生成,且词典中与事件的不同部分对应的关键帧的特征基础向量通常不从移动通过场景112的相同对象生成,且通常由处于场景112之外的位置处的不同组摄像机记录。如下文所述,系统100清除场景的背景并重缩放被识别对象,以识别用于场景112中的新对象的特征向量,其独立于场景112的特定特征且不过度地依赖于不是训练过程的一部分的

单个人或对象的特征。因此,在系统100中,事件处理器104使用数据库106中存储的特征向量词典,以基于场景中在训练过程期间不使用的对象的动作来识别事件,该场景对应于除训练过程期间所使用的位置之外的位置。

[0026] 训练的对象特征和事件数据库106存储数据,所述数据对应于与先前被识别对象相关的多个预定特征和与先前被识别事件相关的特征移动序列。例如,数据库106存储特征向量数据,其对应于存在于场景112中并由视频摄像机108A-108N记录的人和其他对象的被识别形状。特征数据可包括以对应于视频摄像机108A-108N的视角和位置的角度从围绕场景的不同角度和位置观察的同一特征。事件数据包括针对场景中的一个或多个对象的一个或多个被识别特征的预定移动序列。例如,数据库106中的事件数据可包括对应于正在行走的人的一系列特征。走过场景112的另一人呈现类似特征。特征随着行走时人的腿和其他身体部位移动而变化。数据库106使用一个或多个非易失性和易失性数字数据存储装置来实施,包括、但不限于磁性硬盘、光驱、固态存储装置、静态和动态随机存储存储器(RAM)装置和其他合适的数字数据存储装置。

[0027] 在视频监控系统100中,摄像机108A-108N记录场景112的视频图像数据,识别与记录视频中的对象对应的特征数据,并将特征数据和视频数据的一部分传输至事件处理器104。以摄像机108A为例,每个摄像机包括传感器140、特征提取处理器144、存储器148和网络装置152。传感器140包括一个或多个感测元件,比如记录场景112的视频的电荷耦合装置(CCD)或互补金属氧化物半导体(CMOS)图像传感器,且传感器140配置成例如以单色、彩色或近红外生成来自场景112的数字图像数据。在另一实施例中,摄像机包括用于检测远红外频带中的图像的红外传感器。在一些实施例中,传感器140还集成有透镜、反射镜和本领域已知的其他摄像机光学装置。特征提取处理器144包括一个或多个数字处理器,比如中央处理单元(CPU)、图形处理单元(GPU)、数字信号处理器(DSP)、专用集成电路(ASIC)等,其配置成执行存储的程序指令,以处理来自传感器140的图像数据和使用一个或多个特征提取技术识别场景112中的一个或多个对象的特征向量。存储器120存储用于特征提取处理器144的程序指令,且可选地将传感器140在摄像机操作过程中生成的视频数据的缓冲存储在存储器148中。如下文所述,处理站160响应于识别出摄像机108A-108N中的一个记录了事件而可选地生成对缓冲的视频数据的请求。在一个实施例中,摄像机108A-108N中的网络装置152通过无线数据网络(例如,无线局域网(WLAN)或无线宽域网(WWAN))传输数据至处理站160中的相应的网络装置164。

[0028] 在许多数据网络、包括无线网络中,将所有记录的视频数据和其他数据、包括提取的特征数据从每个摄像机传输至处理站160中的事件处理器104所需的带宽需要大量网络带宽。摄像机108A-108N可选地包括可见的、近红外或远红外照明源,且在一些实施例中摄像机包括用于弱光操作的图像增强器。

[0029] 摄像机108A-108N中的每个均包括特征提取处理器144,以便实施图像处理 and 特征提取处理。如下文更详细描述,摄像机108A-108N传输以定期间隔传输的同步帧中的视频中的对象的完全特征向量数据。特征数据包括数据向量,数据向量描述每个帧中所生成的视频数据中的对象的一个或多个特征。如上所述,同步帧是视频数据的帧,其中,摄像机中的处理器生成在视频数据的该帧中被识别的每个特征的完全特征数据。同步帧在摄像机操作过程中以定期间隔生成,且在同步帧之间生成的视频数据帧被称作中间帧。在视频数据的

每个中间帧之间,摄像机仅使用稀疏特征编码方案来传输特征更新,从而极大地降低了用于传输特征向量更新至事件处理器104的数据量和带宽需求。

[0030] 处理站160中的事件处理器104在操作过程中可选地请求来自摄像机108A-108N的一个或多个的完全视频数据。例如,响应于事件的识别,处理器104请求来自摄像机108A-108N中的一个或多个的视频数据,且视频输出装置168将视频显示给操作者,以便查阅。操作者可选地生成对来自其他摄像机108A-108N中的一个或多个的视频的附加请求。因此,在一个操作模式下,摄像机108A-108N的子集传输完全视频数据至处理器104,同时其他摄像机仅传输特征数据和特征更新数据。如上所述,摄像机108A-108N中的每个中的存储器120包括内部数据存储装置,其配置成将视频数据缓冲预定时间段,以使处理器104能够请求存储在摄像机中的附加视频数据。例如,摄像机108B中的存储器120包括数字数据存储装置,数字数据存储装置保留用于场景112的记录视频的先前10分钟的缓冲。摄像机108B生成并传输用于场景112中存在的对象(包括移动对象)的特征向量数据,并传输特征向量数据至处理器104。如果受关注事件出现在场景112中,则处理器104的操作者请求与事件出现所经历的被识别时间对应的完全视频数据,且摄像机108B从数据存储装置取回请求的视频。因此,即使摄像机108B不传输完全视频数据至处理器104,处理器104也可选地取回视频数据,以用于系统100中的选择的受关注事件。

[0031] 在系统100中,数据库106包括训练模型,训练模型用于从摄像机108A-108N传输至中心处理站160的特征向量元数据识别受关注事件的出现。训练在使用系统100以对场景进行监视之前实施,且训练过程通常在与场景112的位置不同的位置处在受控的条件下实施。在一个实施例中,中心处理站160和事件处理器104配置成实施训练过程,而在另一实施例中单独的计算系统实施训练过程,且来自训练过程的数据存储在训练的对象特征和事件数据库106中,以便在系统100的操作过程中使用。

[0032] 训练过程包括一系列试验,其中,人或其他对象实施对应于受关注事件的动作,且动作从多视角被记录为视频。手动注释过程包括一个或多个注释者从视频序列中的每个中选择有限数量的关键帧,以便帮助生成对于每个受关注事件中所出现的人或对象移动的训练模型。在一个实施例中,训练期间对于关键帧的手动选择过程包括容易使用界面。该过程被简化成可由机械土耳其工人(mechanical turk worker)来进行。提供给土耳其工人指令以注释数据,从而获得关键帧。在一个实施例中,尽管用于选择关键帧的训练过程手动地实施,但是特征提取过程和训练词典数据的附加生成在无需人介入的情况下以自动方式实施。

[0033] 例如,在一个实施例中,数字处理装置接收训练数据中的来自特定受关注事件的多个视频序列的视频数据关键帧。在一个配置中,所述多个视频序列包括在受关注事件中从实施单个动作的单个人或对象的不同位置和角度取得的视频。所述多个视频序列还包括在受关注事件中在多个试验期间实施动作的多个人或对象的记录,以改进训练数据的幅度和精度。每个试验由主体在他或她面向不同方向时和在摄像机的视场中的不同位置处实施。在用于系统100的一个训练过程中,试验使用八个不同的定向、如相对于摄像机以弧度

0 、 $\frac{\pi}{4}$ 、 $\frac{\pi}{2}$ 、 $\frac{3\pi}{4}$ 、 π 、 $\frac{5\pi}{4}$ 、 $\frac{3\pi}{2}$ 和 $\frac{7\pi}{4}$ 来实施。

[0034] 训练过程使用分别被生成以描述事件c的出现的一组M视频序列来生成模型,该模

型包括对于一个或多个事件c的外观特征描述符参数模板和变形参数。例如,事件c包括人正在踢他或或她的腿的动作,且训练数据包括从实施该踢腿的多视角记录的、来自一个或多个人训练主体的踢腿的M视频序列。针对给定的事件c的训练组被称为 $\{D_q\}$ ($q=1,2,\dots,M$)。训练过程使用打分函数 $s(p_q|D_q,w^o)=\langle w^o, \Phi(D_q,p_q) \rangle$,其中, w^o 是向量,该向量包括训练过程提炼的作为训练模型的一部分的所有外观和变形参数,且 $\Phi(D_q,p_q)$ 是对应于特定标签 p_q 的相应的外观和变形能量。

[0035] 在某些监视系统实施例中,视频监控过程不仅需要识别单个受关注事件,而且还需要识别多个受关注事件并区分不同的受关注事件。在一个实施例中,训练过程对于每个受关注事件使用多类分类(one-vs-all)学习策略,并在给出在训练过程期间生成的其他受关注事件的硬色调(hard negative)示例的情况下生成共同地检测和识别任何特定受关注事件的模型参数。在一个实施例中,训练过程使用采用下列对象学习函数的支持向量机(SVM)机制:

$$[0036] \quad \operatorname{argmin}_{w^c, \{\xi_q\}, \{\eta_{q,q'}\}} \frac{1}{2} \|w^c\|^2 + \lambda_1 \sum_q \xi_q + \lambda_2 \sum_{q,q'} \eta_{q,q'}$$

$$[0037] \quad \text{s.t. } \forall q, \langle w^c, \Phi(D_q, p_q) - \Phi(D_q, \hat{p}) \rangle \geq \Delta(p_q, \hat{p}) - \xi_q$$

$$[0038] \quad \forall q, q', \langle w^c, \Phi(D_q, p_q) - \Phi(D_{q'}, p_{q'}) \rangle \geq \Delta(p_q, p_{q'}) - \eta_{q,q'}$$

[0039] 在上述SVM机制方程中, λ_1 和 λ_2 是在优化模型期间最小化松弛值(slack value)的用户限定式缩放参数。导向关键帧标签为 $\hat{p}$ 的约束是指成本惩罚函数(cost penalization function),或“损失”函数 $\Delta(p_q, \hat{p})$,其中,关键帧标签 $\hat{p}$ 基于训练过程期间生成的所观察的(“地面实况”)关键帧 p_q 而被惩罚(penalize)。非负松弛项(non-negative slack term) ξ_q 提供了抗约束违反的附加稳定性。导向地面实况标签 p_q 的约束表明,对于特定动作的第q个样本给定任何地面实况标签 p_q 的情况下,在过滤通过另一违规容纳较链损失(violation accommodating hinge-loss)项 $\eta_{q,q'}$ 之后,训练数据中任何其他受关注事件的第q'个样本的任何地面实况标签 $p_{q'}$ 产生较低的分值。

[0040] 损失函数 $\Delta(p_q, \hat{p})$ 在训练过程期间用于反映特定假设标签 $\hat{p}$ 在多大程度上匹配预定地面实况标签 p_q 。在一个实施例中,损失函数是二进制损失函数,其中,在 $\hat{p}$ 匹配 p_q 的情况下, $\Delta(p_q, \hat{p}) = 1$,否则, $\Delta(p_q, \hat{p}) = 0$ 。

[0041] 上述训练过程生成具有外观参数和变形参数的模型,其可用于分类在监视系统100的操作过程中后来所观察的多个受关注事件。然而,训练过程易于为受关注事件中的某些动作分配较高权重,从而可导致对于某些受关注事件的误分类。该偏差使用从训练模型生成的分数数据的中值来估计,使用预定训练数据作为输入,如下列方程所提出的: $b^c = \operatorname{median}\{S(p_1|D_1, w^c), \dots, S(p_M|D_M, w^c)\}$ 。在一个实施例中,训练过程估计与不同的受关注事件c相关的偏差 b^c 。在系统100中,偏差数据存储在数据库106中,且用于在事件识别过程期间将分数标准化,以降低误分类受关注事件的可能性。

[0042] 图2描述了过程200,其用于操作监视系统100,以便生成记录的视频中的特征向量数据并将特征向量数据传送至中央处理系统,从而用于识别对象和受关注事件。过程200在

训练过程已经生成对应于预定数量的受关注事件的用于数据库106的模型参数后进行。在下文的说明中,提及过程200实施动作或函数是指操作处理器(包括摄像机和中央处理系统中任一或两者中的处理器)以执行被编程的指令,以便结合视频监控系统中的其他部件来实施该动作或函数。出于描述性目的,结合图1的视频监控系统100来描述过程200。

[0043] 在过程200期间,视频摄像机108A-108N中的一个或多个生成场景112的记录的視頻,且每个摄像机中的特征提取处理器144实施从视频图像数据的背景减除(框204)。在图2中,出于描述性目的而示出了摄像机108A和108B。每个摄像机从不同位置记录场景112的视频图像数据。摄像机108A和108B相应地记录视频,比如图像帧202A和202B。每个摄像机中的特征提取处理器144减除与在生成视频期间不移动或变化的场景112静态部分对应的图像数据部分,比如场景112中的墙壁或地面。图像206A和206B描述场景112中具有黑背景的人,其代表减除了视频中的背景图像数据。背景减除维持了静态场景的动态演变的背景图像。该背景图像由于一天中场景的光线微小变化而缓慢演变,或对象可从场景放置或清除。动态背景图像用作与由摄像机传感器捕获的每个新图像进行比较的参考图像。摄像机108A和108B中的每个中的特征提取处理器144识别捕获的图像与参考图像的区别,以便提取一个或多个前景对象的轮廓影像(silhouette)。拆分的前景轮廓影像对应于场景中的不同的对象或人,且被分配以不同的识别码。

[0044] 过程200继续,从而摄像机108A和108B中的每个中的特征提取处理器144从图像数据中的前景对象提取特征(框208)。在前景像素位置处的摄像机传感器强度(intensities)在减除背景后对于对象的每个轮廓影像被提取,以形成对于每个对象的前景图像。每个摄像机中的处理器在前景图像上生成最小区域的矩形边界框,且处理器将图像区域调整大小至预定的固定分辨率(resolution)。特征提取处理器144在固定分辨率图像处生成网格,以便在网格中形成分别包括相同像素数的每个框。特征提取处理器144识别每个网格-框中的图像梯度,且特定特征向量在每个网格-框中在图像梯度的直方图中被识别。一旦独立特征向量对于网格中的每个框被识别,特征向量就彼此增补(append)以形成一大的特征向量,例如使用具有HOG描述符的固定大小的5x5网格排列。因此,一个固定大小的特征向量对于图像中的每个前景对象被识别。

[0045] 如上所述,包括前景图像的边界框被调整大小,以便生成固定分辨率图像。例如,高度和大小不同的、或在离摄像机的两个不同距离处的两个人可使用从场景112的视频生成的特征向量来进行比较。因此,在固定分辨率的前景图像上提取特征向量的过程提供了照明不变性、尺度不变性和一些旋转不变性。

[0046] 过程200继续,从而每个摄像机将特征数据描述符向量压缩并传输至事件处理器104(框212)。由于场景中的人的姿势随时间过去而逐渐变化,因此在相继的帧上他们相应的特征向量之间存在高度相关性。图像210A和210B描绘图像中的特征,从而摄像机108A和108B中的每个中的特征提取处理器进行编码,以便传送至处理站160。摄像机108A和108B中的特征提取处理器144随压缩方案实施相关性,且在相继的帧上仅特征向量中的小更新被压缩和传输。特征提取处理器144使用稀疏编码机制来压缩特征向量更新。特征提取处理器144在视频数据的同步帧期间定期地重生成完全特征向量,以计及场景112中的新对象并防止从稀疏特征向量生成过程形成过度的噪声误差。实施稀疏编码和压缩的优点包括降低了传输至事件处理器104的数据量,且相关性方法跟踪前景中的每个独立人或移动对象,由此

能够预测对象移动的路径。摄像机108A和108B中的每个使用每个摄像机中的网络装置152以压缩格式将对于同步帧的完全特征向量数据以及稀疏特征向量数据传输至处理站160中的网络装置164。

[0047] 在随系统100使用的摄像机的一个实施例中,摄像机108A和108B中的每个对于传输完全特征描述符数据的同步帧期间在场景中所识别的每个对象以5x5x32排列的特征描述符数据传输800字节数据。另外,特征描述符的稀疏性实现了特征描述符数据的附加压缩。因此,除非中心处理站160响应于识别出由摄像机108A和108B中的一个或两者观察的受关注事件而生成对完全视频数据的请求,否则摄像机108A-108B仅传输元数据至中心处理站160。相比而言,使用现有技术的H.264视频压缩算法对于是特征描述符数据的大约两倍数量级的640x480像素分辨率帧的彩色图像数据提供了每图像64K字节的平均比特率。

[0048] 过程200继续,从而处理站160中的事件处理器104接收来自摄像机108A和108B的压缩的特征向量数据,并解压缩特征向量数据(框216)。在单个无线摄像机与中心处理站进行通信的情况下,解压缩算法互补于以上提出的压缩算法。如果多于一个无线摄像机传输数据至中心处理站,则执行联合的解压缩方案,从而使用来自一个摄像机的信息来预测对于其他摄像机的更新。在联合的解压缩期间,处理站160从由两个或更多个摄像机针对场景112中的对象生成的多个稀疏特征向量重建完全特征向量。与独立解码每个摄像机的单独数据相比,联合的解压缩方案最小化了解压缩中的误差。

[0049] 过程200继续,从而识别来自摄像机的解压缩特征向量数据中的人或对象(框220)。在一个操作模式下,监控系统100配置成识别对应于人的特征向量并监控该人的动作。在不同配置中,其他实施例配置成识别其他对象(包括机动车辆或除人以外的其他动物)的动作。一些前景特征向量可能对应于人,而其他的可能对应于其他对象(比如小汽车、动物、自行车等)。特征和事件数据库106存储对应于人的且在训练过程期间生成的特征向量的组,以用于视频监控系统100。事件处理器104使用数据库106中的预定训练数据将对应于场景112中的人的特征向量从其他对象过滤。在一个实施例中,过滤对象以识别人物的过程使用对象分类器来实施。

[0050] 在过程200的一些实施例中,事件处理器104配置成从摄像机108A-108N所接收的元数据中识别当被识别对象(比如人)实施动作时发生的特定事件。在图2的示例性实施例中,视频数据包括多个帧,其中,场景112中的人240实施了踢腿动作。处理器104使用在预定时间段上从摄像机108A和108B接收的特征向量来实施踢腿事件的事件识别。事件识别过程是时间性的,因为事件在对应于视频数据的多个帧的多个特征向量中随着时间出现,且由于特征向量来自从不同位置记录事件的多个摄像机而是多视角的。事件识别过程从一个人或更多人的活动来识别事件,即使事件在图像的不同部分中被记录、以不同方向定向,以及即使事件以不同的速度实施。另外,处理器参考存储在数据库106中的所有预定事件来实时地识别事件。如果受关注事件被识别,事件处理器104就生成警告,以提示操作人员查阅事件。处理器104接收来自摄像机108中的一个或多个的事件视频,以便向操作人员提供事件视频回放。处理器104可选地生成事件分类,以便操作人员在观看视频数据之余进行查阅。例如,在图2的示例中,事件处理器104生成表示人已经实施了踢腿的警告(框224),且可选地生成对事件视频的请求,以便通过视频输出装置168显示,从而由操作人员进行查阅。

[0051] 如上所述,系统100包括训练的对象特征和事件数据库106,其存储在训练过程期

间对于多个受关注事件所识别的特征向量数据。使用图2的踢腿事件为例,数据库106存储在训练过程期间从记录一个或多个人实施踢腿的视频所生成的特征向量。该踢腿的记录被称为“样例”,且样例在计算机视觉和机器学习群体中通常叫做“训练数据”。充足的训练数据针对受关注类别的每个事件而被记录。例如,在一个实施例中,训练过程收集在多个试验上实施受关注事件中的每个的多个主体的数据。

[0052] 在训练过程期间,一些事件图像帧被选择为关键帧。对于每个动作,预定数量的关键帧(比如六个关键帧)从每个试验的视频数据手动地被选择。关键帧代表提供了关于被实施的动作的最大信息的姿势/姿态帧。使用上述过程200中针对特征向量提取的相同方法,特征向量针对每个关键帧中的人使用子系统被提取。关键帧特征向量形成了训练数据库。在系统100中,从训练数据生成的事件词典随训练的对象特征和事件数据库106存储。

[0053] 在视频监控过程期间,将与单个人的动作对应的每个特征向量与存储在事件数据库106中的特征向量进行比较。然而,两个特征向量对于不同动作的单个帧而言可能非常相似。例如,人行走的单个帧可能不能区别于人跑动的单个帧。因此,对于每个动作,将查询图像序列的特征向量与关键帧序列进行比较,以便清除关于监控系统中所记录的动作的模糊性。另外,来自多个摄像机的信息需要被溶合,以清除来自不同位置中的多个摄像机所生成的特征向量的模糊性,因为一些动作可由于隐藏(occlusion)而对一些摄像机视野而言是不可见的。例如,垂直于摄像机定向的人的一个手臂对该摄像机而言是不可见的,且因此朝着同一人的其他摄像机将捕获他的隐藏手臂。

[0054] 在系统100中,处理站160中的事件处理器104使用针对每个受关注事件的图形模型,以随时间变化和从多视角识别事件。图形模型构型是一种在多个摄像机视野上捕获多个关键帧之间的相互影响的概率性模型。在一个实施例中,模型包括M个关键帧和N个摄像机视野,以用于图像中的一共 $N \times M$ 个节点。不同配置的图像包括节点之间的多个连接布置。连接的每个选择对于识别事件而言具有不同的特性。图像的边缘对于时间边缘中的动作而编码关键帧之间的时间差,以及沿着空间边缘的前景对象边界框空间距离。

[0055] 图3描绘了图像300中的节点的一种布置,以用于使用多数票决模型识别事件。在多数票决模型中,仅对应于摄像机中的每个的节点在时间上线性地连接起来。因此,图3的图像300包括对于观察事件的每个摄像机的一个链。任何事件出现的可能性的推断使用标准方法、如动态编程针对每个链而独立地实施。在图3的配置中,每个摄像机在决定特定事件是否被观察到方面被分配以相同的权重。在系统100中,事件处理器104接收来自摄像机108A-108N的特征向量数据,并识别来自每个摄像机的特征向量图像是否形成表示事件已经被观察到的多数情形。如果来自多数摄像机的特征向量图像不同意事件已经出现,事件处理器104就识别出没有事件出现(未出现事件),且不请求来自摄像机的完全视频。

[0056] 在图3中,对于每个摄像机,对应于由摄像机识别的关键帧的节点形成为单独的子图像。帧的视频序列表示为D,且任何特定受关注事件表示为定向的图像G,其中,每个节点是对于受关注事件而言的关键帧。图像G由通过边缘E连接的节点V的集合表示。V节点的数量表示为M。在图像中,任何给定的节点 $i \in \{1 \dots M\}$ 具有定位(anchor position) $p_i = (x_i, y_i, t_i)$,其中, (x_i, y_i) 代表围绕图像中的对象所生成的边界框的中心中的像素位置,且 t_i 代表视频序列中的帧号,其作为针对帧何时生成而言的时间参考。图像中的边缘E指定了哪对关键帧节点被约束成具有相关性。例如,图3中,用于每个摄像机的节点通过链中的边缘彼

此连接,对应于特定受关注事件所对应的关键帧的基于时间的序列,且关键帧以预定序列出现。在其他实施例中,该机制更一般,且图像中的边缘不需要是连续的。例如,一些事件可包括对于特定动作的可变次数的重复。对于重复一次或多次的动作,图像可选地包括在关键帧之间形成循环(cycle)的跳跃边缘。

[0057] 在一些实施例中,事件处理器104使用可变形关键帧模型来识别关键帧之间的对象的特征描述符的变化。图3中,事件处理器104生成分数,其对应于每个图像生成分数的可能性, $S(p|D,w) = \sum_{t \in V} \langle w_t, \Phi_{app}(D, p_t) \rangle + \sum_{t_i, j \in E} \langle w_{ij}, \Phi^{def}(p_i, p_j) \rangle$ 其中, $app(D, p_t)$ 是在帧时刻 t_i 检测的对象的HOG或其他特征描述符,且 $\Phi^{def}(p_i, p_j)$ 基于从摄像机中的一个或多个接收的特征描述符元数据中的变化模式化(model)了成对的帧(帧 i 和 j) 之间的对象变形。对于由单个摄像机所生成的一系列图像帧,该变形被表述为: $\Phi^{def}(p_i, p_j) = [dx; dx^2; dy; dy^2; dt; dt^2]$, 其中, $dx = x_i - x_j$ (x位置上的变化), $dy = y_i - y_j$ (y位置上的变化), 且 $dt = t_i - t_j$ (时间上的变化)。为了将视频的帧的特征向量匹配于数据库106的词典中的模板 w_i , 事件处理器104识别与视频 D 中的位置 p_i 的处的特征向量的最大内积响应(inner product response)。两个帧之间的变形权重 w_{ij} 模式化了该模型中在时间上关键帧的对之间的马氏距离(Mahalanobis distance)。用于马氏距离的参数在模型训练期间生成并存储在数据库106中。

[0058] 图4描绘了另一图像配置400, 其中, 一个摄像机充当根节点, 所述根节点在时间上连接至由系统100中的其他摄像机所生成的关键帧。如图4所示, 所有其他摄像机视野连接至来自所选择的摄像机的节点。在该情况下也使用动态编程来进行推断。图4的配置被称为多视场链图形模型。在图4的示例中, 摄像机2被选择为根节点。图4中, 摄像机404被选择为参考摄像机。在一个实施例中, 系统100使用单个摄像机作为参考摄像机, 比如在单个摄像机与其他摄像机相比具有更高分辨率的系统中, 或对于与系统100中的其他摄像机相比定位在特定位置、以便障碍更少地识别场景中的事件的关键帧的摄像机而言。在另一实施例中, 系统100基于先于观察场景的其他摄像机而检测到事件中的关键帧的摄像机而自动地分配参考摄像机。因此, 在系统100中, 对于一个受关注事件而言的参考摄像机可不同于对于另一受关注事件而言的参考摄像机。摄像机404观察事件并生成用于该事件的关键帧, 比如关键帧406A-406M。摄像机408和412对于事件的关键帧(比如关键帧410A-410M和414A-414M)相应地生成附加图像数据。

[0059] 如图4所示, 摄像机中的每个生成来自事件的关键帧, 但是摄像机不必须同时地生成关键帧。例如, 图4中, 与分别来自摄像机410和412的相应的关键帧410B和414B相比, 参考摄像机404更早地生成关键帧406B。事件处理器104接收与来自摄像机的这些关键帧中的每个对应的特征数据, 并识别出与来自参考摄像机404的关键帧406B对应的关键帧410B和414B, 即使关键帧不是恰好同时地生成。在一个实施例中, 事件处理器104将时间约束施加于帧, 这意味着, 当关键帧彼此在相对较短的时间段内出现时, 事件处理器104识别来自与相同事件的不同视野对应的不同摄像机的关键帧。例如, 在一个实施例中, 事件处理器104向由其他摄像机生成的关键帧施加时间加权的比例(scale), 以便识别该关键帧与来自参考摄像机的关键帧相比对应于同一受关注事件的同一部分的可能性(likelihood)。例如, 如果关键帧410B在关键帧406B的100毫秒内出现, 则加权比例分配关于两个关键帧彼此对应的高可能性(例如90%), 而1秒的较长延迟具有关于两个关键帧彼此对应的相应的低可

能性(例如25%)。

[0060] 在一个实施例中,事件处理器104将针对图3的单个摄像机所描述的分数识别过程延伸至包括观察事件的参考摄像机和一个或多个附加摄像机的多个摄像机。图4的图像描绘了随着带定向边缘的关键帧的节点,定向边缘连接针对来自参考摄像机404的关键帧系列的节点404A-404N。图像400还包括连接至来自非参考摄像机408和412的关键帧节点的边缘,其连接至参考摄像机中的相应关键帧。如上所述,事件处理器104基于来自记录同一事件的不同摄像机的关键帧中的每个的检测之间的时间接近性来识别来自不同摄像机的关键帧之间的边缘连接。

[0061] 在图4的多摄像机配置中,事件处理器104以与图3所示的单摄像机配置类似的方式来生成分数 $S(p|D,w)$ 。然而,变形函数 ϕ^{def} 在来自多个摄像机的关键帧数据被用于生成分数时发生变化。为了模式化来自从多个角度和位置记录场景中的对象的摄像机的关键帧之间的变形,事件处理器104使用场景中对摄像机中的每个而言都可见的固定式或“同形(homographic)”特征。例如,在许多监视系统配置中,所有摄像机都具有场景中的单个地平面(ground plane)的视野。地平面不是必须地需要是人行走所在的物理地面,而是从不同角度和位置观察对象时充当用于识别对象和时间的参考点的共同几何平面。事件处理器104认定同形性 H_i ,以用于观察地平面的参考摄像机 r 和任何其他摄像机 l 之间的共同地平面。同形性是用于将平面的一个视野中的像素和相应的特征映射(map)于另一视野的线性变换,且事件处理器104使用同形性来识别不同摄像机的视野中的对象之间的距离。另外,当事件处理器104识别典型地与地平面接触的人的动作时,将围绕每个场景中的对象所形成的边界框的底边角(bottom corner)连接起来的线的中心充当对于场景中对象的3D位置的代替物(proxy)。因此,在地平面或其他同形元素处于不同摄像机的视野内且对象保持近似于同形元素的情况下,事件处理器104可识别从所述不同摄像机观察的同一对象的位置。在监视系统100中,多个摄像机保持共同地平面的视野和地平面上的对象,以用于事件处理器104使用单应变换来识别不同摄像机的视野中的对象的位置。

[0062] 给定摄像机 l 的视野中地平面上像素的齐次坐标 $\mathbf{c}^i = (x^i, y^i, 1)^T$,像素的位置相对于参考摄像机 r 被估计为 $\mathbf{r}^i = H_l^T \mathbf{c}^i$ 。对于这两个视野的变形函数被限定为: $\phi^{\text{def}}(\mathbf{r}^i, \mathbf{r}^j) = [dx, dx^2, dy, dy^2]$,其中, $[dx, dy] = (\mathbf{r}^i - H_l^T \mathbf{r}^j)^T$ 。在一个实施例中,变形函数被模式化为弹簧函数(spring function),其中,实施变形的代价对应于拉伸弹簧所需的力的大小。图6描绘了场景中的同一人(人604)的不同视野602A、602B和602C之间的变形约束的图形表示,该场景具有对三个不同摄像机可见的共同地平面608。事件处理器104在分别对应于视野602A-602C中的每个的特征描述符数据中相应地识别边界框606A-606C的底部中心处的相应的位置612A-612C。边界框位置612A-612C对于单应变换分别与地平面608接触。

[0063] 图5描绘了另一图像500,其中,从摄像机中的每个生成的关键帧彼此连接。在图5的机制中,每个节点都跨越时空连接至其相邻物,如图5所示。对于该模型的推断可使用迭代动态编程来进行,且多个空间迭代散布在时间迭代之间。在一个配置中,事件处理器104将图像图5处理为一组类似于图4的图像,其中,在图像500的处理的一个迭代期间,每个摄像机都被当作参考摄像机。事件处理器104随后联合对于通过图像500的不同迭代的分数,以便识别对应于来自不同摄像机的关键帧的最可能的事件。在一个实施例中,迭代数是修

正先验的(fixed a priori)。图5的配置被称为多视场,其中,对应于所有摄像机的节点彼此连接。

[0064] 在过程200期间,中心处理站160使用上述图形模型在来自摄像机108A和108B的特征向量元数据中使用事件处理器104和数据库106中的预定模型处理所检测的关键帧以识别特定受关注事件(比如图2所示的踢腿事件),并响应于识别出受关注事件而采取行动。事件处理器104实施推断操作以从数据库106中的预定组的受关注事件识别“最佳”事件 c^* 和相应的标签 p^* 。“最佳”事件 c^* 是指在给定来自摄像机的关键帧元数据的所观察序列的情况下出现的可能性最高的事件和相应的其他事件的未出现。标签 p^* 是指在训练过程期间确定的并存储在数据库106中的、与事件 c^* 关联的人或机器可读的标识符。事件处理器104通过最大化分数函数来识别 c^* 和 p^* ,如下列方程所给出的: $\{c^*, p^*\} = \operatorname{argmax}_{p, c \in \{1 \dots C\}} (S(p|D, w^c) - b^c)$,其中, w^c 是存储在数据库106中的对于事件 c 中的一个的模板,且 b^c 是训练过程期间所生成的偏差因子,其从原始分数被减除以计及模型训练过程期间的偏差。

[0065] 应当理解,上述特征以及其它特征和功能、或者它们的替代性方案,可期望地被组合到许多其它不同的系统、应用场合或方法中。随后可由本领域的技术人员作出的各种目前未预见或未料想的替代性方案、修改、改变或改进也意图由权利要求所包含。

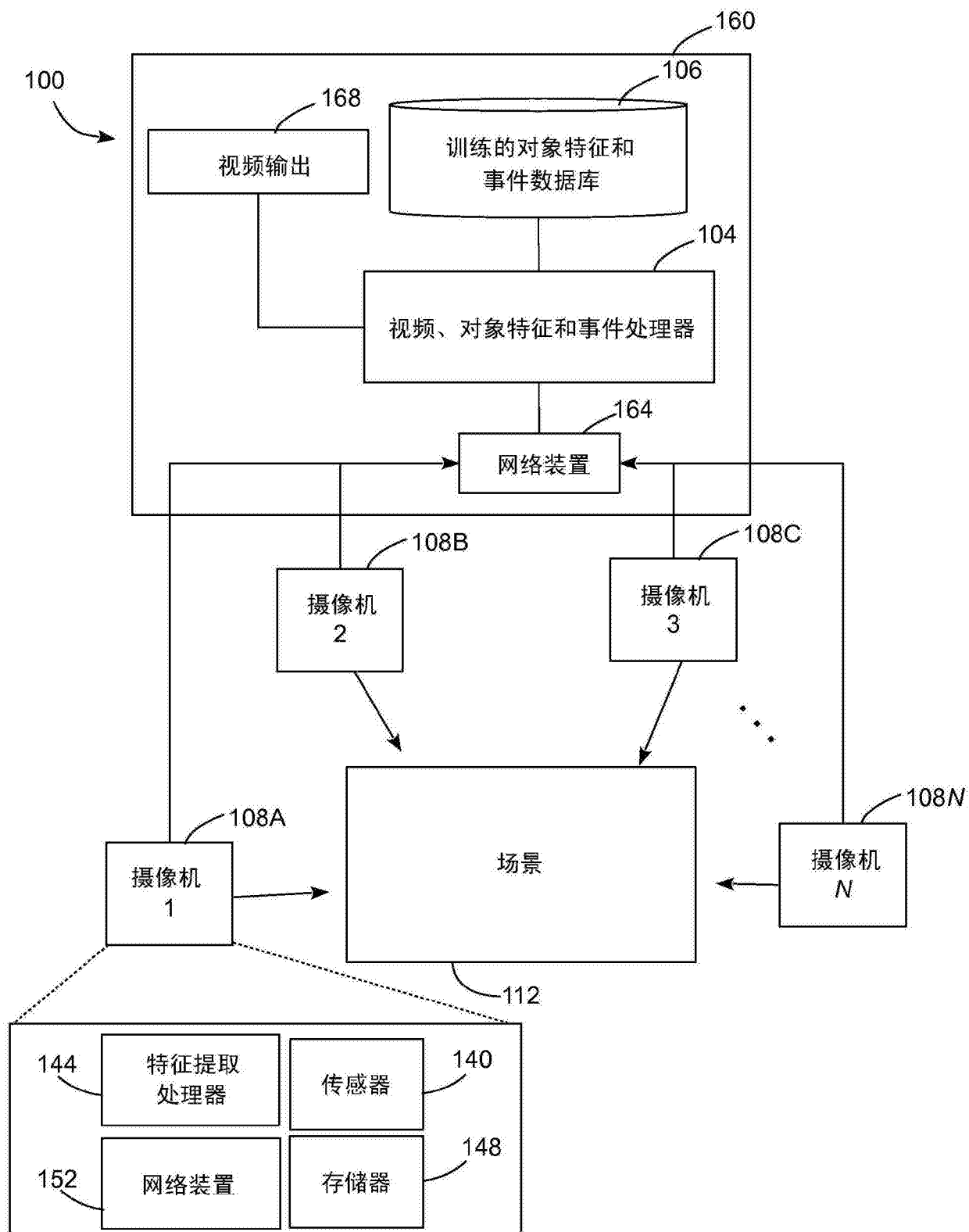


图1

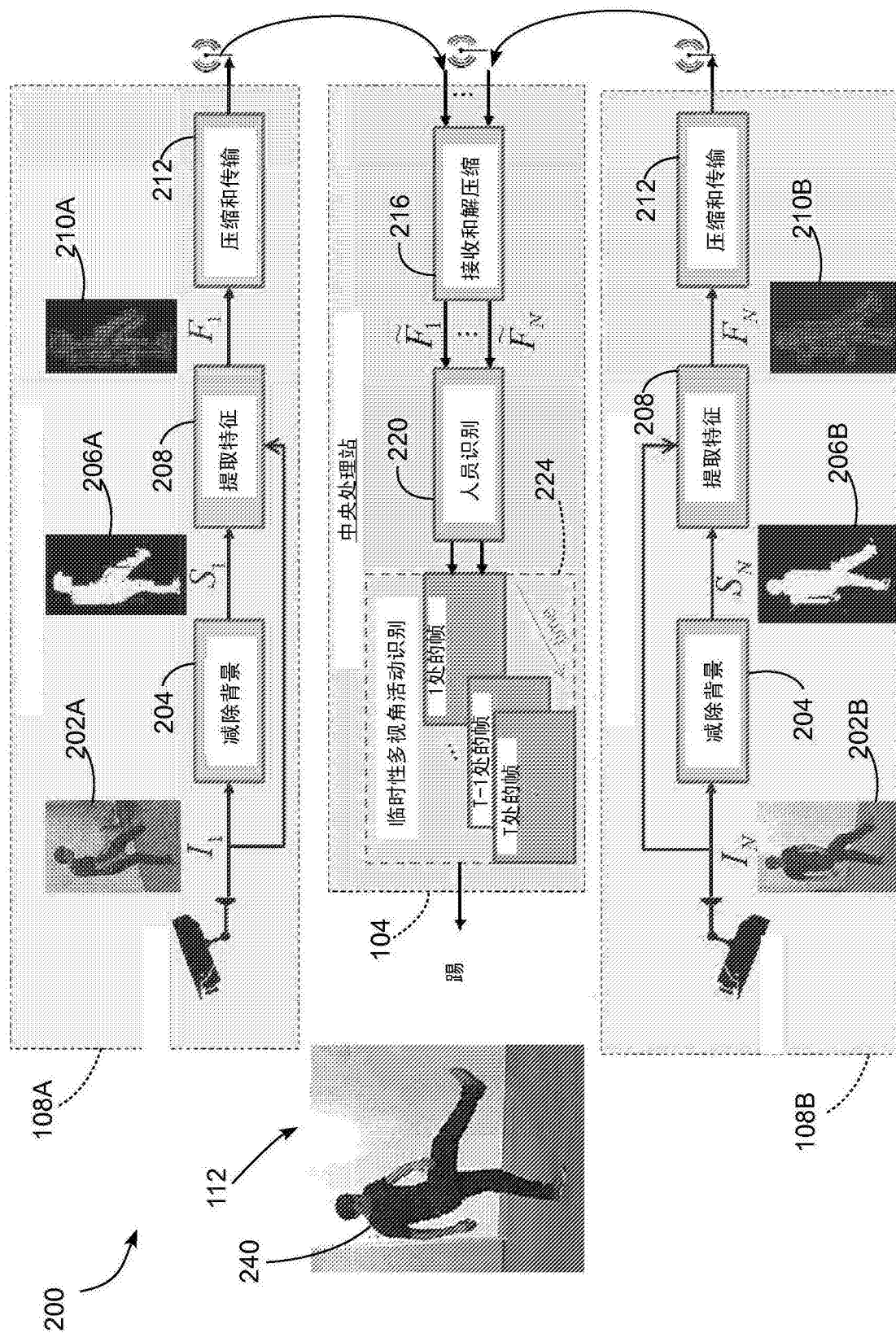


图2

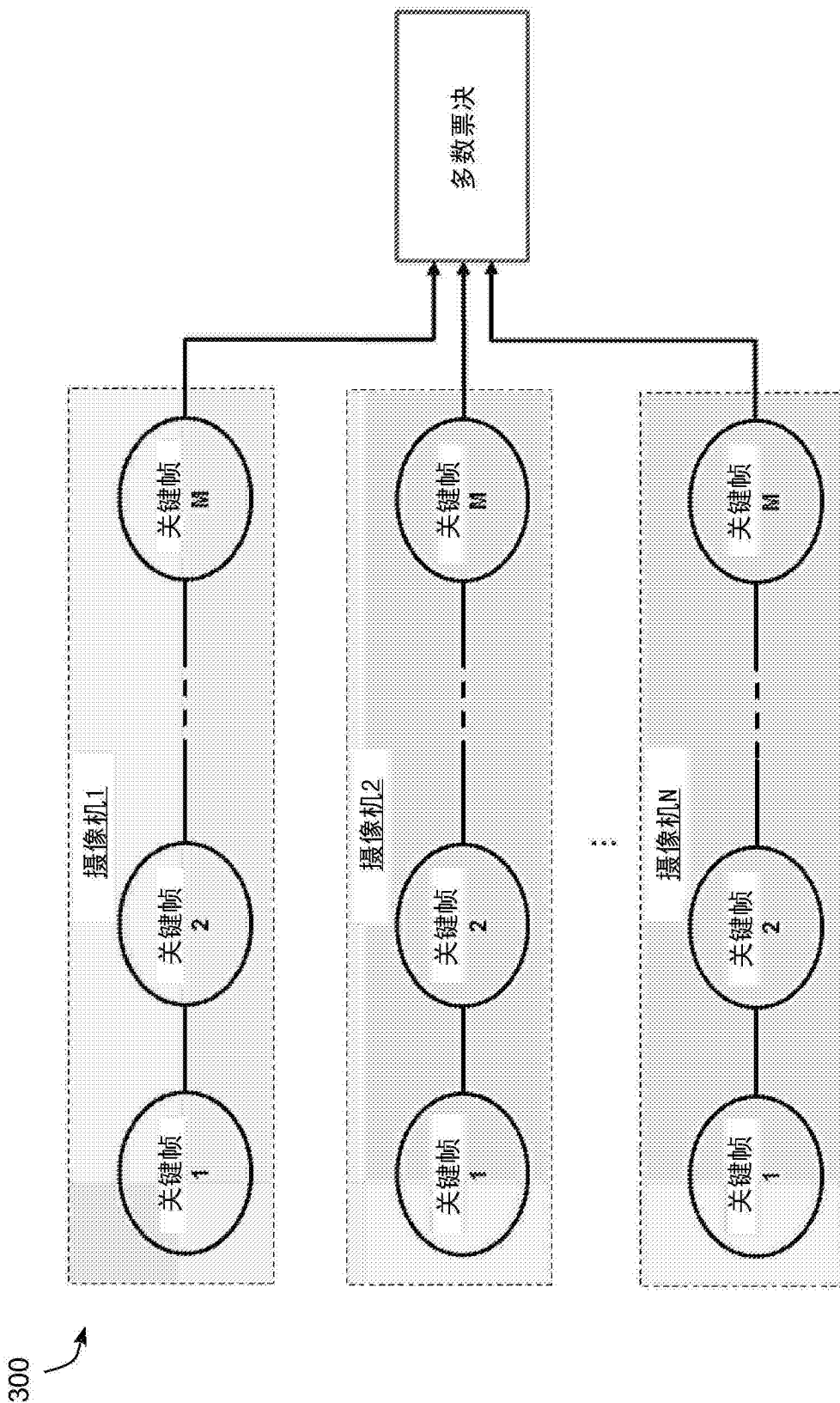


图3

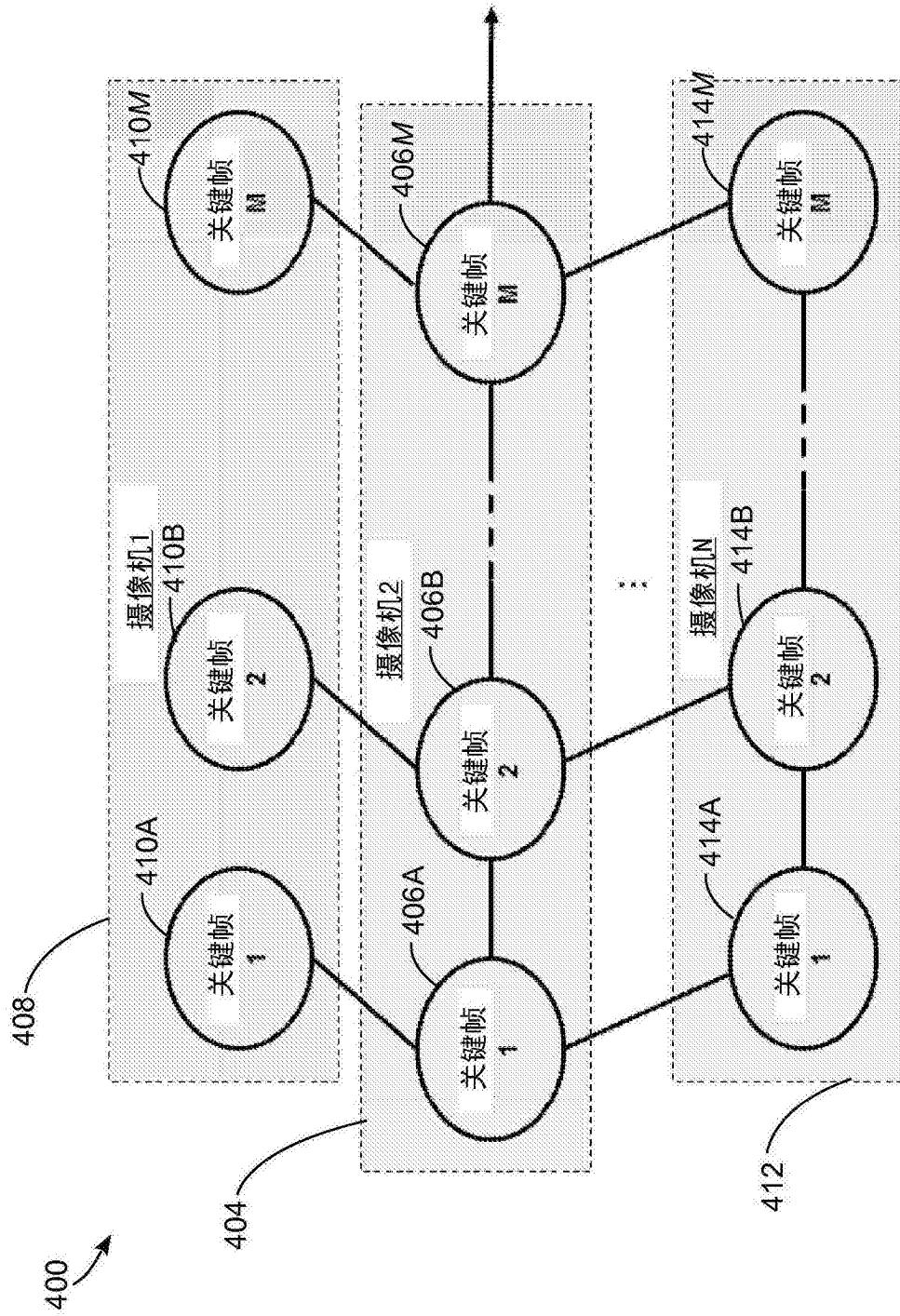


图4

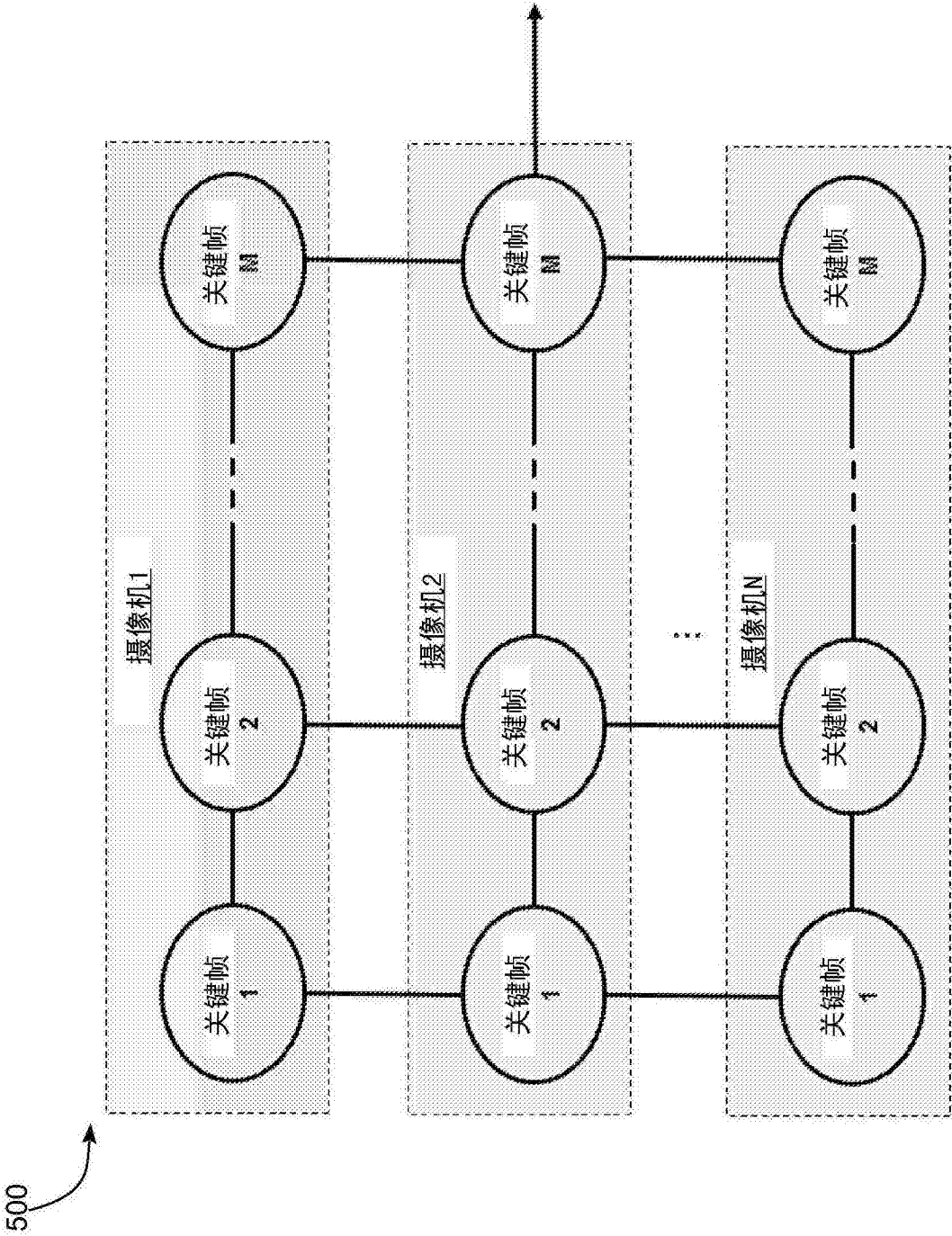


图5

图6 - 全连接的场图形模型

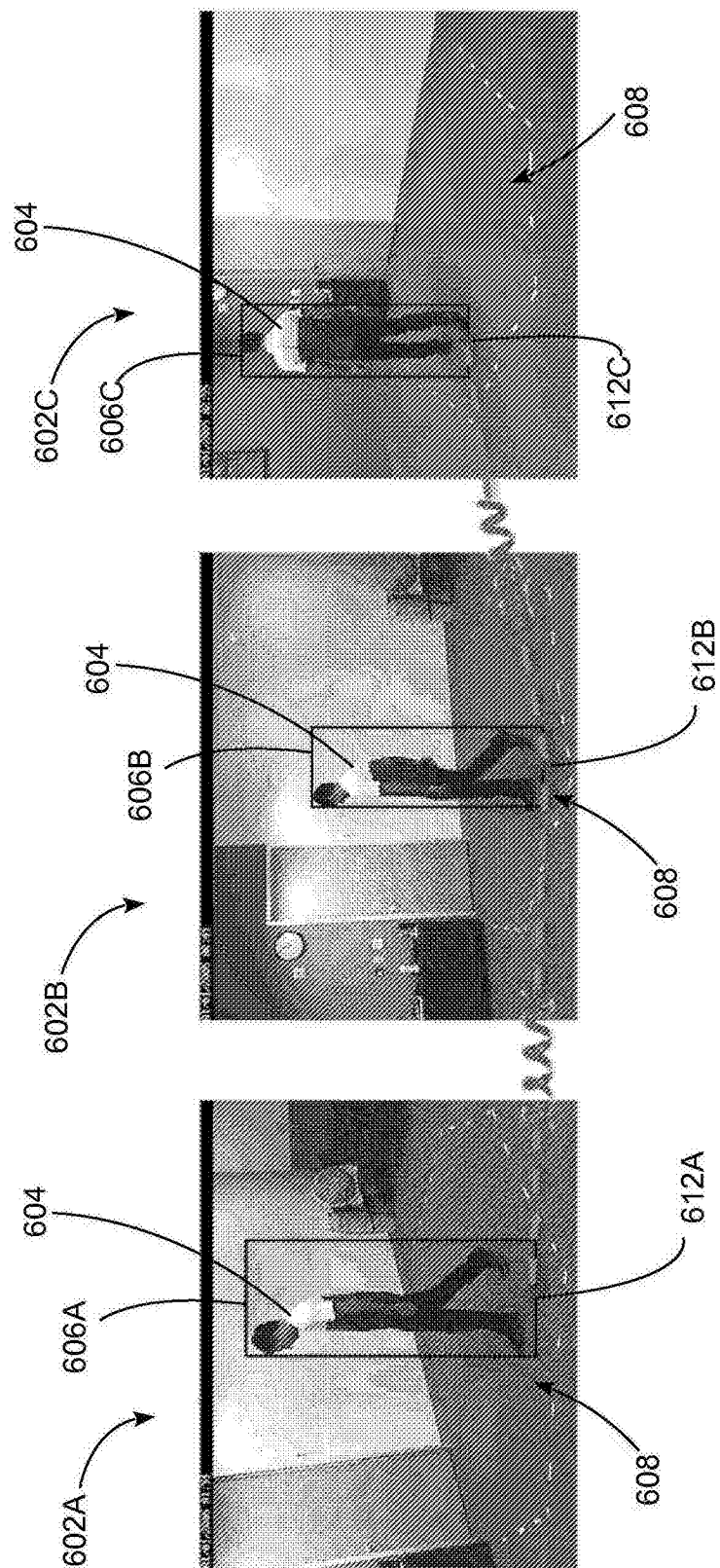


图6